Journal of
Geophysics
Zeitschrift für
Geophysik
Volume 46 1979


Springer International

Journal of Geophysics – Zeitschrift für Geophysik

This journal was founded by the Deutsche Geophysikalische Gesellschaft on the initiative of L. Mintrop in 1924 as the Zeitschrift für Geophysik and edited by G. Angenheister from Vol. 1–18 (1944). It reappeared in 1954 edited by B. Brockamp from Vol. 19–26 (1960), and edited by W. Dieminger and J. Untiedt from Vol. 27 (1961). After Vol. 40 (1970) the title was changed to Journal of Geophysics – Zeitschrift für Geophysik.

Published: Vols. 19–39 by Physica-Verlag, Würzburg, from Vol. 40 by Springer Berlin, Heidelberg, New York.

Springer-Verlag Berlin Heidelberg New York

Printed in Germany by Universitätsdruckerei H. Stürtz AG Würzburg

Author Index

Subject Index

J. Geophys. 46, 1–14, 1979

Journal of Geophysics

Wave Propagation in Stratified Anisotropic Media

An Algorithm for the Computation of the Reflection and Transmission Coefficients as Well as of the Fields*

L. Bossy

Université Catholique de Louvain, B-1348 Louvain-la-Neuve and
Institut Royal Météorologique, B-1180 Bruxelles, Belgium

Abstract. An algorithm based on the accurate computation of the matrizants of elementary slabs of a stratified medium and also on the properties of the propagation matrix is developed. It leads to the step by step computation of the reflection and transmission matrices of the medium and to the computation of the independent solutions of the differential equation. The multiplication of the matrizants and thus the swamping of the solutions are avoided.

The algorithm is explained in the scope of the special problem of the propagation of E.M. waves in the ionosphere; it works also provided the transmission matrix has the required properties.

Key words: Wave propagation – Anisotropic media – Reflection coefficient – Transmission coefficient.

Introduction

There are numerous studies devoted to the propagation of light and of radio waves through stratified media. Bibliographical references can be found in the textbook of Born and Wolf (1970) and in the review paper of Budden (1969)[1].

The main goal of these studies is either the determination of the coefficients of reflection and transmission of the medium or the determination of the fields with respect to the altitude or both.

The mathematical problem to be solved is the integration of a vectorial differential equation of the form:

$$\frac{d}{dz}\mathbf{f}+ik_0\mathbf{T}\mathbf{f}=\mathbf{0}$$

* Dedicated to Professor Dr. K. Rawer on the occasion of his 65th birthday

[1] Among the more recent papers, it is worth mentioning those of: Altman and Cory (1969a and b); Altman et al. (1970); Altman and Fijalkow (1970); Altman and Postan (1971); Berreman (1972); Honig and den Engelsen (1977); Nagano et al. (1975); Teitler and Henvis (1970)

0340-062X/79/0046/0001/$02.80

where

f is a vector with an even number ($2p$) of components;

k_0 is a wave number of reference;

T is a square matrix ($2p \times 2p$) which depends on k_0, on the direction of the wave normal in the vacuum and on the physical parameters of the medium at each height. Moreover, this matrix has the property that one half of its eigenvalues have a negative imaginary part and the other half have a positive imaginary part. Such matrices will be named matrices of class τ_{2p}.

For instance, the matrix **T** is of class τ_2 for isotropic media; it is at least of class τ_4 for anisotropic ones.

The purpose of this study is to develop an algorithm based on the matrizants (or transfer matrices) of elementary layers on which the matrix **T** is approximated by means of a matrix polynomial and on the use of the fact that the matrix **T** is of class τ_{2p} in order to avoid the numerical swamping of the solutions during the integration.

The statement of the method will be made in the frame of the propagation of E.M. waves in an horizontally stratified ionosphere, using the language familiar in the field.[2]

An important part of the algorithm has already been stated in an earlier paper (Bossy and Claes, 1974).

1. Vectorial Differential Equation

The differential equation governing the propagation of E.M. waves in a stratified medium is derived from the Maxwell equations

$$\begin{aligned} &\operatorname{curl} \mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B} = -\frac{1}{c}\boldsymbol{\mu}\frac{\partial}{\partial t}\mathscr{H} \\ &\operatorname{curl} \mathbf{H} = \frac{\partial}{\partial t}\mathbf{D} \quad \text{or} \quad \operatorname{curl} \mathscr{H} = \frac{1}{c}\boldsymbol{\varepsilon}\frac{\partial}{\partial t}\mathbf{E} \end{aligned} \tag{1.1}$$

with

$$\begin{aligned} &\mathscr{H} = \sqrt{\frac{\mu_0}{\varepsilon_0}}\mathbf{H}, \quad \mathbf{B} = \mu_0 \boldsymbol{\mu}\mathbf{H} = \frac{1}{c}\boldsymbol{\mu}\mathscr{H} \\ &\mathbf{D} = \varepsilon_0 \boldsymbol{\varepsilon}\mathbf{E}, \quad \sqrt{\frac{\mu_0}{\varepsilon_0}}\mathbf{D} = \frac{1}{c}\boldsymbol{\varepsilon}\mathbf{E} \end{aligned} \tag{1.2}$$

where the tensors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ depend only on the height variable z.

---

[2] The basic formulation for non horizontal stratifications can be found in Rawer and Suchy (1967, p. 147)

Expressing the fields in the form:

$$\begin{aligned}\mathbf{E} &= \mathbf{e}(z)\exp i[\omega t - k_0(n_1 x + n_2 y)]\\ \mathscr{H} &= \mathbf{h}(z)\exp i[\omega t - k_0(n_1 x + n_2 y)]\end{aligned} \tag{1.3}$$

where

$$\begin{aligned}k_0 &= \omega/c\\ \mathbf{n}_0 &= n_i \mathbf{e}_i = \sin\theta_0\cos\Phi_0\,\mathbf{e}_1 + \sin\theta_0\sin\Phi_0\,\mathbf{e}_2 + \cos\theta_0\,\mathbf{e}_3\end{aligned} \tag{1.4}$$

are defined with respect to the vacuum, one can transform (1.1) in a system of 6 equations; namely in a matricial form:

$$\begin{bmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{e}\\ \mathbf{h}\end{bmatrix} + ik_0\begin{bmatrix}\mathbf{0} & \boldsymbol{\mu}\\ -\boldsymbol{\varepsilon} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{e}\\ \mathbf{h}\end{bmatrix} = 0 \tag{1.5}$$

where the matrix $\mathbf{D}$ represents the operator

$$\mathbf{D} = \begin{bmatrix} 0 & \dfrac{d}{dz} & -ik_0 n_2\\ \dfrac{d}{dz} & 0 & ik_0 n_1\\ ik_0 n_2 & -ik_0 n_1 & 0\end{bmatrix}. \tag{1.6}$$

The third and sixth equations are algebraic; they allow the elimination of e_z and h_z, so that finally the fourvector

$$\mathbf{f} = (e_x, e_y, h_x, h_y)^T \tag{1.7}$$

is the solution of the homogeneous differential equation (Clemmow and Heading, 1954)

$$\frac{d}{dz}\mathbf{f} + ik_0\,\mathbf{T}\mathbf{f} = \mathbf{0} \tag{1.8}$$

where

$$\mathbf{T} = \begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2\\ \mathbf{T}_3 & \mathbf{T}_4\end{bmatrix} \tag{1.9}$$

with[3]

$$\begin{aligned}T_{1,jk} &= -n_j\varepsilon_{3k}/\varepsilon_{33} + (-1)^{j+k+1} n_{3-k}\mu_{3-j,3}/\mu_{33}\\ T_{2,jk} &= \varepsilon_{3kl} n_j n_l/\varepsilon_{33} + (-1)^{j+1}(\mu_{3-j,k}\mu_{33} - \mu_{3-j,3}\mu_{3k})/\mu_{33}\\ T_{3,jk} &= -\varepsilon_{3kl} n_j n_l/\mu_{33} + (-1)^{j}(\varepsilon_{3-j,k}\varepsilon_{33} - \varepsilon_{3-j,3}\varepsilon_{3k})/\varepsilon_{33}\\ T_{4,jk} &= -n_j\mu_{3k}/\mu_{33} + (-1)^{j+k} n_{3-k}\varepsilon_{3-j,3}/\varepsilon_{33}.\end{aligned}$$

[3] The indices j, k and l are equal to 1 or 2 and the elements ε_{mnp} are such that $\varepsilon_{mnp} = \varepsilon_{npm} = \varepsilon_{pmn} = -\varepsilon_{mpn} = -\varepsilon_{nmp} = -\varepsilon_{pnm}$ with $\varepsilon_{123} = 1$

As can be seen, the matrix $\mathbf{T}$ depends only on the pulsation ω, on the direction parameters n_1 and n_2 and on the constitutive tensors of the medium. One remarks that trace $\mathbf{T}=0$ for every medium where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are diagonal tensors; this is for instance the case for any isotropic medium.

More precisely, for such isotropic media one has

$$\boldsymbol{\varepsilon}=\varepsilon\mathbf{I}, \quad \boldsymbol{\mu}=\mu\mathbf{I} \tag{1.10}$$

where $\mathbf{I}$ is the identity matrix, so that

$$\mathbf{T}=\begin{bmatrix} 0 & 0 & \dfrac{n_1 n_2}{\varepsilon} & \mu-\dfrac{n_1^2}{\varepsilon} \\ 0 & 0 & -\mu+\dfrac{n_2^2}{\varepsilon} & -\dfrac{n_1 n_2}{\varepsilon} \\ -\dfrac{n_1 n_2}{\mu} & -\varepsilon+\dfrac{n_1^2}{\mu} & 0 & 0 \\ \varepsilon-\dfrac{n_2^2}{\mu} & \dfrac{n_1 n_2}{\mu} & 0 & 0 \end{bmatrix}. \tag{1.11}$$

On account of the particular structure of $\mathbf{T}$, (1.8) can be put in the form of the differential equation

$$\frac{d}{dz}\begin{bmatrix} e_x \pm i\sqrt{\mu/\varepsilon}\, h_x \\ e_y \pm i\sqrt{\mu/\varepsilon}\, h_y \end{bmatrix} + ik_0 \begin{bmatrix} \mp i\dfrac{n_1 n_2}{n} & \mp i\dfrac{n^2-n_1^2}{n} \\ \pm i\dfrac{n^2-n_2^2}{n} & \pm i\dfrac{n_1 n_2}{n} \end{bmatrix} \begin{bmatrix} e_x \pm i\sqrt{\mu/\varepsilon}\, h_x \\ e_y \pm i\sqrt{\mu/\varepsilon}\, h_y \end{bmatrix} = 0 \tag{1.12}$$

where

$$\varepsilon\mu=n^2=(M-i\chi)^2. \tag{1.13}$$

The secular equation corresponding to the matrix $\mathbf{T}$ in (1.12) is

$$\lambda^2=n^2-(n_1^2+n_2^2)$$

its eigenvalues have the property required by the matrices of class τ_2.

In the general case, the matrix $\mathbf{T}$ of (1.8) is of class τ_4. As a matter of fact, if one considers the propagation in a homogeneous medium having the physical characteristics of the medium at the level $\bar{z}$, one has to solve the equation with constant coefficients

$$\frac{d}{dz}\mathbf{f}+ik_0\mathbf{T}(\bar{z})\,\mathbf{f}+\mathbf{0} \tag{1.14}$$

the solution of which is (Gantmacher, 1966, p. 119)

$$\mathbf{f}=\mathbf{f}(\bar{z})\, e^{-ik_0\mathbf{T}(\bar{z})(z-\bar{z})}. \tag{1.15}$$

Quite generally, this solution can be expressed in terms of linear combinations of the exponential functions

$$e^{-ik_0 q_\alpha(z-\bar{z})} \qquad (\alpha=1,\ldots,4)$$

where the q_α are the eigenvalues of the matrix $\mathbf{T}$. Booker (1936) has shown that in the present case two eigenvalues have a negative imaginary part and that the two corresponding eigenvectors represent waves the energy-flow of which is directed upwards and that the two other eigenvalues have positive imaginary parts with corresponding waves carrying energy downwards.

This specific property of the matrices $\mathbf{T}$ will be systematically used in the development of the algorithm.

2. Matrizant of an Elementary Layer

It is supposed that the medium considered is situated between the levels z_0 and $z_s (z_0<z_s)$ and that the integration domain is divided into s sub-domains or elementary layers through intermediary levels situated at $z_1, z_2, \ldots, z_{s-1}$.

On the nth elementary layer $z\in[z_{n-1}, z_n]$, one defines the matrizant $\mathbf{M}(z_{n-1}, z)$ through the relation (Volland, 1962b)

$$\mathbf{f}(z_{n-1})=\mathbf{M}(z_{n-1}, z)\,\mathbf{f}(z) \tag{2.1}$$

so that, introducing in (1.8), one gets

$$\frac{d}{dz}\mathbf{M}-ik_0\mathbf{M}\mathbf{T}=\mathbf{0} \quad \text{with } \mathbf{M}(z_{n-1}, z_{n-1})=\mathbf{I}. \tag{2.2}$$

One knows (Rawer and Suchy, 1967, p. 162) that if a polynomial development of $\mathbf{T}$ on $[z_{n-1}, z_n]$ is used such that

$$\mathbf{T}=\sum_{j=0}\mathbf{T}_j(z-z_{n-1})^j \tag{2.3}$$

then, the corresponding potential development of $\mathbf{M}$ can be written as

$$\mathbf{M}=\sum_{k=0}\mathbf{M}_k(z-z_{n-1})^k; \qquad \mathbf{M}_0=\mathbf{I} \tag{2.4}$$

where the matrices $\mathbf{M}_k$ are obtained through the recurrence relation

$$\mathbf{M}_k=\frac{ik_0}{k}\sum_{m=0}^{k-1}\mathbf{M}_{k-m-1}\mathbf{T}_m; \qquad \mathbf{M}_{k<0}=\mathbf{0}. \tag{2.5}$$

The matrizant ${}_n\mathbf{M}$ related to the nth slab is the matrix defined by

$${}_n\mathbf{M}=\mathbf{M}(z_{n-1}, z_n)=\sum_{k=0}\mathbf{M}_k(z_n-z_{n-1})^k. \tag{2.6}$$

One could verify (Bossy, 1971) that with an approximative representation of **T** by a polynomial of the fifth order it is possible for electronic distributions like those of the **D** region of the ionosphere to work with slabs of 1 km. thickness and to obtain ${}_n\mathbf{M}$ with a relative precision better than 10^{-7}. One could also ascertain that the number of terms needed for the development of ${}_n\mathbf{M}$ did practically not depend on the degree of the polynomial representation of **T**.

These remarks are no longer true in the neighbourhood of the levels where **T** diverges; it is then necessary to consider that z varies along a path distorted in the complex z-plane and that, after the circuit around the singularity, the integration proceeds again along the real axis (Budden, 1969). Proceeding in this way, the set of matrizants which can be computed corresponds either to a continuous profile (with some discontinuities in the first derivatives at the subdivision levels) or to a profile with discontinuities at some subdivision levels. Apart the fact that it needs the representation of the matrix **T** by the development (2.3), the volume of computation required exceeds not very much those needed in the method of Inoue and Horowitz (1966) but, for identical slab thicknesses, the approximation of the matrizants is much better.

The multiplicative property of the matrizants such that

$$\mathbf{f}(z_0)=\mathbf{M}(z_0,z_n)\,\mathbf{f}(z_n)={}_1\mathbf{M}\,{}_2\mathbf{M}\,{}_3\mathbf{M}\ldots{}_n\mathbf{M}\,\mathbf{f}(z_n) \tag{2.7}$$

is well known; one knows also (Rawer and Suchy, 1967, p. 158) that the matrizants are nothing else than particular wronskians of the differential equation. Noting the fact that, for the matrices **T** of class τ_{2p}, the particular solutions of the differential equation show extremely important relative variations; one knows that the solutions, the modulus of which grow, tend to mix with the solutions the modulus of which decrease with the effect that the independence of the solutions is destroyed as well as the significance and the usefulness of the wronskian. Therefore, it is fundamental that the integration algorithm avoids the product of matrizants if one needs to be safe from the swamping of the solutions.

3. Propagator and Diffusion Matrix

One considers an elementary slab $z\in[a,b]$, situated between two semi-infinite uniform media. The matrizant **M** related to this slab and the propagation matrices $\mathbf{T}_I$ and $\mathbf{T}_S$ of the uniform media $I(z\leqq a)$ and $S(z\geqq b)$ are known; are also known the matrices $\mathbf{Q}_I$ and $\mathbf{Q}_S$ the columns of which are the eigenvectors respectively of $\mathbf{T}_I$ and $\mathbf{T}_S$.

One knows that the relations

$$\mathbf{f}(a)=\mathbf{Q}_I\mathbf{c}(a);\qquad \mathbf{f}(b)=\mathbf{Q}_S\mathbf{c}(b) \tag{3.1}$$

define the decomposition of the vectors **f** with respect to the eigenvectors of **T** as they appear in **Q**. If the media I and S have been selected in such a way that there exist two upgoing and two downgoing waves, then the components of the vectors **c** are the amplitudes of these waves.

The relation between these amplitudes at the levels a and b is

$$\mathbf{c}(a)=\mathbf{Q}_I^{-1}\,\mathbf{f}(a)=\mathbf{Q}_I^{-1}\,\mathbf{M}\mathbf{f}(b)=\mathbf{Q}_I^{-1}\,\mathbf{M}\,\mathbf{Q}_S\,\mathbf{c}(b)=\mathbf{P}\mathbf{c}(b) \tag{3.2}$$

where (Lacoume, 1967) the matrix $\mathbf{P}$ is called the propagator related to the slab $[a,b]$ situated between the two uniform media I and S.

If one classes the eigenvectors in such a way that the two first columns of $\mathbf{Q}$ contain the upgoing waves and the two last the downgoing waves; then the fourvectors $\mathbf{c}$ can be split in two bivectors namely the bivector $\mathbf{u}$ for the upgoing waves and the bivector $\mathbf{d}$ for the downgoing waves. For instance, at the levels a and b, one has

$$\mathbf{c}(a)=\begin{bmatrix}\mathbf{u}_a\\ \mathbf{d}_a\end{bmatrix};\qquad \mathbf{c}(b)=\begin{bmatrix}\mathbf{u}_b\\ \mathbf{d}_b\end{bmatrix}. \tag{3.3}$$

The diffusion matrix $\mathbf{S}$ (Volland, 1962a) connects the amplitudes of the waves leaving a slab with the amplitudes of the waves entering this slab in the form

$$\begin{bmatrix}\mathbf{d}_a\\ \mathbf{u}_b\end{bmatrix}=\mathbf{S}\begin{bmatrix}\mathbf{u}_a\\ \mathbf{d}_b\end{bmatrix}=\begin{bmatrix}\mathbf{R}_a^b & \mathbf{D}_b^a\\ \mathbf{D}_a^b & \mathbf{R}_b^a\end{bmatrix}\begin{bmatrix}\mathbf{u}_a\\ \mathbf{d}_b\end{bmatrix}. \tag{3.4}$$

The reflection matrices $\mathbf{R}$ and the transmission matrices $\mathbf{D}$ are 2×2 matrices; they are related to the 2×2 $\mathbf{P}_i$ matrices obtained when partitioning $\mathbf{P}$ according to

$$\mathbf{P}=\begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_2\\ \mathbf{P}_3 & \mathbf{P}_4\end{bmatrix} \tag{3.5}$$

through the relations (Volland, 1968)

$$\mathbf{R}_a^b=\mathbf{P}_3\mathbf{P}_1^{-1},\quad \mathbf{D}_a^b=\mathbf{P}_1^{-1},\quad \mathbf{D}_b^a=\mathbf{P}_4-\mathbf{P}_3\mathbf{P}_1^{-1}\mathbf{P}_2,\quad \mathbf{R}_b^a=-\mathbf{P}_1^{-1}\mathbf{P}_2. \tag{3.6}$$

It appears that as well the propagator as the diffusion matrix depend on the structure of the medium through the matrizant and on the choice of the contiguous uniform media through the eigenvectors of their propagation matrices. It is worth remembering that the choice of these contiguous media is arbitrary as far as their eigenvectors correspond to an equal number of ascending and descending waves.

4. Diffusion Matrix of Two Contiguous Slabs

Starting with the diffusion matrices of two contiguous slabs, one looks for the expression of the diffusion matrix of the sum of the slabs in terms of those of the individual slabs. More precisely, one considers the slab $[a,b]$ situated between the uniform media I and S and the slab $[b,c]$ situated between the media $I^x\equiv S$ and S^x for which one has

$$\begin{bmatrix}\mathbf{d}_a\\ \mathbf{u}_b\end{bmatrix}=\begin{bmatrix}\mathbf{R}_a^b & \mathbf{D}_b^a\\ \mathbf{D}_a^b & \mathbf{R}_b^a\end{bmatrix}\begin{bmatrix}\mathbf{u}_a\\ \mathbf{d}_b\end{bmatrix};\qquad \begin{bmatrix}\mathbf{d}_b\\ \mathbf{u}_c\end{bmatrix}=\begin{bmatrix}\mathbf{R}_b^c & \mathbf{D}_c^b\\ \mathbf{D}_b^c & \mathbf{R}_c^b\end{bmatrix}\begin{bmatrix}\mathbf{u}_b\\ \mathbf{d}_c\end{bmatrix} \tag{4.1}$$

and one looks for the expression of

$$\begin{bmatrix} \mathbf{d}_a \\ \mathbf{u}_c \end{bmatrix} = \begin{bmatrix} \mathbf{R}_a^c & \mathbf{D}_c^a \\ \mathbf{D}_a^c & \mathbf{R}_c^a \end{bmatrix} \begin{bmatrix} \mathbf{u}_a \\ \mathbf{d}_c \end{bmatrix} \tag{4.2}$$

in terms of the known diffusion matrices with the condition that the slab $[a,c]$ be situated between the uniform media I and S^x.

One starts from the two equations

$$\begin{aligned} \mathbf{u}_b - \mathbf{R}_b^a \mathbf{d}_b &= \mathbf{D}_a^b \mathbf{u}_a \\ -\mathbf{R}_b^c \mathbf{u}_b + \mathbf{d}_b &= \mathbf{D}_c^b \mathbf{d}_c \end{aligned}$$

eliminating $\mathbf{d}_b$, one obtains

$$(\mathbf{I} - \mathbf{R}_b^a \mathbf{R}_b^c)\, \mathbf{u}_b = \mathbf{D}_a^b \mathbf{u}_a + \mathbf{R}_b^a \mathbf{D}_c^b \mathbf{d}_c$$

or

$$\mathbf{u}_b = (\mathbf{I} - \mathbf{R}_b^a \mathbf{R}_b^c)^{-1} (\mathbf{D}_a^b \mathbf{u}_a + \mathbf{R}_b^a \mathbf{D}_c^b \mathbf{d}_c)$$

introducing $\mathbf{u}_b$ in

$$\mathbf{u}_c = \mathbf{D}_b^c \mathbf{u}_b + \mathbf{R}_c^b \mathbf{d}_c$$

one has

$$\begin{aligned} \mathbf{u}_c &= \mathbf{D}_b^c (\mathbf{I} - \mathbf{R}_b^a \mathbf{R}_b^c)^{-1} (\mathbf{D}_a^b \mathbf{u}_a + \mathbf{R}_b^a \mathbf{D}_c^b \mathbf{d}_c) + \mathbf{R}_c^b \mathbf{d}_c \\ &= \mathbf{D}_a^b \mathbf{u}_a + \mathbf{R}_c^a \mathbf{d}_c \end{aligned}$$

so that

$$\begin{aligned} \mathbf{R}_c^a &= \mathbf{R}_c^b + \mathbf{D}_b^c (\mathbf{I} - \mathbf{R}_b^a \mathbf{R}_b^c)^{-1} \mathbf{R}_b^a \mathbf{D}_c^b \\ &= \mathbf{R}_c^b + \mathbf{D}_b^c \mathbf{R}_b^a (\mathbf{I} - \mathbf{R}_b^c \mathbf{R}_b^a)^{-1} \mathbf{D}_c^b \,^4 \\ \mathbf{D}_a^c &= \mathbf{D}_b^c (\mathbf{I} - \mathbf{R}_b^a \mathbf{R}_b^c)^{-1} \mathbf{D}_a^b. \end{aligned} \tag{4.3}$$

One gets in a similar way

$$\begin{aligned} \mathbf{R}_a^c &= \mathbf{R}_a^b + \mathbf{D}_b^a (\mathbf{I} - \mathbf{R}_b^c \mathbf{R}_b^a)^{-1} \mathbf{R}_b^c \mathbf{D}_a^b \\ &= \mathbf{R}_a^b + \mathbf{D}_b^a \mathbf{R}_b^c (\mathbf{I} - \mathbf{R}_b^a \mathbf{R}_b^c)^{-1} \mathbf{D}_a^b \\ \mathbf{D}_c^a &= \mathbf{D}_b^a (\mathbf{I} - \mathbf{R}_b^c \mathbf{R}_b^a)^{-1} \mathbf{D}_c^b. \end{aligned} \tag{4.4}$$

With the aid of these relations, it is possible to determine step by step the submatrices of the diffusion matrices using operations which avoid the multiplication of matrizants, and which, except for the inverses, use only matrices the terms of which have modulus less than unity.

---

[4] Because of the identity

$$(\mathbf{I} - \mathbf{A}\mathbf{B})^{-1} \mathbf{A} \equiv \mathbf{A} (\mathbf{I} - \mathbf{B}\mathbf{A})^{-1}.$$

From a formal point of view, these relations are analogous to those obtained by Altman and Cory (1969b) in their application of the thin-film optical method (Born and Wolf, 1970, p. 51–70).

5. Reflection and Transmission Matrices of a Medium

One considers a stratified medium between the levels z_0 and z_s divided into s elementary slabs. With each subslab (say the nth), one associates a lower uniform medium I_n and an upper uniform medium S_n; the set of these media is subject to the conditions $S_n \equiv I_{n+1}$ and $I_1 = S_s$ the latter being related to the vacuum.

If one writes $\mathbf{u}_n = \mathbf{u}(z_n)$ and $\mathbf{d}_n = \mathbf{d}(z_n)$, one knows that

$$\begin{bmatrix} \mathbf{d}_o \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{R}_o^n & \mathbf{D}_n^o \\ \mathbf{D}_o^n & \mathbf{R}_n^o \end{bmatrix} \begin{bmatrix} \mathbf{u}_o \\ \mathbf{d}_n \end{bmatrix} = \mathbf{S}_n \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{d}_n \end{bmatrix} \tag{5.1}$$

where according to (4.3 and 4) the submatrices of S_n are obtained through the relations

$$\begin{aligned}
\mathbf{R}_o^n &= \mathbf{R}_o^{n-1} + \mathbf{D}_{n-1}^o (\mathbf{I} - \mathbf{R}_{n-1}^n \mathbf{R}_{n-1}^o)^{-1} \mathbf{R}_{n-1}^n \mathbf{D}_o^{n-1} \\
&= \mathbf{R}_o^{n-1} + \mathbf{D}_{n-1}^o \mathbf{R}_{n-1}^n (\mathbf{I} - \mathbf{R}_{n-1}^o \mathbf{R}_{n-1}^n)^{-1} \mathbf{D}_o^{n-1} \\
\mathbf{D}_o^n &= \mathbf{D}_{n-1}^n (\mathbf{I} - \mathbf{R}_{n-1}^o \mathbf{R}_{n-1}^n)^{-1} \mathbf{D}_o^{n-1} \\
\mathbf{D}_n^o &= \mathbf{D}_{n-1}^o (\mathbf{I} - \mathbf{R}_{n-1}^n \mathbf{R}_{n-1}^o)^{-1} \mathbf{D}_n^{n-1} \\
\mathbf{R}_n^o &= \mathbf{R}_n^{n-1} + \mathbf{D}_{n-1}^n (\mathbf{I} - \mathbf{R}_{n-1}^o \mathbf{R}_{n-1}^n)^{-1} \mathbf{R}_{n-1}^o \mathbf{D}_n^{n-1} \\
&= \mathbf{R}_n^{n-1} + \mathbf{D}_{n-1}^n \mathbf{R}_{n-1}^o (\mathbf{I} - \mathbf{R}_{n-1}^n \mathbf{R}_{n-1}^o)^{-1} \mathbf{D}_n^{n-1}.
\end{aligned} \tag{5.2}$$

Starting with $\mathbf{R}_o^o = \mathbf{0}$ and $\mathbf{D}_o^o = \mathbf{I}$ and applying s times the relations (5.2), one gets:

$$\begin{bmatrix} \mathbf{d}_o \\ \mathbf{u}_s \end{bmatrix} = \begin{bmatrix} \mathbf{R}_o^s & \mathbf{D}_s^o \\ \mathbf{D}_o^s & \mathbf{R}_s^o \end{bmatrix} \begin{bmatrix} \mathbf{u}_o \\ \mathbf{d}_s \end{bmatrix} \tag{5.3}$$

where $\mathbf{R}_o^s$ and $\mathbf{D}_o^s$ are respectively the reflection and the transmission matrices of the medium when the source is located in the medium I_1 while $\mathbf{R}_s^o$ and $\mathbf{D}_s^o$ are the corresponding matrices when the emission takes place in the medium S_s. Particularly, if the columnvectors of the matrix $\mathbf{Q}$ in the vacuum are put in the following order: upgoing parallel, upgoing perpendicular, downgoing parallel and downgoing perpendicular, one gets with the usual notations

$$\mathbf{R}_o^s = \begin{bmatrix} {}_{\parallel}\mathbf{R}_{\parallel} & {}_{\parallel}\mathbf{R}_{\perp} \\ {}_{\perp}\mathbf{R}_{\parallel} & {}_{\perp}\mathbf{R}_{\perp} \end{bmatrix}; \quad \mathbf{D}_o^s = \begin{bmatrix} {}_{\parallel}\mathbf{D}_{\parallel} & {}_{\parallel}\mathbf{D}_{\perp} \\ {}_{\perp}\mathbf{D}_{\parallel} & {}_{\perp}\mathbf{D}_{\perp} \end{bmatrix}. \tag{5.4}$$

6. Independent Solutions of the Differential Equation

If the application of the algorithm is concerned with the fourvector $\mathbf{f}$ (and of the fields) at the subdivision levels, one has previously (a) to store all the

submatrices $\mathbf{R}_n^o$ and $\mathbf{D}_o^n$, (b) to compute the submatrices corresponding to the integration from the top to the bottom of the medium; namely:

$$\begin{aligned}
\mathbf{R}_s^n &= \mathbf{R}_s^{n+1} + \mathbf{D}_{n+1}^s (\mathbf{I} - \mathbf{R}_{n+1}^n \mathbf{R}_{n+1}^s)^{-1} \mathbf{R}_{n+1}^n \mathbf{D}_s^{n+1} \\
&= \mathbf{R}_s^{n+1} + \mathbf{D}_{n+1}^s \mathbf{R}_{n+1}^n (\mathbf{I} - \mathbf{R}_{n+1}^s \mathbf{R}_{n+1}^n)^{-1} \mathbf{D}_s^{n+1} \\
\mathbf{D}_s^n &= \mathbf{D}_{n+1}^n (\mathbf{I} - \mathbf{R}_{n+1}^s \mathbf{R}_{n+1}^n)^{-1} \mathbf{D}_s^{n+1} \\
\mathbf{D}_n^s &= \mathbf{D}_{n+1}^s (\mathbf{I} - \mathbf{R}_{n+1}^n \mathbf{R}_{n+1}^s)^{-1} \mathbf{D}_n^{n+1} \\
\mathbf{R}_n^s &= \mathbf{R}_n^{n+1} + \mathbf{D}_{n+1}^n (\mathbf{I} - \mathbf{R}_{n+1}^s \mathbf{R}_{n+1}^n)^{-1} \mathbf{R}_{n+1}^s \mathbf{D}_n^{n+1} \\
&= \mathbf{R}_n^{n+1} + \mathbf{D}_{n+1}^n \mathbf{R}_{n+1}^s (\mathbf{I} - \mathbf{R}_{n+1}^n \mathbf{R}_{n+1}^s)^{-1} \mathbf{D}_n^{n+1}
\end{aligned} \tag{6.1}$$

starting with $\mathbf{R}_s^s = \mathbf{0}$ and $\mathbf{D}_s^s = \mathbf{I}$ and to store all the submatrices $\mathbf{R}_n^s$ and $\mathbf{D}_s^n$.

Then, the formation of the two pairs of solutions proceeds as follows:

(a) for the first pair, one admits the limiting condition

$$\mathbf{d}_s = \mathbf{0} \tag{6.2}$$

and then, one looks for solutions corresponding to a source situated under the level z_0. The integration then happens, not with initial conditions, but taking this limiting condition into account; in this way, the growing of unwanted solutions is avoided.

Under this condition, one has at the extreme levels:

$$\mathbf{d}_o = \mathbf{R}_o^s \mathbf{u}_o, \qquad \mathbf{u}_s = \mathbf{D}_o^s \mathbf{u}_o \tag{6.3}$$

and at each intermediate level z_n

$$\mathbf{d}_o = \mathbf{R}_o^n \mathbf{u}_o + \mathbf{D}_n^o \mathbf{d}_n, \qquad \mathbf{u}_n = \mathbf{D}_o^n \mathbf{u}_o + \mathbf{R}_n^o \mathbf{d}_n$$

so that

$$\begin{aligned}
\mathbf{d}_n &= (\mathbf{D}_n^o)^{-1} (\mathbf{d}_o - \mathbf{R}_o^n \mathbf{u}_o) = (\mathbf{D}_n^o)^{-1} (\mathbf{R}_o^s - \mathbf{R}_o^n) \mathbf{u}_o = \mathbf{B}_o^n \mathbf{u}_o \\
\mathbf{u}_n &= \mathbf{D}_o^n \mathbf{u}_o + \mathbf{R}_n^o \mathbf{B}_o^n \mathbf{u}_o = \mathbf{A}_o^n \mathbf{u}_o .
\end{aligned} \tag{6.4}$$

If one notes that owing to (4.4)

$$\mathbf{R}_o^s - \mathbf{R}_o^n = \mathbf{D}_n^o \mathbf{R}_n^s (\mathbf{I} - \mathbf{R}_n^o \mathbf{R}_n^s)^{-1} \mathbf{D}_o^n$$

one obtains for $\mathbf{B}_o^n$ the expression

$$\mathbf{B}_o^n = \mathbf{R}_n^s (\mathbf{I} - \mathbf{R}_n^o \mathbf{R}_n^s)^{-1} \mathbf{D}_o^n .$$

Then, one gets

$$\begin{aligned}
\mathbf{A}_o^n &= \mathbf{D}_o^n + \mathbf{R}_n^o \mathbf{B}_o^n = [\mathbf{I} + \mathbf{R}_n^o \mathbf{R}_n^s \ (\mathbf{I} - \mathbf{R}_n^o \mathbf{R}_n^s)^{-1}] \mathbf{D}_o^n \\
&= (\mathbf{I} - \mathbf{R}_n^o \mathbf{R}_n^s)^{-1} \mathbf{D}_o^n \ ^5
\end{aligned} \tag{6.5}$$

---

[5] According to the identity

$$\mathbf{I} + \mathbf{AB}(\mathbf{I} - \mathbf{AB})^{-1} \equiv (\mathbf{I} - \mathbf{AB})^{-1}$$

and

$$\mathbf{B}_o^n = \mathbf{R}_n^s \mathbf{A}_o^n \tag{6.6}$$

(b) for the second pair, one admits the limiting condition

$$\mathbf{u}_o = \mathbf{0} \tag{6.7}$$

corresponding to a source placed above the medium.

Then one has at the extreme levels:

$$\mathbf{d}_o = \mathbf{D}_s^o \mathbf{d}_s, \quad \mathbf{u}_s = \mathbf{R}_s^o \mathbf{d}_s \tag{6.8}$$

and at each intermediate level

$$\begin{aligned} \mathbf{d}_n &= \mathbf{A}_s^n \mathbf{d}_s = (\mathbf{I} - \mathbf{R}_n^s \mathbf{R}_n^o)^{-1} \mathbf{D}_s^n \mathbf{d}_s \\ \mathbf{u}_n &= \mathbf{B}_s^n \mathbf{d}_s = \mathbf{R}_n^o \mathbf{A}_s^n \mathbf{d}_s. \end{aligned} \tag{6.9}$$

The linearity of the differential equation allows the grouping of the two pairs in the form of:

$$\mathbf{c}(z_n) = \begin{bmatrix} \mathbf{u}_n \\ \mathbf{d}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_o^n & \mathbf{B}_s^n \\ \mathbf{B}_o^n & \mathbf{A}_s^n \end{bmatrix} \begin{bmatrix} \mathbf{u}_o \\ \mathbf{d}_s \end{bmatrix} \tag{6.10}$$

with, at the extreme levels,

$$\mathbf{c}(z_o) = \begin{bmatrix} \mathbf{u}_o \\ \mathbf{d}_o \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_o^s & \mathbf{D}_s^o \end{bmatrix} \begin{bmatrix} \mathbf{u}_o \\ \mathbf{d}_s \end{bmatrix}; \quad \mathbf{c}(z_s) = \begin{bmatrix} \mathbf{u}_s \\ \mathbf{d}_s \end{bmatrix} = \begin{bmatrix} \mathbf{D}_o^s & \mathbf{R}_s^o \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{u}_o \\ \mathbf{d}_s \end{bmatrix}.$$

Finally, the fourvectors $\mathbf{f}(z_n)$ are obtained from

$$\mathbf{f}(z_n) = \mathbf{Q}(z_n)\,\mathbf{c}(z_n) = \mathbf{Q}(z_n) \begin{bmatrix} \mathbf{A}_o^n & \mathbf{B}_s^n \\ \mathbf{B}_o^n & \mathbf{A}_s^n \end{bmatrix} \begin{bmatrix} \mathbf{u}_o \\ \mathbf{d}_s \end{bmatrix}. \tag{6.11}$$

7. Choice of the Media I_n and S_n. Interpolation

In the present algorithm, the media I_n and S_n are only introduced in order to define four eigenvectors corresponding, from a physical point of view, to a pair of ascending waves and to another pair of descending waves. These eigenvectors constitute a vectorial basis in terms of which the solution is decomposed at each intermediate level; these choice of these vectors (and of the media) is largely a matter of convenience; only physical reasons can restrict this choice. It has evidently no effect on the final results (6.11) of the integration; only the matrices of diffusion bear the mark of the adopted choice. In most algorithms (Inoue and Horowitz, 1966; Altman and Cory, 1969b; Nagano et al., 1975), the media I_n and S_n are defined using the values of the physical parameters at the levels z_{n-1} and z_n. Then, the solution is decomposed in terms of the characteristic waves at the level reached during the integration. These waves are determined starting from the eigenvalues of the matrix $\mathbf{T}(z_n)$ (in ionospheric propagation, they are

the roots of the Booker equation) and then forming the corresponding eigenvectors. This type of decomposition bears interest for the physical interpretation of the propagation conditions. However, this interest must not be overrated.

In practice, the numerical treatment gains in simplicity and speed when each of the media I_n and S_n is the same uniform medium. In the special case of electromagnetic waves, it is advisable to use the vacuum for which no problem of normalisation of the eigenvectors arises.

If the fields are decomposed with respect to components linearly polarized in the plane of incidence ($\|$ component) and perpendicular to it ($\perp$ component), one uses the matrix:

$$\mathbf{Q}=\begin{bmatrix} \cos\theta_o\cos\Phi_o & \sin\Phi_o & -\cos\theta_o\cos\Phi_o & \sin\Phi_o \\ \cos\theta_o\sin\Phi_o & -\cos\Phi_o & -\cos\theta_o\sin\Phi_o & -\cos\Phi_o \\ -\sin\Phi_o & \cos\theta_o\cos\Phi_o & \sin\Phi_o & -\cos\theta_o\cos\Phi_o \\ \cos\Phi_o & \cos\theta_o\sin\Phi_o & -\cos\Phi_o & -\cos\theta_o\sin\Phi_o \\ \|\uparrow & \perp\uparrow & \|\downarrow & \perp\downarrow \end{bmatrix} \tag{7.1}$$

and one obtains, with

$$\mathbf{u}_o=\begin{bmatrix}\mathbf{u}_{o\|}\\ \mathbf{u}_{o\perp}\end{bmatrix} \quad \text{and} \quad \mathbf{d}_o=\begin{bmatrix}\mathbf{d}_{o\|}\\ \mathbf{d}_{o\perp}\end{bmatrix} \tag{7.2}$$

the matrices

$$\mathbf{R}_o^s=\begin{bmatrix}{}_{\|}\mathbf{R}_{\|} & {}_{\|}\mathbf{R}_{\perp}\\ {}_{\perp}\mathbf{R}_{\|} & {}_{\perp}\mathbf{R}_{\perp}\end{bmatrix} \quad \text{and} \quad \mathbf{D}_o^s=\begin{bmatrix}{}_{\|}\mathbf{D}_{\|} & {}_{\|}\mathbf{D}_{\perp}\\ {}_{\perp}\mathbf{D}_{\|} & {}_{\perp}\mathbf{D}_{\perp}\end{bmatrix} \tag{7.3}$$

defining respectively the reflection and the transmission of the considered medium limited from both sides by the vacuum. For circular polarizations (r=right, l=left), one gets directly (Budden, 1961)

$${}_c\mathbf{R}_o^s=\begin{bmatrix}{}_r\mathbf{R}_r & {}_r\mathbf{R}_l\\ {}_l\mathbf{R}_r & {}_l\mathbf{R}_l\end{bmatrix}=\mathbf{U}^{-1}\,\mathbf{R}_o^s\,\mathbf{U}; \quad {}_c\mathbf{D}_o^s=\begin{bmatrix}{}_r\mathbf{D}_r & {}_l\mathbf{D}_r\\ {}_r\mathbf{D}_l & {}_l\mathbf{D}_l\end{bmatrix}=\mathbf{U}^{-1}\,\mathbf{D}_o^s\,\mathbf{U} \tag{7.4}$$

with

$$\mathbf{U}=2^{-1/2}\begin{bmatrix}1 & 1\\ i & -i\end{bmatrix}. \tag{7.5}$$

Having computed the diffusion matrix related to a medium S_n at the level z_n, it is not difficult to obtain the diffusion matrix related to an other medium S_n' at the same level. If $\mathbf{Q}'$ is the matrix with the eigenvectors of S_n' one forms the matrix and the partition

$$\mathbf{Q}\,\mathbf{Q}'^{-1}=\begin{bmatrix}\mathbf{V}_1 & \mathbf{V}_2\\ \mathbf{V}_3 & \mathbf{V}_4\end{bmatrix} \tag{7.6}$$

and one applies the relations

$$\mathbf{R}'^{n}_{o} = \mathbf{R}^{n}_{o} + \mathbf{D}^{o}_{n}\mathbf{V}_3\mathbf{D}'^{n}_{o}$$
$$\mathbf{D}'^{n}_{o} = (\mathbf{V}_1 - \mathbf{R}^{o}_{n}\mathbf{V}_3)^{-1}\mathbf{D}^{n}_{o}$$
$$\mathbf{D}'^{o}_{n} = \mathbf{D}^{o}_{n}(\mathbf{V}_3\mathbf{R}'^{o}_{n} + \mathbf{V}_4)$$
$$\mathbf{R}'^{o}_{n} = (\mathbf{V}_1 - \mathbf{R}^{o}_{n}\mathbf{V}_3)^{-1}(\mathbf{R}^{o}_{n}\mathbf{V}_4 - \mathbf{V}_2). \tag{7.7}$$

Finally, in regions where the fields vary very fast with height, it can be useful to compute the fields at closer levels in order to describe or to interpret the results of the integration. The easiest process for this interpolation consists in the computation of the matrizants related with the subslabs and to use the relation (2.1). No swamping has to be feared in this case.

Concluding Remarks

All the parts of the algorithm have been programmed in FORTRAN V and tested on the UNIVAC 1100/40 of the Institut Royal Météorologique, including the path distortion around the singularity of **T** which arises at great height where the collision frequency is very low.

There arose no numerical difficulties for the ionization profiles (sum of Chapman functions) used for the *D*- and *E*-regions; the only difficulties were met with the resolution of the Booker equation and the subsequent formation of the eigenvectors, especially (as is well known) as soon as the roots are not well separated.

Therefore, the systematic use of the vacuum as uniform medium in the case of E.M. waves brings advantages as regards speed, usefulness and also accuracy. Further, if one wants to appeal to characteristic waves at any level, it is sufficient to use (7.7); this will have no influence on the results of the present algorithm.

Finally, this algorithm is suitable for the numerical treatment of any differential equation like (1.8) provided **T** is of class τ_{2p}. Such a case arises, after a change of variable, in the theory of gravity waves within the atmosphere (Volland, 1969, p. 500).

References

Altman, C., Cory, H.: The simple thin-film optical method in electromagnetic wave propagation. Radio Sci. **4** (5), 449–457, 1969a

Altman, C., Cory, H.: The generalized thin-film optical method in electromagnetic wave propagation. Radio Sci. **4** (5), 459–470, 1969b

Altman, C., Cory, H., Fijalkow, E.: Coupling processes in the *D*- and *E*-regions at low and very low frequencies – I: Frequencies less than critical: a study of the daytime ionosphere. J. Atmos. Terr. Phys. **32** (8), 1457–1473, 1970

Altman, C., Fijalkow, E.: Coupling processes in the *D*- and *E*-regions at low and very low frequencies – II: Frequencies greater than critical: a study of the night-time ionosphere. J. Atmos. Terr. Phys. **32** (8), 1475–1488, 1970

Altman, C., Postan, A.: Coupling processes in the *D*- and *E*-regions at low and medium frequencies – III: The 'coupling echo' and the transition through 'critical coupling'. J. Atmos. Terr. Phys. **33** (2), 329–341, 1971

Berreman, D.W.: Optics in stratified and anisotropic media: 4×4 matrix formulation. J. Opt. Soc. Am. **62** (4), 502–510, 1972

Booker, H.G.: Oblique propagation of electromagnetic waves in a slowly-varying non-isotropic medium. Proc. R. Soc. London, Ser. A: **155**, 235–257, 1936

Born, M., Wolf, E.: Principles of Optics, 4th edn. pp. 51–70. Oxford: Pergamon Press 1970

Bossy, L.: La détermination précise du matrizant pour la résolution des équations de Maxwell dans une ionosphère stratifiée. Ann. Soc. Sci. Bruxelles **85** (1), 29–38, 1971

Bossy, L., Claes, H.: A numerical method for the integration of the equations governing the propagation of E.M. waves in a plane stratified magnetoionic medium, ELF-VLF Radio Wave Propagation, J.A. Holtet, ed.: pp. 111–116. Dordrecht: D. Reidel 1974

Budden, K.G.: Radio waves in the ionosphere. pp. 90–91. Cambridge: University Press 1961

Budden, K.G.: Full wave solutions for radio waves in a stratified magnetoionic medium. Alta Freq., 28 n° speziale, 167–179, 1969

Clemmow, P.C., Heading, J.: Coupled form of the differential equations governing radio propagation in the ionosphere. Proc. Cambridge Philos. **50**, 319–333, 1954

Gantmacher, F.R.: Théorie des matrices, vol. 2. Paris: Dunod 1966

Honig, E.P., den Engelsen, D.: Reflection from stratified anisotropic media: an alternative method. Opt. Acta **24** (1), 89–95, 1977

Inoue, Y., Horowitz, S.: Numerical solution of full-wave equation with mode coupling. Radio Sci. **1** (8), 957–970, 1966

Lacoume, J.L.: Etude de la propagation des ondes électromagnétiques dans un magnétoplasma stratifié. Ann. Géophys. **23** (4), 477–483, 1967

Nagano, I., Mambo, M., Hutatsuishi, G.: Numerical calculation of electromagnetic waves in an anisotropic multilayered medium. Radio Sci. **10** (6), 611–618, 1975

Rawer, K., Suchy, K.: Radio-observations of the ionosphere, Handbuch der Physik, XLIX/2, Geophysik III. Berlin, Heidelberg, New York: Springer 1967

Teitler, S., Henvis, B.W.: Refraction in stratified anisotropic media. J. Opt. Soc. Am. **60** (6), 830–834, 1970

Volland, H.: Die Streumatrix der Ionosphäre. Arch. Elektr. Übertragung **16** (7), 328–334, 1962a

Volland, H.: Kopplung und Reflexion von elektromagnetischen Wellen in der Ionosphäre. Arch. Elektr. Übertragung **16** (10), 515–524, 1962b

Volland, H.: Die Ausbreitung langer Wellen. Braunschweig: Vieweg 1968

Volland, H.: The upper atmosphere as a multiply refractive medium for neutral air motions. J. Atmos. Terr. Phys. **31** (4), 491–514, 1969

Received October 31, 1978; Accepted May 2, 1979

J. Geophys. 46, 15–21, 1979

Journal of Geophysics

Radio Pulse Dispersion in the Ionosphere*

A.K. Paul

Space Environment Laboratory, NOAA Environmental Research Laboratories, Boulder, Colorado 80303, USA

Abstract. A computer simulation of the propagation of a radio pulse shows that the best estimate of the virtual heights can be obtained by the phase change with frequency (principle of stationary phase). This is especially important for those frequencies where the penetrated part of the ionosphere is highly dispersive indicated by a rapid change of the virtual heights with frequency.

Key words: Ionosphere – Virtual height pulse dispersion.

Introduction

The travel time of a radio pulse reflected from the ionosphere can be measured in two different ways. For the classical approach a certain characteristic of the pulse shape is selected and the time difference between the appearance of this characteristic in the transmitted pulse and the reflected echo is measured. The other approach is based on the principle of stationary phase and the travel time can be derived from the phase change over a small frequency increment. Using the Dynasonde Wright (1977) compared the two methods. Systematic differences were found between the two measurements in those parts of the ionogram where the virtual height varies rapidly with frequency. In the following we show some further examples of those discrepancies and also the results of a computer simulation leading to the conclusion that the stationary phase principle gives a much better estimate of the virtual height than any characteristic of the pulse shape.

Dynasonde Data

In one standard mode of operation the Dynasonde transmits four pulses with a radio frequency f and the echoes are received in sequence at four different

* Dedicated to Professor Dr. K. Rawer on the occation of his 65th birthday

0340-062X/79/0046/0015/$01.40

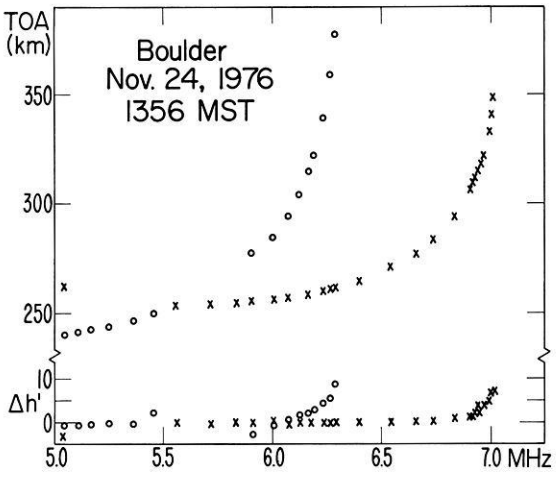


Fig. 1. Virtual heights and discrepancies between TOA (derived from the leading edge of an echo) and DPH (derived from the phase change with frequency), *Boulder, November 24, 1976* (*circles:* ordinary wave; *crosses:* extraordinary wave)

receiving antenna, which are linearly polarized and located at the corners of a square. This sequence is then repeated at a slightly higher frequency $f+\Delta f$, where Δf is always 8 kHz. For each echo the time of arrival TOA is counted to the instant of maximum second derivative of the leading edge of the pulse and data are also obtained on echo amplitude and phase. The latter two quantities are measured at a later time, approximately in the center of the echo (Wright and Pitteway, 1977). The phase resolution is 4° C and correspondingly a phase change of one unit over 8 kHz gives a resolution of 0.208 km for the virtual height (DPH) derived from the stationary phase principle. The TOA virtual height has practically the same resolution of 0.200 km, but is actually the average of eight values measured at the four antenna for the two frequencies. The TOA values are corrected for amplitude dependent delays in the system.

Four phase differences, one for each antenna, are obtained over the 8 kHz frequency interval. Figure 1 shows the high frequency portion of an ionogram where ordinary and extraordinary traces approach their critical frequencies. In the upper part the TOA values are plotted and in the lower portion the difference

$$\Delta h' = \text{DPH-TOA}.$$

We see that in the flat portions of each trace this difference is very small and increases rapidly when the slope of the trace becomes very large. Figure 2 shows that this effect is not restricted to steep positive slopes, it also appears when the virtual heights are rapidly decreasing with frequency. In this ionogram, a well pronounced $F1$ region maximum was present and only the $F2$ region portion of the ordinary component is shown. In both cases the differences between TOA and DPH amount to 10 km or more. Those differences can be significant in the calculation of electron density profiles, especially in the extrapolation to obtain the parameters of the maximum of a layer.

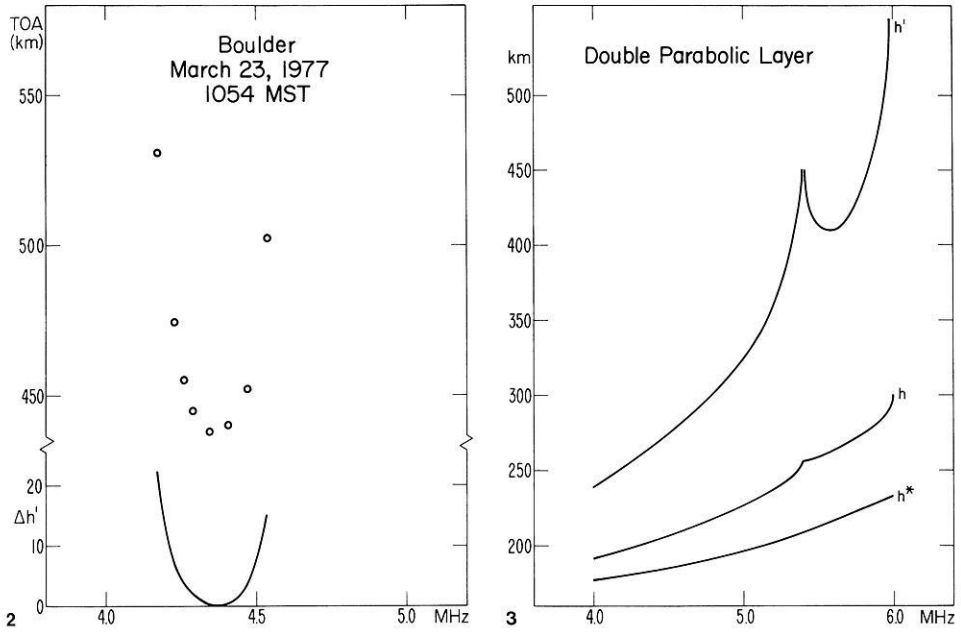


Fig. 2. Virtual heights and discrepancies between TOA and DPH (ordinary wave), *Boulder, March 23, 1977*

Fig. 3. Virtual heights (h'), real heights (h) and phase height (h^*) of a *double parabolic layer*

Computer Simulation

The problem of the propagation of electromagnetic waves through a dispersive medium was first treated by Sommerfeld (1914) and Brillouin (1914). Several authors (e.g., Rydbeck, 1942; Budden, 1961; Wait, 1965) studied specific problems of the pulse propagation in the ionosphere using in general the following procedure: A rectangular radio pulse is decomposed by the Fourier transform into its spectral components. The propagation of the pulse through the ionosphere adds a frequency dependent phase term for the argument of the individual waveletts. The phase term is then expanded into a Taylor series and higher order terms of this series are neglected. The integral describing the echo returned from the ionosphere can then be expressed by well known functions. Those studies concentrated mainly on the propagation velocity and the deformation of the pulse shape, while the effect on the phase of the echo was neglected.

In order to resolve the experimental discrepancy between the virtual heights derived from the echo shape and those derived from the change of phase with frequency, the propagation of a radio pulse in the ionosphere was simulated numerically. The results were amplitude and phase as a function of time. In our calculations the radio pulse has a cosine shape and is given by

$$f(t)=\begin{cases}0 & -\infty \leqq t \leqq -T \\ \frac{1}{2}\left(1+\cos \pi \frac{t}{T}\right) \cos 2\pi f_0 t & -T \leqq t \leqq T \\ 0 & T \leqq t \leqq \infty\end{cases} \tag{1}$$

where T is the nominal pulse width (measured at half maximum amplitude level) and f_0 is the carrier frequency. A cosine pulse appears to be more realistic than a rectangular pulse since the transmitter network including antenna will always smooth the edges of an originally rectangular pulse. It also has the advantage of a faster decrease of the amplitude density of the spectral components with increasing distance of the frequency from the carrier frequency and we avoid the Gibb's phenomena which causes an artificial additional distortion of the echo shape as apparent in the papers by Rydbeck (1942) and Budden (1961).

The amplitude density of the pulse as obtained by the Fourier integral is then

$$a(f)=\frac{1}{4\pi}\left(\frac{\sin 2\pi(f-f_0)\,T}{(f-f_0)(1-4(f-f_0)^2\,T^2)}+\frac{\sin 2\pi(f+f_0)\,T}{(f+f_0)(1-4(f+f_0)^2\,T^2)}\right). \tag{2}$$

The echo reflected from the ionosphere is then given by

$$g(t)=\int_{-\infty}^{+\infty} a(f)\exp[i(2\pi f t-\phi(f))]\,df \tag{3}$$

where the phase $\phi(f)$ is proportional to the phase height $h^*(f)$

$$\phi=2\pi p=\frac{4\pi}{C} f h^*=\frac{4\pi}{C} f \int_0^{h_r} \mu\, dz \tag{4}$$

(μ refractive index, h_r reflection height, p phase in cycles).
Since μ depends only of f^2 we see that

$$\phi(-f)=-\phi(f) \tag{5}$$

and according to (2) that

$$a(-f)=a(f).$$

Therefore we can write instead of (3)

$$\begin{aligned} g(t) &= 2\int_0^{\infty} a(f)\cos 2\pi[ft-p(f)]\,df \\ &= 2R\left(\int_0^{\infty} a(f)\exp[i\,2\pi(ft-p(f))]\,df\right). \end{aligned} \tag{6}$$

Assuming that p can be expanded into a Taylor series we write

$$p = p(f_0) + \frac{dp}{df}(f - f_0) + \delta p \tag{7}$$

where δp includes all the higher order terms. Since by definition

$$\frac{dp}{df} = \tau$$

where τ is the total travel time of the pulse we can write instead of (6)

$$g(t) = 2R\left(\exp[i2\pi(f_0\tau - p(f_0))]\int_0^\infty a(f)\exp[i2\pi(f(t-\tau) - \delta p)]\,df\right). \tag{8}$$

If δp can be neglected the integral in (8) is then the inverse Fourier transform and gives the original pulse at the later time $t = \tau$. The factor in front of the integral in (8) is simply a phase shift of the carrier frequency corresponding to the difference between virtual height and phase height.

Instead of a series expansion we use the exact expression of p for a model electron density profile neglecting the earth's magnetic field. The model consists of two half parabolas with different critical frequencies and is shown in Fig. 3 together with the virtual and phase heights (Paul, 1967). For the numerical evaluation of (6) we used the following formulation

$$g(t) = 2R\left(\exp(2\pi i(f_0 t - p_0))\int_0^\infty a(f)\exp[2\pi i(\Delta f t - \Delta p)]\,df\right) \tag{9}$$

with $\Delta f = f - f_0$ and $\Delta p = p(f) - p(f_0)$.

The integral in (9) gives both the amplitude of the echo and its phase deviation from the nominal value p_0 as a function of time while the high frequency oscillation of the carrier frequency is contained in the factor in front of the integral.

The integral in (9) was evaluated numerically for several radio frequencies f_0. Some of the results are shown in Fig. 4. In the upper part of this figure a portion of the virtual height curve of Fig. 3 is shown, in the lower part the echo shape for three frequencies (indicated in the virtual height curve by the vertical arrows). The amplitudes of the echos are shown on a relative virtual height scale with zero in the nominal center of the pulse. The vertical dashed lines indicate the half-amplitude width of the undistorted pulse. The pulse form at the minimum of the virtual height trace ($f = 5.55$ MHz) is practically undistorted. At the other two frequencies, at $f = 5.44$ where the virtual height is rapidly decreasing with frequency and at $f = 5.95$ where the virtual height is rapidly increasing the echo shapes are flattened out, the amplitude is lower and the pulse width wider. A rough estimate of the product of the square of the maximum amplitude times the pulse width at half amplitude level shows that this quantity is equal for all three pulses shown, which means that the energy is conserved as expected. For each of the frequencies shown in Fig. 4 the

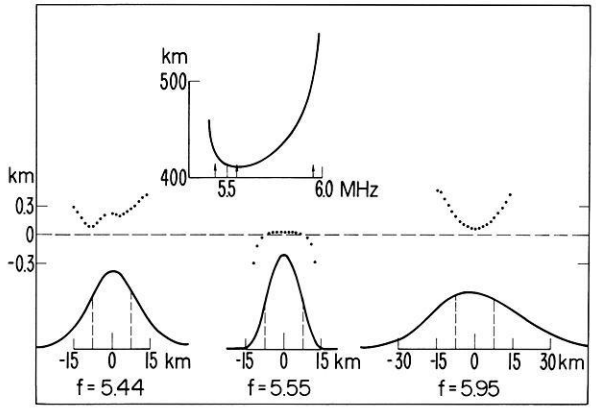


Fig. 4. Dispersion effect on phase change with frequency (*middle, dotted curves*) and on amplitude (*bottom*) for three frequencies reflected from the double parabolic model. The high frequency portion of the virtual heights of the model is shown in the *upper part* of the figure where the three frequencies are indicated by *arrows*

calculation was repeated for a second frequency 8 kHz higher simulating the Dynasonde mode of operation. At a given frequency f_0 the phase deviation from the nominal value $\phi(f_0)$ as defined in (4) can be significant. However, the change of this deviation with frequency is rather small, and therefore causes only small errors in the virtual heights, if those are derived from the phase change with frequency. Those errors are shown in the middle of Fig. 4 by the dotted curves, if the phase change over a 8 kHz interval is used. The horizontal scale is the same as for the echo shapes in the lower part of this figure. In the three cases shown the error of the virtual height (DPH) caused by the dispersion effect is less than 200 m.

More details are presented in Table 1, where the dispersion errors are listed for different methods of measuring the virtual height of the echo. The first line shows the average error over the nominal pulse width around the center of the echo obtained by the phase change over the 8 kHz frequency interval (see center line of Fig. 4). The following three lines give the errors, if a characteristic of the echo shape is used, first the location of the maximum, then the location of the steepest slope and in the last line the location of the point where the amplitude reaches half the value of its maximum. Those values were obtained by numerical interpolation and/or differentiation.

The phase change with frequency causes by far the smallest error. Next best results are obtained, if the time of maximum of the amplitude is used, but at least the error for $f = 5.95$ is not negligible anymore. Even larger errors occur if the steepest slope or the half amplitude level are used especially close to the critical frequency ($f = 6.0$ MHz). The discrepancies in the last line are comparable with the experimental errors found from the Dynasonde data (Figs. 1 and 2).

The results of the computer simulation indicate that the best estimates of the virtual heights are obtained if the principle of the stationary phase is applied. Any characteristic of the leading edge of the echo may lead to very large errors

Table 1. Virtual height error (calculated–theoretical height)

Frequency	5.44 MHz	5.55 MHz	5.95 MHz
Phase	0.196 km	0.022 km	0.118 km
Max. amplitude	−0.629 km	0.135 km	− 1.253 km
Steepest slope	−0.223 km	0.181 km	−10.982 km
Half amplitude	−3.089 km	0.039 km	−12.326 km

near the critical frequency. For profile calculations (e.g., Howe and McKinnis, 1967) this will in turn cause erroneous values for peak parameters.

The use of the maximum of the amplitude as a characteristic would reduce those errors significantly, but may not be practical.

It also should be pointed out that according to the computer simulation the amplitude decreases rapidly with increasing dispersion, an effect which has to be corrected for in deviative absorption measurements.

References

Brillouin, I.: Über die Fortpflanzung des Lichtes in dispergierenden Medien 2. Ann. Phys. **4**, 44, pp. 203–240, 1914

Budden, K.G.: Radio Waves in the Ionosphere. Cambridge: University Press 1961

Howe, H.H., McKinnis, D.E.: Ionospheric electron-density profiles with continuous gradients and underlying ionization corrections. II. Formulation for a digital computer. Radio Sci. **2**, No. 10, pp. 1135–1158, 1967

Paul, A.H.: Ionosphere electron density profiles with contineous gradients and underlying ionisation corrections. 1. The mathematical–physical problem of true height determination from ionograms. Radio Sci. **2**, No. 10, 1127–1133, 1967

Rydbeck, O.E.H.: A theoretical survey of the possibility of determining the distribution of the free electrons in the upper atmosphere, Chalmers Tek. 3, 1942

Sommerfeld, A.: Über die Fortpflanzung des Lichtes in dispergierenden Medien 1. Ann. Phys. **4**, 44, pp. 177–202, 1914

Wait, J.R.: Propatation of pulses in dispersive media. Radio Sci. **69D**, No. 11, 1387–1401, 1965

Wright, J.W.: Development of systems for remote sensing of ionospheric structures and dynamics: The dynasonde data acquisition and dynamic display system, NOAA Special Report, Space Environment Laboratory, NOAA Environmental Research Laboratories, Boulder, Colorado 80303, 1977

Received October 31, 1978; Revised Version January 23, 1979; Accepted March 29, 1979

J. Geophys. 46, 23–33, 1979

Journal of Geophysics

Ionospheric Absorption and Profiles of Electron Density and Loss-Rate in the Lower Ionosphere*

K.M. Kotadia and A. Gupta

Physics Department, Gujarat University, Ahmedabad 380009, India

Abstract. A1-absorption measurements on 1.8, 2.2 and 2.5 MHz frequencies are being made at Gujarat University, Ahmedabad (23°N, 72.6°E; mag. dip 34°N) since April, 1972. Partial reflection measurements were started in 1974 by *Physical Research Laboratory* (PRL) and the Ionosonde has been in operation since 1952 just nearby. The rocket experiment facilities were established in India at the equatorial station Thumba in 1963. In this paper, we have combined the results of all the above ground-based and rocket-borne experiments to determine the electron density and effective loss-rate profiles in the height range 65–110 km for quiet-sun low solar activity period. Cira (1972) Atmospheric model is used for concentrations of various gases, pressure and temperature. Profiles of nitric oxide and metastable molecular oxygen used in our computations are those due to Meira (1971) and Huffman et al. (1971).

It is shown that the *N-h* profile so obtained indicates a stratification intermediate between the *D*-layer maximum and the *E*-layer, and this is clearer at smaller *solar zenith angle* (SZA) times. This stratification seems to have a maximum electron density around 10^4 electrons/cm^3 at a height of about 95 km. The effective loss-rate is found to be greater at heights below 80 km and smaller at heights above 90 km than the dissociative recombination coefficient to the extent of an order of magnitude depending on the time of the day. The results are discussed in context to the negative ions, heavy metallic ions and the gas composition.

Key words: Ionospheric absorption – Electron density – Loss rate – Lower ionosphere.

1. Introduction

The ionospheric absorption of a radio wave at a given frequency is a function of the refractive index, electron density and the collision frequency in the medium

* Dedicated to Professor Dr. K. Rawer on the occasion of his 65th birthday

0340-062X/79/0046/0023/02.20

through which the wave travels. The refractive index at a particular height is itself determined by the electron density N, electron collision frequency and the effective frequency ($f \pm f_H$). The collision frequency depends on the neutral particle concentration, electron and ion densities and electron temperature. The collision frequency that is important in the lower ionosphere is that of electrons with neutral molecules.

Here, the ionospheric absorption data are used to fill the gap suitably in the electron density profile obtained by other methods such as rocket, partial reflections, and the ionosonde measurements. At Ahmedabad, we have at the Gujarat University an INDO-GDR collaboration A1-absorption project in operation since 1972. We have also a joint cooperative project with Physical Research Laboratory (PRL) on partial reflections since 1974 and the ionosonde is being run by the P.R.L. since 1952. Quite a large number of rocket flights carrying different kinds of payloads have been made since the establishment in 1963 of the rocket launching centre at the equatorial station, namely Thumba (8°33′N, 76°56′E; $I = 0.8°$S) but these were scattered instances as regards days, solar zenith angle and solar activity.

For estimation of effective loss-rate, we need to know the total electron-ion pair production rate which requires the knowledge of concentrations of various gas species and the fluxes of various ionizing radiations. In particular, here we have produced average results for the quiet-sun conditions during minimum solar activity ($F_{10.7} < 80$ U; $1\ \mathrm{U} = 10^{-22}\ \mathrm{W\ m^{-2}\ Hz^{-1}}$) at solar zenith angle (SZA) 30°–65° in the months of January, February, and March, 1975.

2. Atmospheric Model

The first atmospheric model giving pressure, density, mean molecular mass, temperature etc. was proposed in 1962 (U.S. standard Atmosphere 1962). Since then, many new observations have been made using ground-based, balloon and spacecraft techniques and improved computer modelling methods. The latest model used in our computations is the Cira, 1972 atmospheric model which is essentially based on the works of Groves (1971) and Jacchia (1971). In the calculation of ion-production rate from 60 km to 110 km altitudes, we have to consider the atmospheric model at height upto 160 km since the principal ionizing radiations responsible in the lower ionosphere begin to be attenuated on penetrating to heights from the level of 160 km downwards. The major constituents of the gas being ionized are N_2, O_2, and O and the minor constituents are the nitric oxide, NO, and the metastable oxygen, $O_2(1\Delta g)$. The latter two, though minor, are the main gases contributing to the middle D-region ionization. The models for N_2, O_2, and O are adopted from the Cira (1972), that of NO is due to Meira (1971) and the one for $O_2(1\Delta g)$ is due to Huffman et al. (1971). The profile of [NO] is believed to be nearly the same at heights above 70 km at different SZA's. Meira's model is still considered reasonably good for calculation of ion-production rate.

3. Collision Frequency

Following relations are used to calculate the collision frequency at different heights

$$\nu_{en} = 6.5 \times 10^5 \, p \quad \mathrm{s}^{-1}$$

Piggott and Thrane (1966). According to Banks (1966),

$$\nu_{ei} = 3.6 \times \mathrm{N} \, \mathrm{Te}^{-1.5} \ln(2 \times 10^4 \, \mathrm{Te}^{1.5}/\mathrm{N}^{0.5})$$

and

$$\nu_{eo} = [0] \; (1.88 \times 10^{-10} \, \mathrm{Te}^{0.5}).$$

In the above relations, p is pressure in Pa (or nt.m^{-2}), N is electron density in m^{-3}, Te is electron temperature in °K and [0] atomic oxygen concentration in m^{-3}. ν_{ei} and ν_{eo} are relatively unimportant in the lower ionosphere. There is some diurnal variation in pressure ranging from 0.1 to 0.2 Pa at 95 km with maxima at about 1400 and 0200 h and minima at 0800 and 2000 h but it reduces to very small variation at 65 km. Also at low latitudes, there is a seasonal variation in pressure with a minimum value 0.075 Pa in January and maximum value 0.105 Pa in September at 95 km altitude, with another subsidiary maximum in April. These variations in pressure produce changes accordingly in the collision frequency. The contribution to absorption of radio waves due to ionization and collision frequency below 70 km is very small.

4. Ion-Production Rate

The photo-ionization rate of a molecular or atomic gas species due to the absorption of ionizing radiations entering into the atmosphere is calculated by using the formula

$$q(z, \chi) = \phi(\lambda_j, z, \chi) \sum_j \gamma_j(\lambda) \, \sigma_{aj}(\lambda) \, n_j(z)$$

where j denotes j^{th} constituent of the gas, σ_{aj}, γ_j and n_j are respectively the absorption cross-section, ionization efficiency and number density of the j^{th} gas constituent, λ_j is the largest wavelength of the radiation capable of ionizing that constituent. $\phi(\lambda_j, z, \chi)$ is the total radiation flux of wavelength λ_j at solar zenith angle χ that is available after penetration to the height z. This flux at wavelength λ is given by the expression

$$\phi_\lambda(z, \chi) = \phi_{o\lambda} \exp\left[-\sum_j \sigma_{aj} \int_z^\infty n_j(l) \, dl\right]$$

in which $\phi_{o\lambda}$ is the flux of radiation of wavelength λ at the top of the atmosphere. The integral in the bracket gives the total number of particles

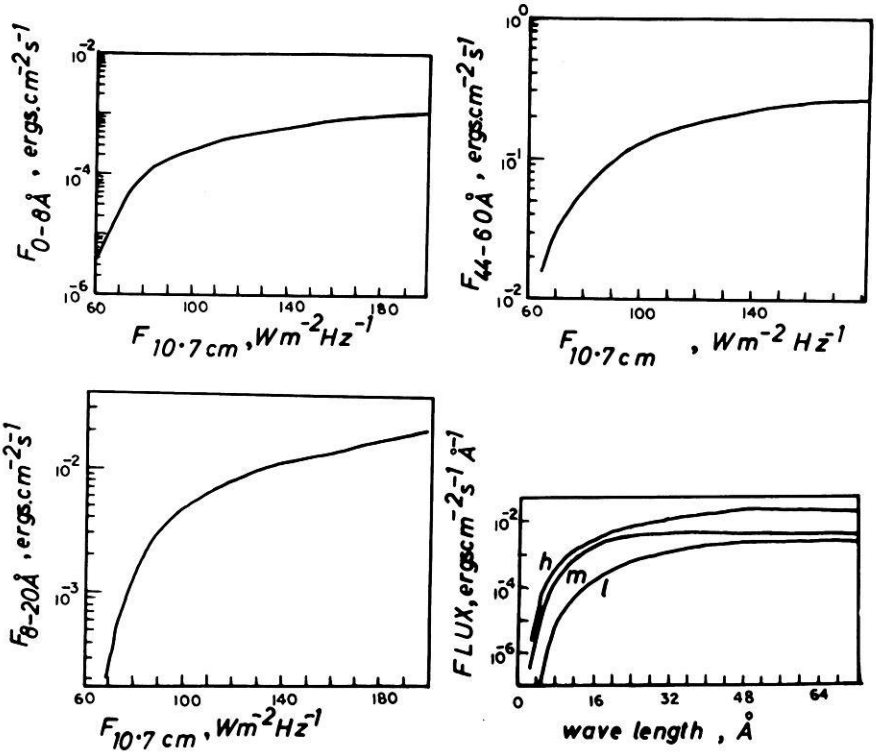


Fig. 1. Solar X-ray fluxes in 0–8 Å, 8–20 Å, and 44–60 Å bands against 10.7 cm solar radio flux, and X-ray spectrum for high, medium and low solar activity in the wavelength range 0 to 72 Å. $F_{10\cdot7}$ is in units of 10^{-22}

$S(z, x)$ of j^{th} gas constituent of spherically symmetric atmosphere in a unit column along the direction of the sun-rays from the level z of interest towards the sun (here upto 50 km radially above height z since the radiations concerned are absorbed in that region). This was calculated for different SZA values.

5. Ionizing Radiations and Cross-Sections

The total ion-production rate for all gases due to different radiations passing through them is obtained by summing up the q-rates for each gas produced by the radiations of wavelength ranging from 1 Å–1215·7 Å. The fluxes of these ionizing radiations, the absorption cross-sections and ionization-efficiency of a particular gas due to a particular radiation are now available in the richly published literature as well as Institutional Reports (Hinteregger et al., 1965; Hinteregger, 1970; Swider, 1969; Manson, 1972; Hall et al., 1965; Vidal-Madjar et al., 1973; Heroux and Higgins, 1977; Ivanov-Kholodnyi and Firsov, 1974; Horan, 1970; Loidl and Schwentek, 1977; Schmidtke, 1976). An average picture of the variation of flux in the X-ray spectrum with solar activity is shown in Fig. 1. The flux of the radiation at 1–3 Å remains below the measurable level for

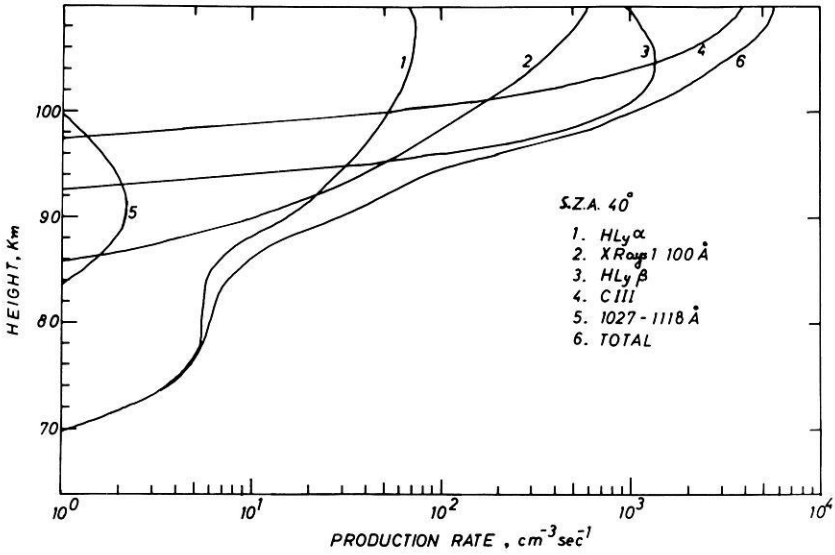


Fig. 2. Ion-production rate of various gas constituents in the height range 60–110 km due to different radiations for solar zenith angle 40°

most of the solar cycle period, and the ionization produced by such radiation at heights below 70 km does not contribute significantly to the radio wave absorption. A table of a few representative values of these fluxes was given in our earlier paper (Chhipa and Kotadia, 1978). The ion-production rates including the effects of background galactic cosmic rays and scattered geocoronal Lyman-alpha and Lyman-beta radiations were calculated for different solar zenith angles and here only one of them is shown in Fig. 2 for SZA 40°.

6. *N-h* Profiles

In-situ measurements of electron density have been made by various methods of rocket-probing. However, these have been under different conditions at different places. They seem to be so scattered in time and place that the material is insufficient for deducing a picture of general behaviour of the ionosphere, and particularly of the lower ionosphere where the situation is complicated by collisions, ion kinetics, uncertain loss processes and turbulent wind motions. However, the regular ground-based experiments such as ionosonde, absorption, partial reflections, wave-interaction, incoherent back-scatter and field-strength measurements at *vlf*, *lf*, *mf*, and *hf* can provide good amount of material to study the day-to-day variations and regular features of the ionosphere. As such, these experiments would continue to be powerful supporting tools in the investigations of the middle and the upper atmosphere. Here we have tried to combine the results obtained by ionosonde, A1-absorption and partial reflections for the height range 74–110 km, and rocket data at lower heights where

the ionospheric conditions are believed not to undergo significant changes with time and solar activity. *N-h* profiles available from several rocket flights in the height range 65–74 km are adjusted suitably for the low solar activity such that their average becomes continuous with the mean *N-h* profile obtained from partial reflections measurements. This brings the profile upto 88 km altitude. From the ionosonde *foE* and *hmE* data ($hmE = 105$ km around noontime), the *N-h* profile is drawn from *E*-layer maximum down to a height one-half the semithickness below the *hmE* level assuming parabolic distribution of electron density. *hmE* is supposed to vary according to the relation

$$hmE = 105 + H \ln(\sec \chi)$$

where the scale height H may change from 8 km at noon to 15 km near the sunrise and sunset times. The gap in the *N-h* profile in the height range 88 km to the lower half-thickness region of the *E*-layer has to be filled in such a way that the computed absorption and the virtual height of reflection at three radio frequencies agree reasonably with the observed values of absorption and virtual height. This will lead to the profile construction upto the height at which the electron density corresponds to the highest frequency, 2.5 MHz, used in our A1-absorption measurements.

In the computation of absorption coefficient, use is made of the Sen-Wyller generalised magnetoionic theory without recourse to QL or QT approximation. The computer program also takes care of converting the true height into virtual height. The lower part of the ionosphere under question is divided into slabs of thickness 1 km and absorption in each such slab is found. The slab-thickness is reduced as we go higher up, finally taking slabs of thickness 10 m near the level of reflection. The total absorption is then found by adding up absorption calculated for all such slabs and then multiplying the sum by 2 for up and down journey of the wave. The lowest frequency in our case is 1.8 MHz. Assuming on the basis of intuitive thinking, a kind of exponential profile passing through the point of reflection at 1.8 MHz and the points at ± 1.5 km, the upper point having greater electron density then at the lower point, absorption *Ldb* and h' are computed and compared with the observed values. If they do not agree, the points above and below the level of reflection are shifted upwards or downwards in height or laterally to lower or higher values of N (which amounts to changing the exponential shape of the profile) so that ultimately after a number of calculations and recalculations, the calculated and measured L_{db} and h' at the stated frequency agree well within experimental error. The process is then repeated in a similar way for the other two frequencies 2.2 and 2.5 MHz. The profile finally arrived at up to the reflection level at 2.5 MHz is then smoothly joined up with the *E*-layer profile obtained from the ionosonde data. Figure 3 gives the average *N-h* profiles thus completed for the month of March, 1975 at SZA 30°, 42°, and 65° C. It may be seen that the thus-found *N-h* profiles show the *D*-layer maximum electron density around 10^3 to 2×10^3 electrons/cm^3 at 82 to 85 km depending on the solar zenith angle. They also indicate another maximum at about 95 km with N around 10^4 electrons/cm^3. The cusp of such a maximum is clearer at lower solar zenith angles, but it is almost absent at high

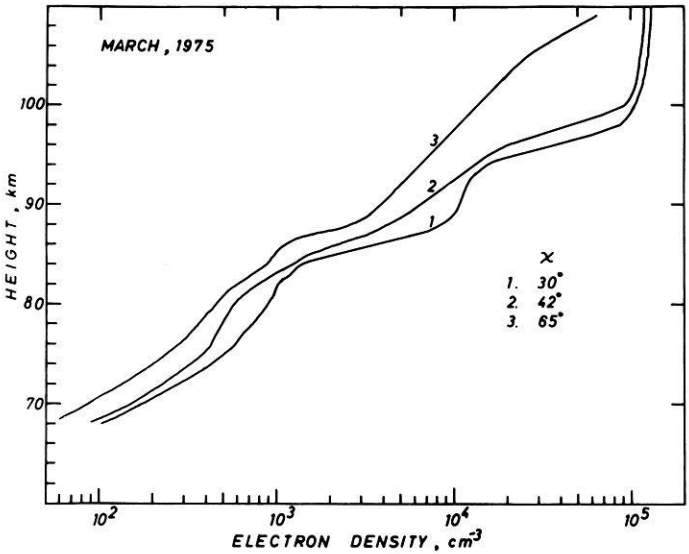


Fig. 3. Electron density distribution with height at SZA 30°, 42°, and 65° in March, 1975 under quiet-sun low solar activity conditions

values of SZA in winter months. Figures 4 and 5 give the N-h profiles for January and February, 1975 and the features of the stratification intermediate between the D-layer maximum and the bottom of the E-layer as described above are amply clear in these profiles also. Such intermediate stratifications were also reported by Bremer and Singer (1977) from their studies on A3 *lf* and *mf* absorption measurements. Mechtly and Bowhill (1972) had earlier noticed mild inflection at around 95 km in their rocket-measured N-h profiles during quiet-sun periods.

7. Effective Loss-Rate of Ionization

Having got the N-h and Q-h profiles, we now proceed to determine the $\alpha_{\text{eff}}-h$ profiles using the relation $\alpha_{\text{eff}}=Q/N^2$ where α_{eff} is the effective loss-rate or recombination coefficient. In the lower ionosphere, this is more accurately expressed as $\alpha_{\text{eff}}=(1+\lambda)\ (\alpha_d+\lambda\alpha_i)$ where $\lambda=N^-/N$, α_d is dissociative recombination coefficient, α_i is ion-ion neutralization rate and N^- is negative ion concentration. The ratio Q/N^2 is calculated for different heights and the results are shown in Fig. 6. It also gives mean α_{eff} for SZA 30°–50°. The effective α_d, i.e., total of electron recombination rates with all the molecular positive ions is based on theoretical models (not shown in the figure) rapidly increases at heights below 90 km and attains a constant value below 80 km as well as in the 90–95 km height region (Taubenheim et al., 1975). The transition region within the 80–90 km region may move up with increase in solar zenith angle. It is interesting to note that α_{eff} also shows a transition region similar to that for α_d.

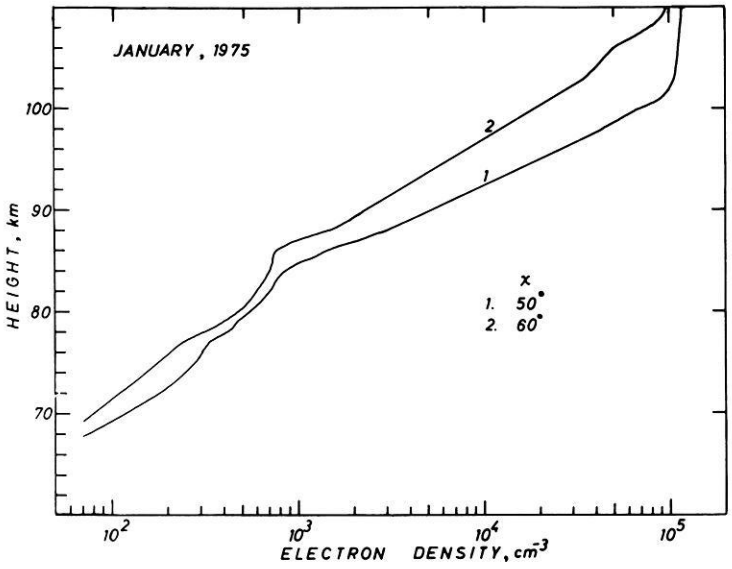


Fig. 4. Same as in Figs. 3, but at SZA 50° and 60° in January, 1975

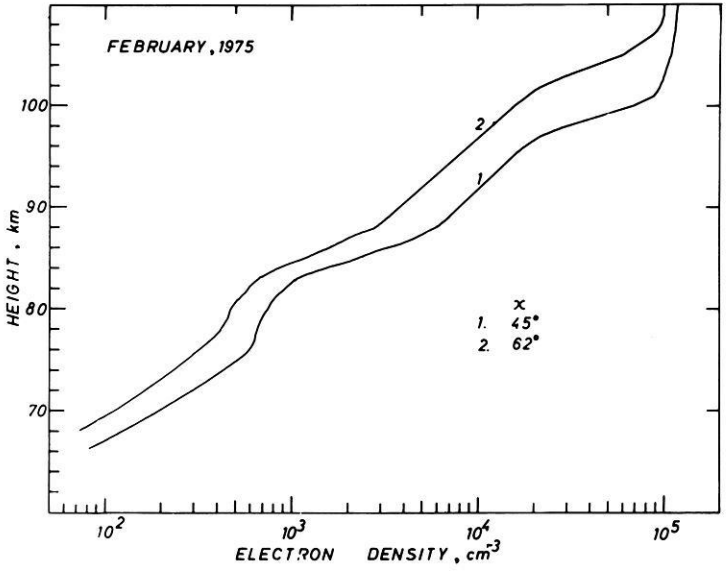


Fig. 5. Same as in Fig. 3, but at SZA 45° and 62° in February, 1975

However, it is found to be lower than α_d at heights above 90 km varying widely to the extent of an order of magnitude with SZA. In contrast, α_{eff} is found to be larger than α_d at heights below 80 km, also varying widely with SZA. Thus it seems that the recombination rates at heights above and below the transition region are not simple or similar processes, such as that of dissociative recom-

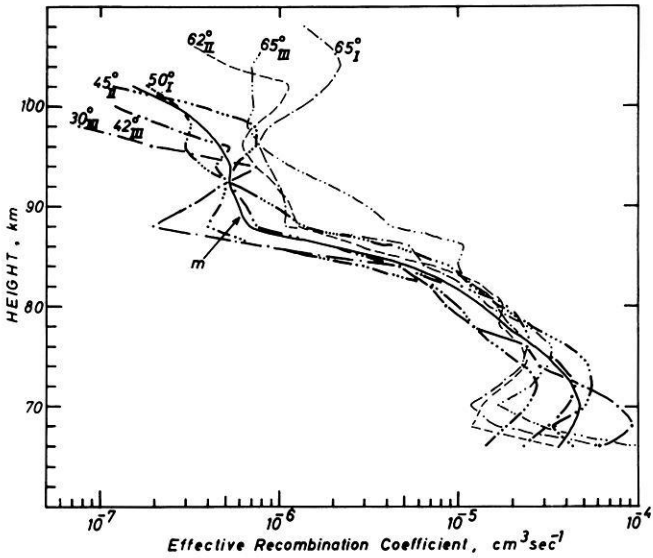


Fig. 6. Effective loss-rate calculated from the profiles of ion production rate and electron density for different solar zenith angles. Suffixes I, II, and III stand for respective months. The *full line curve* marked *m* gives mean α_{eff} for SZA 30°–50°

bination of molecular positive ions. The processes probably involve complex chemistry of cluster ions, negative ions and heavy metallic ions.

8. Conclusions and Discussion

The height-distribution profiles of electron density (N-h) and effective loss-rate ($\alpha_{eff} - h$) are constructed for the height range 65–110 km combining the results of rocket-borne N-h measurements and ground-based partial reflections, A1-absorption and ionosonde measurements for quiet-sun low solar activity conditions at different solar zenith angles in the months of January, February, and March, 1975. It is shown that all the N-h profiles so obtained point to a stratification intermediate between the D-layer maximum and the bottom of the E-layer at about 95 km with Nm around 10^4 electrons/cm^3. Such intermediate maxima had also been observed in some of the rocket flights.

The effective loss-rate is found to be smaller at heights 90–100 km and larger at heights 70–80 km than the dissociative recombination coefficient, varying to the extent of an order of magnitude with change in solar zenith angle and the season.

The low values of α_{eff} at heights above 90 km may be attributed to the presence of some metallic ions which reduce on the average the rate of recombination with electrons. There may be some uncertainty in the radiation fluxes and month-to-month changes in the atmospheric model which may result in low values of ion-production rate and hence give low values of Q/N^2. There is

now sufficient evidence that the ionospheric absorption reduces by about 30% on several occasions when a strong low-type reflecting *Es*-layer occurs at about 95 km (Kotadia et al., 1977). The presence of metallic ions is probably a cause for such strong reflecting sheet of ionization. This fact also supports the intermediate cusp of ionization at 95 km and the low value of α_{eff}. The change in α_{eff} with solar zenith angle also suggests the change in gas composition and hence the relative proportions of various kinds of ions. The high values of α_{eff} at heights below 80 km are probably due to the large influence of negative ions which become almost equal to the number of positive ions at about 70 km during the daytime. These are some explanations for the results reported here, but the chemistry of the lower ionosphere and the ion-kinetics which play important roles in the finer structure of N-h and $\alpha_{eff}-h$ profiles are quite complex and they need to be studied in detail. There is also the case for the investigation of metallic ions in the height range 90–95 km.

Acknowledgements. The authors are very grateful to the scientists of Heinrich Hertz Institute, Academy of Sciences, Berlin (GDR) for their cooperation in establishing a joint research project on ionospheric absorption at Gujarat University, Ahmedabad and for regular discussions and suggestions. One of the authors (A. Gupta) received financial assistance from the Council of Scientific & Industrial Research and the University Grants Commission in the form of research fellowships and contingency grants. We express our due thanks to them also. Very helpful cooperation received from Dr. J.S. Shirke and Dr. S.M. Pradhan of P.R.L. in the construction of N-h profiles with their partial reflection data is gratefully appreciated.

References

Banks, P.: Collision frequencies and energy transfer. Planet. Space Sci. **14**, 1085, 1966

Bremer, J., Singer, W.: Diurnal, seasonal, and solar cycle variations of electron densities in the ionospheric *D*- and *E*-regions. J. Atmos. Terr. Phys. **39**, 25, 1977

Chhipa, G.M., Kotadia, K.M.: Electron density and recombination coefficient profiles obtained in relation to ionospheric absorption and rocket measurements. Ann. Geophys. (Paris)

Cira: COSPAR Working Group IV. Berlin: Akademie Verlag 1972

Groves, G.V.: Atmospheric structure and its variations in the region from 25 to 120 km. Environmental Research Paper No. 368, AFCRL Rep. 71-0410, 1971

Hall, L.A., Schweizer, W., Hinteregger, H.E.: Longterm variation of solar extreme ultraviolet fluxes. J. Geophys. Res. **70**, 2241, 1965

Heraux, L., Higgins, J.E.: Summary of full-disk solar fluxes between 250 and 1940 Å. J. Geophys. Res. **82**, 3307, 1977

Hinteregger, H.E.: The extreme ultraviolet solar spectrum and its variation during a solar cycle. Ann. Geophys. **26**, 547, 1970

Hinteregger, H.E., Hall, L.A., Schmidtke, G.: Solar XUV radiation and neutral particle distribution in July, 1963, thermosphere. Space Res. **5**, 1175, 1965

Horan, D.M.: Coronal electron temperatures associated with solar flares. Publication No. 70-22133. Washington D.C.: Catholic University of America 1970

Huffman, R.E., Paulsen, D.E., Larrabee, J.C., Cairns, R.B.: Decrease in *D*-region $O_2(1\Delta g)$ photoionization rates resulting from CO_2 absorption. J. Geophys. Res. **76**, 1028, 1971

Ivanov-Kholodnyi, G.S., Firsov, V.V.: Short wave solar radiation spectrum at different activity levels. Geomagn. Aeron. **14**, 331, 1974

Jacchia, L.G.: Revised static models of the thermosphere and exosphere with empirical temperature profiles. Cambridge, Mass. USA: Smithsonian Astrophysical Observatory Special Rep. No. 332, 1971

Kotadia, K.M., Chhipa, G.M., Taubenheim, J.: Differences between A1 absorption values measured with sporadic-*E* and normal *E*-layer reflections. Indian J. Radio Space Phys. **6**, 1, 1977

Loidl, A., Schwentek, H.: Mesospheric molecular oxygen density, pressure, and temperature profiles obtained from measurements of solar *H* Lyman-α radiation. J. Geophys. **44**, 107, 1977

Manson, J.E.: Measurements of the Solar Spectrum between 30 and 128 Å. Solar Phys. **27**, 107, 1972

Mechtly, E.A., Bowhill, S.A.: Changes of lower ionosphere electron density with solar activity. J. Atmos. Terr. Phys. **34**, 1899, 1972

Meira, L.G. Jr.: Rocket measurements of upper atmospheric nitric oxide and their consequences to the lower ionosphere. J. Geophys. Res. **76**, 202, 1971

Piggott, W.R., Thrane, E.V.: The electron densities in the *E*- and *D*-regions above Kjeller. J. Atmos. Terr. Phys. **28**, 467, 1966

Schmidtke, G.: EUV indices for solar-terrestrial relations. Philadelphia: XIX Plen. Meeting of COSPAR, 1976

Swider, W. Jr.: Ionization rates due to the attenuation of 1–100 Å non-flare solar X-rays in the terrestrial atmosphere. Rev. Geophys. **7**, 573, 1969

Taubenheim, J., Subrahmanyam, C.V., Klein, G.: Aeronomic results of neutral nitric oxide densities in the lower ionosphere inferred from electron density profiles. Varna: XVIII COSPAR Meeting, 1975

Vidal-Madjar, A., Blamont, J.E., Phissamay, B.: Solar lyman alpha changes and related hydrogen density distribution at the earth's exobase (1969–1970) J. Geophys. Res. **78**, 1115, 1973

Received October 31, 1978; Revised Version February 8, 1979; Accepted February 28, 1979

J. Geophys. 46, 35–42, 1979

Journal of Geophysics

Effective Energy Reception of the Electron Gas per Created Ion Electron Pair*

D. Bilitza

Institut für physikalische Weltraumforschung der Fraunhofer-Gesellschaft, Heidenhofstraße 8, D-7800 Freiburg, Federal Republic of Germany

Abstract. The effective energy reception of the electron gas per ionized neutral particle and per created photoelectron (PE) has been calculated from data of the Satellite AEROS and by use of the Multi Satellite and Incoherent Scatter (MSIS) neutral model. Our results show good agreement with the calculations of Nagy et al. (1969) and the model of Swartz and Nisbet (1972). The comparison with the results of Dalgarno et al. (1968) shows the importance of non-local energy loss of the PEs above 350 km. For a height of 400 km we calculated the portion of non-local photoelectrons to 30% and for 450 km to 50%. Finally, analytical model functions depending on height and Nisbet's density ratio have been approximated to the calculated energy reception.

Key words: Heating efficiency – Electron energy equation – Non-local heating – Photo-electron production – AEROS – Temperature gradient – Non-local photo-electrons – MSIS – Heat gain – Heat loss.

1. Introduction

The aeronomy satellite AEROS with its extensive measurement program meets the ideal requirements for a variety of investigations of the heat budget of the electron gas in the ionosphere. Three terms determine the heat (resp. energy) of the electron gas in the stationary case.

- The heat gain P_g/eV m^{-3} s^{-1} by collisions with the hot PEs, which are created by the ionization of the neutral particles by the solar radiation.
- The heat loss P_t/eV m^{-3} s^{-1} by collisions with the cold neutral particles and ions.
- The heat conduction P_c/eV m^{-3} s^{-1} in the electron gas.

As the electrons move along the magnetic field lines driven by the neutral winds, we have also to consider the amount of energy transported by directed move-

* Dedicated to Professor Dr. K. Rawer on the occasion of his 65th birthday

0340-062X/79/0046/0035/$01.60

ment of the particles. Hoegy and Brace (1978) showed, by using experimental data of the satellite AE-C, that the three terms are sufficient to describe the measured electron temperature. Our special interest concerns the question, how much energy is given to the electron gas on the average per created PE. This factor ε/eV, named heating efficiency, gives us the connection between the PE production rate $q/\mathrm{m}^{-3}\,\mathrm{s}^{-1}$ and the heat gain rate P_g.

$$P_g(h) = \varepsilon(h) \cdot q(h). \tag{1}$$

However, above 300 km ε loses this clear explanation of an average energy input to the electron gas per PE. Because of the lower neutral density, the PEs created in such heights can travel several scale heights before they lose their suprathermal energy. That means the heat input to the electron gas is not only due to locally created PEs but also due to PEs cascading from other heights to the local level. In this case ε is the proportionality factor between the local PE production rate and the local heat gain rate of the electron gas.

2. Calculation of ε

The electron energy equation in our simplified form is

$$P_g = P_t + P_c \tag{2}$$

Using Eq. (1), one gets

$$\varepsilon = \frac{P_t + P_c}{q}. \tag{3}$$

P_t contains energy losses due to elastic and inelastic collisions with neutral particles and ions. The loss rates of the electron elastic collisions and excitation of rotational, vibrational and electronic states of molecular and atomic oxygen and molecular nitrogen are given by Rishbeth and Garriott (1969). By far most energy is lost by *fine-structure excitation of atomic oxygen*. Hoegy (1976) showed that the widely used analytical expression for this loss rate which was introduced by Dalgarno and Degges (1968) is almost twice the values he calculated based on a more involved theory. We have approximated the electron temperature dependence of the ratio of the new to the old value given by Hoegy (1976) in his Fig. 1. The following is our analytical expression.

$$\begin{aligned} P_t^{(\mathrm{new})}/P_t^{(\mathrm{old})} &= 0.432 + 0.1752 \cdot \exp[-1.07 \cdot 10^{-3} \cdot (T_e - 1000)] \\ P_t^{(\mathrm{old})} &= 3.4 \cdot 10^{-18} \cdot n(0) \cdot N_e \cdot (T_e - T_n)/T_n \cdot (1 - 7 \cdot 10^{-5} \cdot T_e) \end{aligned} \tag{4}$$

$n(0), N_e/\mathrm{m}^{-3}$ = atomic oxygen and electron density

$T_n, T_e/K$ = neutral and electron temperature.

The loss rate due to Coulomb collisions in an O^+-gas with assumed charge neutrality ($n(O^+) = N_e$) is (Risbeth and Garriott, 1969).

$$P_t^{(O^+)} = 4.8 \cdot 10^{-13} \cdot N_e^2 \cdot (T_e - T_i)/T_e^{3/2}$$
$$T_i/K \text{ ion temperature.} \tag{5}$$

Using calculations of Chapman and Cowling (1970) and Spitzer (1956) the *heat conduction* of the electrons in a fully ionized neutral gas is given by

$$P_c = \sin^2 \psi \cdot 7.7 \cdot 10^7 \frac{d}{dh}\left(T_e^{\frac{5}{2}} \frac{dT_e}{dh}\right) \tag{6}$$

ψ = magnetic dip.

The factor $\sin^2 \psi$ considers the fact that the electrons are forced to move along the magnetic field lines. For the heights that are considered here, the influence of collisions with neutrals on the heat conduction can be ignored.

The *PE production* is calculated by the number of ionizing processes.

$$q(h) = \sum_k n_k(h) \cdot \sum_{\lambda_i} \sigma_k^{(I)}(\lambda_i) \cdot I_{\lambda_i}(h)$$

$$I_{\lambda_i}(h) = I_{\lambda_i}(h_{\text{EUV}}) \cdot \exp\left[-\frac{1}{\cos \chi} \sum_k \sigma_k^{(A)}(\lambda_i) \cdot \int_{h_{\text{RPA}}}^{h_{\text{EUV}}} n_k(h')\, dh'\right] \tag{7}$$

$k = 1, 2, 3$ meaning O, O_2, N_2

$\sigma_k^{(A/I)}(\lambda_i)/\text{m}^2$ absorption and ionization cross sections of the neutrals k for the solar radiation of wave length λ_i

χ = solar zenith angle

$I_{\lambda_i}(h)/\text{m}^{-2}\,\text{s}^{-1}$ photon flux in height h of wave length λ_i

$h_{\text{EUV/RPA}}$ satellite height for the EUV and RPA measurements, respectively.

3. Used Data

The neutral temperature and the densities of O_2, N_2 and O were taken from the MSIS-model (Hedin et al., 1977). The *electron density and temperature* was measured on board the AEROS-A by the Retarding Potential Anlyser (RPA) of Spenner and Dumbs (1974).

Furthermore we used the *photon fluxes* measured by the AEROS-EUV-Spectrometer of Schmidtke et al. (1974).

The *cross sections* were taken from the reviews of Huffman (1969) and Schoen (1969). The *ion temperature* was calculated by assuming thermal balance between heat gained by the ion gas from the hotter electrons [Eq. (5)] and heat lost to the colder neutrals (Risbeth and Garriott, 1969).

$$T_i = \frac{T_n + 6 \cdot 10^6 \cdot X \cdot T_e^{-\frac{1}{2}}}{1 + 6 \cdot 10^6 \cdot X \cdot T_e^{-\frac{3}{2}}} \qquad X = \frac{N_e}{\sum_k n_k}. \tag{8}$$

For this calculation we have not considered the heat conduction in the ion gas which is neglectable in our height range. As one notices from Eq. (6) we need *the first and second height derivatives of the electron temperature* to calculate P_c. From in situ measurements these values can be obtained only for large height

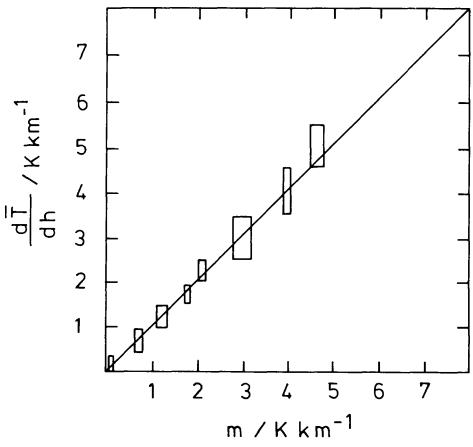


Fig. 1. Comparison of the gradients calculated in different ways (see text)

gradients of the satellite's orbit, as otherwise the results are influenced by the latitudinal and longitudinal gradients. To calculate ε for the global data base, we have gathered the temperatures of the whole AEROS-A mission in 10 degree intervals of geomagnetic latitude (beginning with $-60\ldots-50$) and have approximated these group profiles by a function explained in Appendix I.

For the nighttime when no PEs are produced ($P_g=0$), the temperature gradient can be calculated from the balance between heat conduction and heat loss in the following way. The height profile of the electron temperature can be divided into regions within which the temperature is nearly linear (Appendix I).

$$T_e = T_{e0} + m \cdot (h - h_0). \tag{9}$$

By differentiating Eq. (9) and substituting in Eq. (6) we get

$$P_c = \sin^2 \psi \cdot 7.7 \cdot 10^7 \cdot \tfrac{5}{2} \cdot T_e^{\frac{3}{2}} \cdot m^2. \tag{10}$$

Finally, using Eq. (2), bearing in mind that $P_g=0$, and solving for m we get

$$m = 7.2075 \cdot 10^{-5} \cdot \frac{\sqrt{P_t}}{\sin \psi \cdot T_e^{\frac{3}{4}}} \, K\,km^{-1}. \tag{11}$$

This gives us the possibility to compare the single gradients m with the derivative of the approximated mean function. Figure 1 shows that the mean values of gradient m calculated by using our theoretical considerations fits very well with the empirical approximation, thus justifying our gradient determination. That is obviously also valid for the lower height range where high gradients occur at night.

4. Results

Figure 2 shows the height behaviour of ε, averaged over all latitudes, and the mean absolute deviation from the median value. Figure 2 contains also ε as

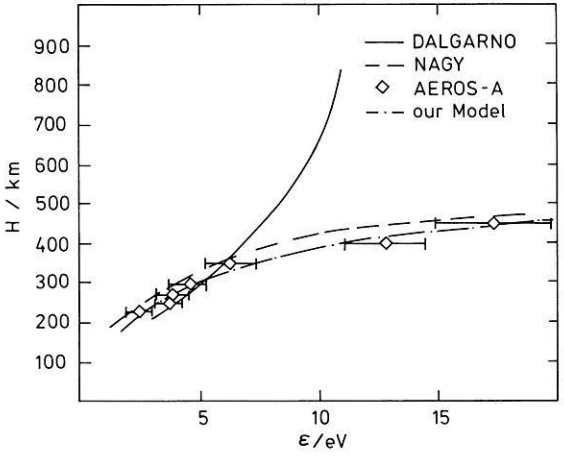


Fig. 2. Height dependance of ε and $\bar{\varepsilon}$ compared to results of other authors

calculated by Nagy et al. (1969) for data of the incoherent scatter station Millstone Hill/USA. The agreement with our values is rather good. The solid line in Fig. 2 represents the energy given to the electron gas per PE, $\bar{\varepsilon}$, which was calculated by Dalgarno et al. (1968) for data from Millstone Hill. Below 350 km the PEs lose their energy totally local, and one can identify the proportionality factor ε as the energy given to the electron gas per PE.

If one wants to extend the energy-input-per-PE-interpretation above 350 km, one has to replace the local PE production rate q by the total amount $\bar{q}$ of PEs per unit volume and second

$$P_g = \bar{\varepsilon} \cdot \bar{q}. \tag{12}$$

Conversely one can calculate the proportion p of non-local PEs q_n with respect to the total rate $\bar{q}$ to zero order by equalizing Eqs. (1) and (12):

$$\begin{aligned} &q/\bar{q} = \bar{\varepsilon}/\varepsilon \quad \text{with} \quad \bar{q} = q + q_n \\ &p = q_n/\bar{q} = 1 - q/\bar{q} = 1 - \bar{\varepsilon}/\varepsilon. \end{aligned} \tag{13}$$

Using the values of Fig. 2 one gets at a height of 400 km a portion of non-local PEs of 30% and at 450 km even 50%.

Figure 3 shows ε depending on the ratio of electron to neutral density and the model for $\bar{\varepsilon}$ established by Swartz and Nisbet (1972). Their model reproduces the fact that the energy of the PEs is divided mainly between electrons and neutrals. That means the energy given to the electron gas depends on the ratio of electron to neutral density. But the losses of the PEs to the electron gas are important only for low PE energies (<1 eV). The main competitors in this energy range are molecular nitrogen and oxygen with their low-lying vibrational states. Therefore the atomic oxygen with higher excitation states is weighted with 0.1. Swartz and Nisbet (1972) restricted their model to the region indicated in Fig. 3, as with decreasing neutral density more and more PEs escape into the protonosphere.

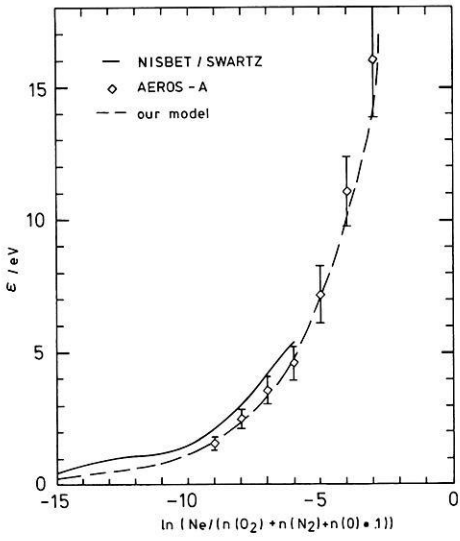


Fig. 3. Density ratio dependance of ε and $\bar{\varepsilon}$ compared to Nisbet's model

This is about the same region for which we can equalize ε and $\bar{\varepsilon}$. This is confirmed by our calculations as shown in Fig. 3. As the concept of the proportionality factor ε simplifies heat budget calculations, we have approximated our values by analytical functions (see Figs. 2 and 3):

$$\varepsilon(h) = 0.307 \cdot \exp(0.00916 \cdot h)$$

$$\varepsilon(x) = 44.7 \cdot \exp(0.364 \cdot X)$$

$$X = \ln \frac{N_e}{n(O_2) + n(N_2) + 0.1 \cdot n(O)}.$$

Both models are valid within the ranges given in Figs. 2 and 3.

Appendix I

Model for the Electron Temperature

For all latitudes we can divide the electron temperature profile in regions with constant height gradient

$$\frac{dT_e}{dh} = \begin{cases} ST_0 & h \ll XS_1 \\ ST_1 & XS_1 \ll h \ll XS_2 \\ \vdots & \vdots \end{cases}. \tag{A7}$$

By connecting these regions with an Epstein step function, that means functions of the following kind

$$f(x)=y_1+\frac{y_2-y_1}{1+\exp\left(-\frac{x-x_1}{d_1}\right)} \tag{A8}$$

with

$$f(x)\begin{cases}\nearrow y_1 & x\ll x_1-d_1\\ \searrow y_2 & x\gg x_1+d_1\end{cases}$$

we get the analytic function

$$\frac{dT_e}{dh}=ST_0+\sum_{k=1}^{M}\frac{ST_k-ST_{k-1}}{1+\exp\left(-\frac{h-XS_k}{D_k}\right)}. \tag{A9}$$

M = number of different regions.
Height integration gives us

$$T_e(h)=\bar{T}+(h-\bar{h})\cdot ST_0$$
$$+\sum_{k=1}^{M}(ST_k-ST_{k-1})\cdot D_k\cdot\ln\left[\frac{1+\exp\left(\frac{h-XS_k}{D_k}\right)}{1+\exp\left(\frac{\bar{h}-XS_k}{D_k}\right)}\right]$$

with the temperature $\bar{T}$ at the reference height $\bar{h}$. The parameters M, $ST_{0\ldots M}$, $XS_{1\ldots M}$ and $D_{1\ldots M}$ have been varied to get the optimum fit to the data.

References

Chapman, S., Cowling, T.G.: The mathematical theory of non-uniform gases, 3rd edn. Cambridge: University Press 1970

Dalgarno, A., Degges, T.C.: Electron cooling in the upper atmosphere. Planet. Space Sci. **16**, 125–127, 1968

Dalgarno, A., McElroy, M.B., Rees, M.H., Walker, J.C.G.: The effect of oxygen cooling on ionospheric electron temperatures. Planet. Space Sci. **16**, 1371–1380, 1968

Hedin, A.E., Reber, C.A., Newton, G.P., Spencer, N.W., Brinton, H.C., Mayr, H.G.: A global thermospheric model based on mass spectrometer and incoherent scatter data MSIS 2. Composition. J. Geophys. Res. **82**, 2148–2156, 1977

Hedin, A.E., Salah, J.E., Evans, J.V., Reber, C.A., Newton, G.P., Spencer, N.W., Kayser, D.C., Alcayde, D., Bauer, P., Cogger, L., McClue, J.P.: A global thermospheric model based on mass spectrometer and incoherent scatter data MSIS 1. N_2 density and temperature. J. Geophys. Res. **82**, 2139–2147, 1977

Hoegy, W.R.: New fine structure cooling rate. Geophys. Res. Lett. **3**, 541–544, 1976

Hoegy, W.R., Brace, L.H.: The importance of electron heat conduction in the energy balance of the *F*-region. Geophys. Res. Lett. **5**, 269–272, 1978

Huffman, R.E.: Absorption cross sections of athmospheric gases for use in aeronomy. Can. J. Chem. **47**, 1823–1830, 1969

Nagy, A.F., Fontheim, E.G., Stolarski, R.S., Beutler, A.E.: Ionospheric electron temperature calculations including protonospheric and conjugate effects. J. Geophys. Res. **74**, 4667–4676, 1969

Rawer, K., Ramakrishnan, S., Bilitza, D.: International Reference Ionosphere 1978, International Union of Radio Science (URSI) Brussels 1978

Risbeth, H., Garriott, O.K.: Introduction to ionspheric physics. London: Academic Press 1969

Schmidtke, G., Schweizer, W., Knothe, M.: The AEROS-EUV Spectrometer. J. Geophys. **40**, 577–584, 1974

Schoen, R.L.: Laboratory measurements of photoionization, photoexcitation, and photodetachment. Can. J. Chem. **47**, 1979–1890, 1969

Spenner, K., Dumbs, A.: The retarding potential analyzer on AEROS-B. J. Geophys. **40**, 585–592, 1974

Spitzer, L.: Physics of fully ionized gases. New York: Interscience 1956

Swartz, W.E., Nisbet, J.S.: Revised calculation of F region ambient electron heating by photoelectrons. J. Geophys. Res. **77**, 6259–6261, 1972

Received November 3, 1978; Revised Version March 16, 1979

J. Geophys. 46, 43–56, 1979

Journal of Geophysics

Empirical Model of Global Electron Temperature Distribution Between 300 and 700 km Based on Data From Aeros-A*

K. Spenner and R. Plugge

Institut für physikalische Weltraumforschung der FhG, D-7800 Freiburg, Federal Republic of Germany

Abstract. An empirical model function of the global electron temperature distribution has been determined based on the measurements of the planar Retarding Potential Analyzer on-board the Aeros-A satellite. The model represents the mean temperature between 300 and 700 km altitude at 0300 LT and 1500 LT depending on latitude, longitude, and height. The model values are compared with all the measured data to show the accuracy achieved and the mean spread of the data for different latitudes. A clear correlation was not found between the electron temperature and geophysical indices such as K_p or sunspot number for the period of low solar activity between January and August, 1973. Seasonal and annual effects could not be detected. The mathematical background and method used to generate the model function is described in the appendix.

Key words: Ionosphere – Electron temperature – Satellite observations – Retarding potential analyzer.

Introduction

The electron temperature shows a characteristic global pattern. The distribution is sometimes masked by geophysical events. However, averaged data show clearly a quite constant pattern as measured by different satellites (Brace, 1970; Dorling and Raitt, 1976; Spenner et al., 1978). In the last ten years a large quantity of temperature data has become available and only a small part of them could be published. A review of the available observations in the F-region was recently given by Schunk and Nagy (1978).

An effective way to handle large amounts of data and to extract an average distribution from them can be achieved by modeling measurements by analytical functions. A representative model allows one to compare different sets of data in a quite general form and is useful in predicting mean values. A global

* Dedicated to Professor Dr. K. Rawer on the occasion of his 65th birthday

0340-062X/79/0046/0043/$02.80

temperature model has been generated with ESRO-4 data (Dorling and Raitt, 1976). It shows the dependence of the temperature distribution on altitude, latitude and local time, but it does not show a longitudinal dependence.

We will present in this paper an additional analytical model of the electron temperature. For this purpose we use the data of the planar Retarding Potential Analyzer (Spenner et al., 1974) on-board the AEROS satellite. The satellite was in orbit between December, 1972 and August, 1973 with an apogee of 800 km and a perigee of 240 km. Its path was approximately polar (97°) covering nearly the whole globe. The sun synchronous orbit allowed us to always achieve temperature data at 0300 LT and 1500 LT. Therefore, the AEROS orbit is ideal for investigating a longitudinal structure at constant local time. The model provides the mean electron temperature distribution in concise form representing more than 10,000 measurements.

Model Functions

The model must describe the temperature as a function of the latitude θ, longitude ϕ and height h. The model function $T(h, \theta, \phi)$ can be expressed by spherical harmonics in the form

$$T(h,\theta,\phi)=\sum_{k=0}^{K}\sum_{n=0}^{N}\sum_{m=-n}^{n} a_{knm}\cdot Y_{nm}(\theta,\phi)\cdot F_k(h) \tag{1}$$

where a_{knm} are parameters of series expansion in spherical harmonics, $Y_{nm}(\theta, \phi)$ spherical harmonics, $F_k(h)$ height functions.

Such a sum describes any complex distribution in a worldwide form. The number of necessary terms K, and N depends on the particular distribution and on the desired accuracy.

The parameters a_{knm} can be determined from the relation

$$a_{knm}=\int_h\int_\phi\int_\theta T_M\, Y_{nm}\, F_k\, d\tau \tag{2}$$

where T_M is the measured temperature and $d\tau$ is a space element of the considered volume. A detailed deduction of Eqs. (1) and (2) is shown in the Appendix.

To some extent a similar attempt to model atmospheric parameters was chosen by the OGO-6 model, which describes neutral temperature and composition (Hedin et al., 1974). It also uses spherical harmonics to describe the latitudinal distributions. However, the longitudinal distribution was not considered, and the height function was not determined in the form described here.

The time dependence is not introduced in Eq. (1) because all available data belong to two constant local times. We determine separately two sets of parameters for day- and nighttime according to Eq. (2). Other possible variations such as seasonal, geomagnetic or solar activity variations are not introduced at this stage. These variations will be analyzed separately in a later section. Therefore, all measurements were used in the analysis without any preselection for special geophysical conditions.

Data Basis

The AEROS Retarding Potential Analyzer was designed to measure a temperature in half a second with a repetition period of 18 s determined by telemetry capacity. This measuring interval provided at least one value within every degree of latitude. Because of problems with the satellite-born triggering system only a few orbits are available with a resolution of one degree. Most of the measurements have a resolution of 10 degrees. The data are measured during complete orbit revolutions. Every second or third orbit was recorded. The data coverage was almost homogeneously distributed in longitude and latitude during a period of some days. However, it took three months to achieve data at all heights between apogee and perigee.

All data measured during the first 100 days of 1973 were grouped in ranges of 10° latitude, 60° longitude and 100 km altitude. The relatively wide longitudinal range was chosen to get at least 10 measurements in each cell. The center of the longitudinal cells was shifted in steps of 10°, to get 36 longitudinal cells. One measurement was always contributed to 6 different longitudinal cells centered at 10° longitude and latitude and at 100 km levels between 300 and 700 km. The particular temperature data in each of the 3070 cells were averaged. Standard deviations were calculated separately for each cell. Data being outside of two standard deviations were not used in order to suppress possible runaways under unfavorable conditions. The relatively small amount of data in each cell and the sometimes quite large temporal changes of the electron temperature suggested that the mean temperature in each range be smoothed by a running weighted average between cells directly adjacent in latitude and height. These smoothed data are assigned $T_M(\theta, \phi, h)$.

Calculation of the Model

In the beginning we used three coordinate systems: geographic, geomagnetic and modified dip coordinates. For each system the standard deviations in all the particular cells were determined and averaged. The smallest mean standard deviation was by far achieved in the geomagnetic system. This suggests that the geomagnetic system gives the most appropriate description. Therefore, further analysis is done only with geomagnetic coordinates.

The parameter a_{knm} were calculated with Eq. (2). The integral was replaced by a sum over the 3070 cells. The appropriate order of the series expansion was determined by a comparison between the model function $T(N, K)$ with the mean temperatures T_M in the particular cells. The mean deviation d, which is given by

$$d = \frac{\Sigma \delta \pm \Sigma |\delta|}{n} \tag{3}$$

where $\delta = (T_M - T)\, T^{-1}$ and n is number of measurements, is shown in Fig. 1 as a function of N, the highest order of spherical harmonics, for $K = 3$. A third order approximation in height h is necessary to reproduce the temperature maximum

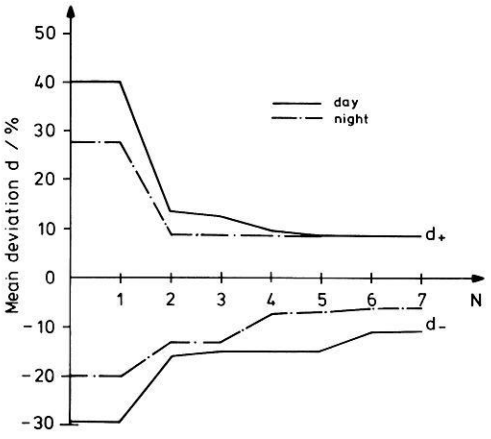


Fig. 1. Deviation between observed mean temperatures and the order N of the model function

at low altitudes in the equatorial area. However, analyzing the height-profile of temperature at a fixed location – except low latitudes – it is usually sufficient to describe T as a linear function of height h. Fig. 1 indicates that the highest useful order for N is 6.

A higher number of parameters did not essentially improve the temperature function. The average temperature is reproduced well and the mean scatter in particular cells is less than 10%. A similar analysis made for K confirms that K greater than 3 does not improve significantly the model function. For all further investigations we use the temperature function T with $N=6$ and $K=3$. However, Fig. 1 suggests that a second order approximation in N gives already an approximation with 15%.

Temperature Distribution

The model function $T(\theta, \phi)$ is presented in Figs. 2 and 3 at two fixed heights at 1500 LT and 0300 LT, respectively. The latitudinal distribution derived earlier by several authors (Brace, 1970; Spenner and Dumbs, 1974; Dorling and Raitt, 1976) is reproduced in the global distribution derived in this study. In addition a longitudinal distribution occurs. One may argue that a longitudinal distribution is only the result of temporal variations. Figures 2 and 3, however, demonstrate clearly that the general longitudinal pattern is quite similar at different heights and during day and nighttime. Ten thousand independent measurements cannot produce such a consistent picture without a geophysical source.

The longitudinal distribution of the model function is shown in more detail (at 500 km altitude) in Figs. 4 and 5. The highest longitudinal temperature variation is approximately 400 K along a fixed magnetic latitude in the southern hemisphere at 1500 LT. The variations at other latitudes are usually between 100 and 300 K. Low temperatures are obtained between 60° and 140° geomagnetic longitude in the southern hemisphere and between 260° and 360° in the northern hemisphere. At 0300 LT the longitudinal variation along a fixed

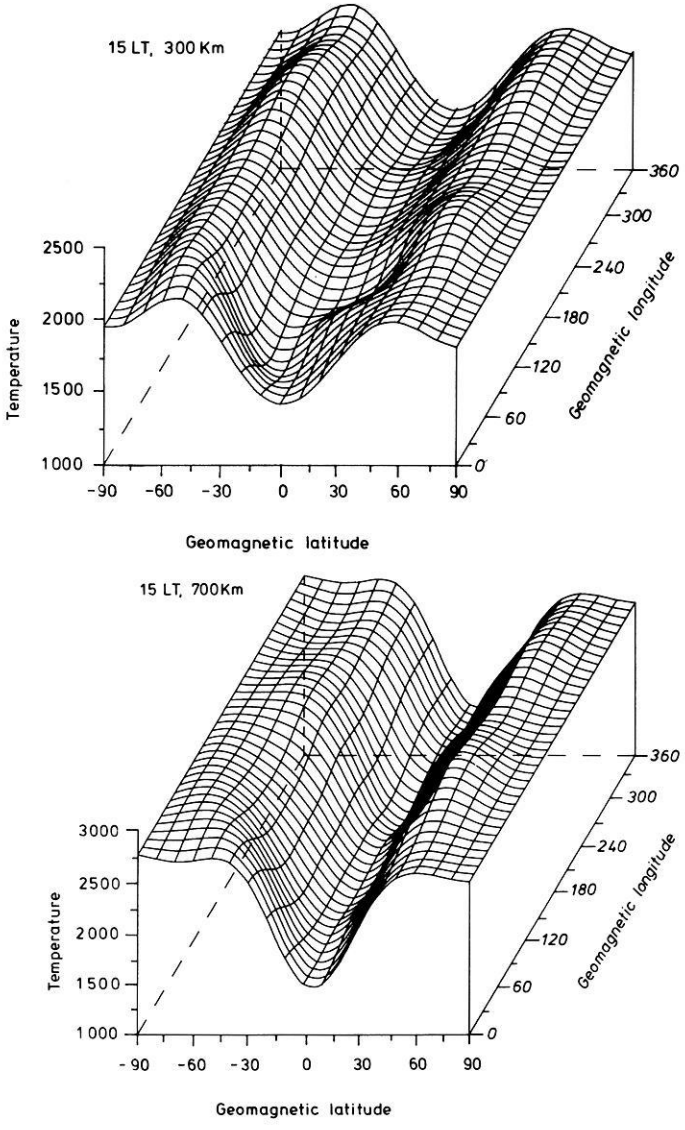


Fig. 2. Model distribution at 1500 LT for 700 and 300 km

magnetic latitude is also well pronounced. The locations of high and low temperature are close to that observed 1500 LT. A longitudinal variation of the neutral gas density was recently reported by von Zahn and Fricke (1978). Comparing the electron temperature variation with the neutral Argon density variation along a fixed latitude range we find that a high electron temperature occurs on locations with a low Argon density.

At 1500 LT the latitudinal temperature distribution shows a deep minimum close to the equator (Fig. 4) at 500 km. In contrast to the daytime distribution there are two small minima on both sides of the equator (Fig. 5) at 0300 LT.

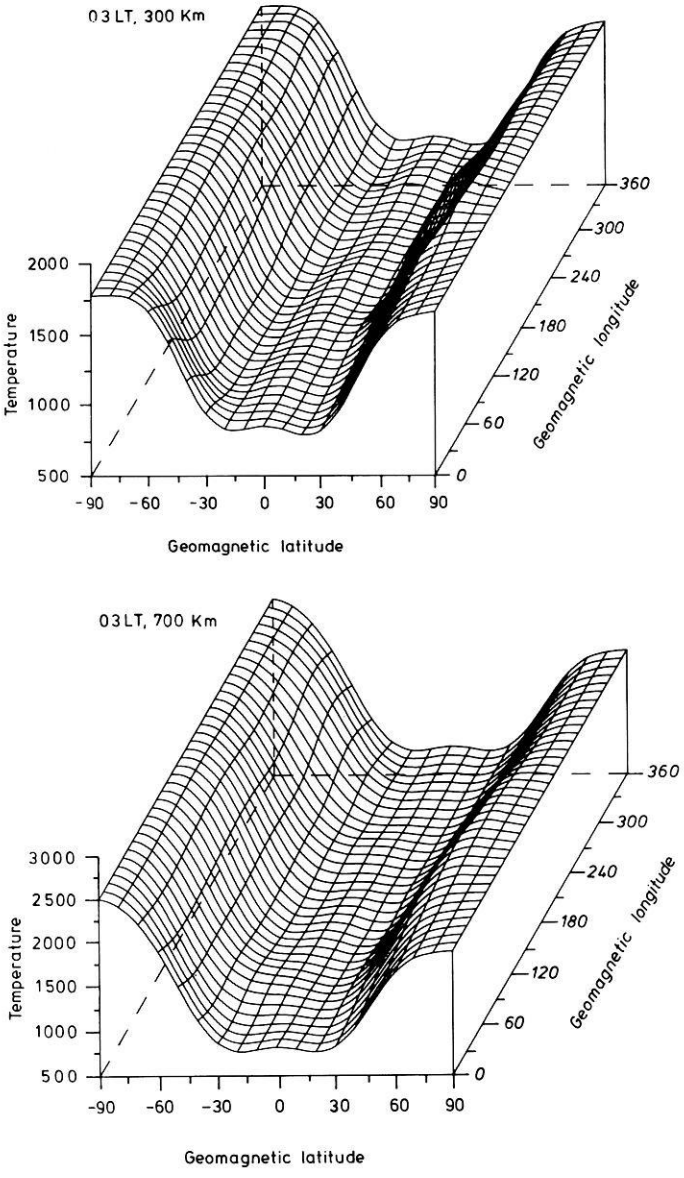


Fig. 3. Model distribution at 0300 LT for 700 and 300 km

The latitudinal and height variation of the temperature is given in Figs. 6 and 7 at 0° geomagnetic longitude. The isotherms at 1500 LT indicate a temperature minimum at approximately 400 km altitude at low latitudes. Little variation with altitudes may be in small error because the data required to derive a height variation were obtained over a relatively long time period under different geophysical conditions. The isotherms represent the averaged distribution in a three months period but they may possibly not reproduce well details in a height profile.

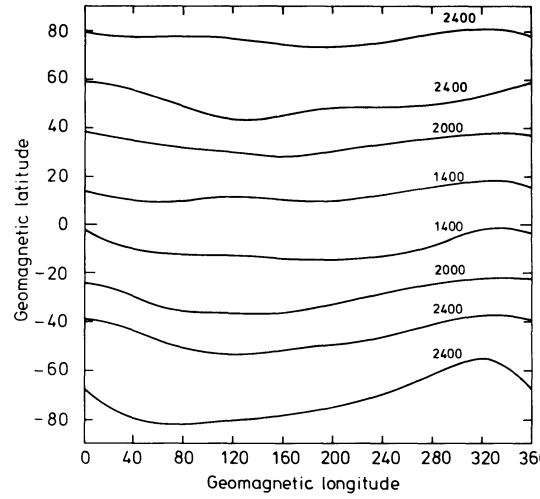


Fig. 4. Isotherms at 500 km altitude and 1500 LT

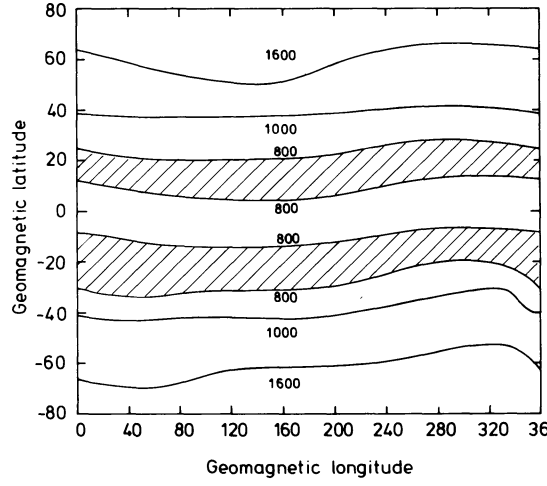


Fig. 5. Isotherms at 500 km altitude and 0300 LT

Intercomparison Between Model and Measurements

We calculated the model temperatures for all individual measurements and compared them with each other. To evaluate how well the model represents the data we have calculated the model temperature for each data point and compared it to the measurement. To demonstrate that the model represents the latitudinal variation adequately, we have determined the mean deviation of the measured values from the model values at different latitudes for both 1500 LT and 0300 LT. The calculated mean difference $\overline{\Delta T}$ between all AEROS measurements and the model was less then 30 K. This confirms that the computed parameters a_{knm} are satisfactory, and that the model function represents the

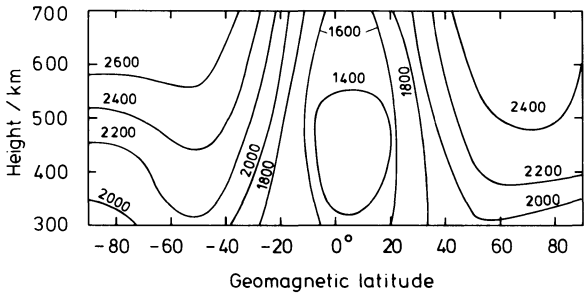


Fig. 6. Isotherms at 0° geomagnetic longitude and 1500 LT

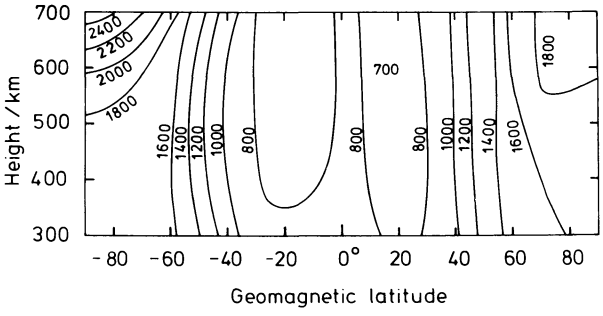


Fig. 7. Isotherms at 0° geomagnetic longitude and 0300 LT

global mean value of the measurements. A more detailed analysis has to demonstrate whether the model describes the mean value of selected data at different latitudes well enough and to determine how large the scatter in the measurements is. For this purpose we determined the mean deviation d_D according to Eq. (3) for different latitudes. δ in Eq. (3) is the difference between the measured temperature and the model function at the same location divided by the model-temperature. The d_D-value gives a relative mean scatter of 7% integrated over all longitudes for the first 100 days in 1973.

Figures 8 and 9 show the boundaries of the determined mean deviations for different latitudes. It is evident that the data spread is increased at higher latitudes. These relative variations during night (Fig. 9) are much higher than during day (Fig. 8) with the exception of the equatorial area. The daytime model is a few percent too high at low latitudes and too low at midlatitudes. Areas with strong temperature gradients are smoothed a little. These small deviations of the average values are not of great importance since the temporal variations given by the scatter are much larger. The uncertainty in predicting temperatures is mainly given by the scatter of particular measurements. The calculated d_D value varies only slightly during the AEROS-A mission period of eight months and does not change the general behavior.

The model function represents an average temperature during the time period considered. Differences ΔT between particular data and the model may reflect geophysical events during a few day period or seasonal and annual effects for a longer time period. We tried to find possible relations between the temperature differences $\overline{\Delta T}$ and geophysical indices. To this end we correlated

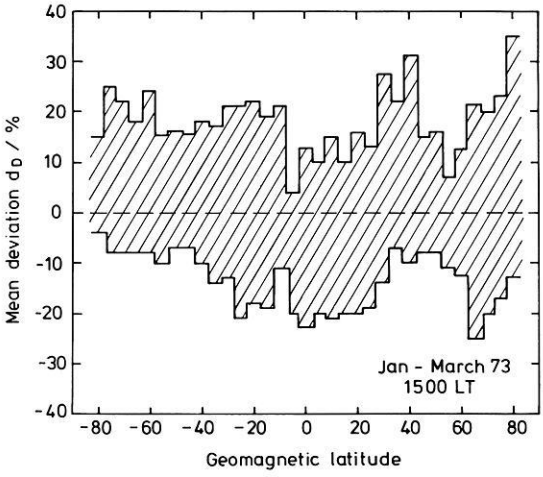


Fig. 8. Mean deviation between model and measured data points at 1500 LT

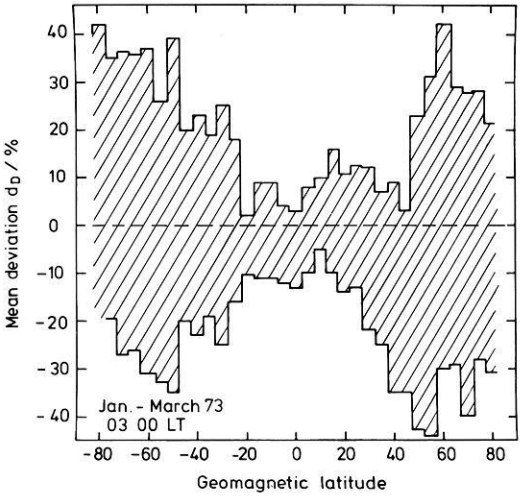


Fig. 9. Mean deviation between model and measured data points at 0300 LT

$\overline{\Delta T}$ with the K_p-value, the A_p-value and with the sunspot number during the eight months period. The analysis was done for different geomagnetic latitude ranges. The result did not show any correlation between these indices. This indicates clearly that the temperatures do not follow the indices in a simultaneous and clear relation. It confirms an earlier investigation (Spenner, 1975), which did not show a significant relation between K_p and electron temperature.

Monthly mean temperature differences $\overline{\Delta T}$ were used to search for a seasonal or annual variation. To obtain data for a period of one year for a Fourier analysis, the measurements of one hemisphere were used again six months later at the other hemisphere mirrored at the geographic equator. This procedure can only provide a rough indication of the annual variation when no complete period of measurement is available. The calculated Fourier coefficients

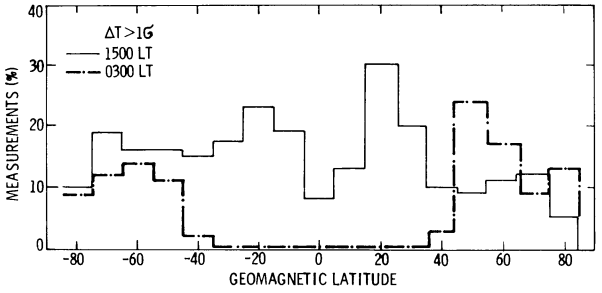


Fig. 10. Relative frequency of measurements observed outside the σ_M boundary

did not show a significant annual or seasonal period either at low, middle or high latitudes. This means that the model function determined for January to March, 1973, must be valid for the second part of the AEROS-A mission from April to August, 1973. Indeed the calculated mean deviation d_D for the period was not considerably different from the values shown already in Figs. 8 and 9. Seasonal and annual effects must be smaller than short time or local changes. As a consequence we argue that these effects do not exceed 15% during low solar activity.

Individual data points are sometimes far away from the mean temperature value. A few such values considerably increase the mean deviation $(\overline{\Delta T}) = 1/n \Sigma |\Delta T|$ and the standard deviation σ between model and measurements. The determined σ_M value between particle measurements and the mean temperature was 440 K and 490 K for day and nighttime, respectively, where

$$\sigma_M^2 = \frac{\Sigma (T - \overline{\Delta T})^2}{n_R} \tag{4}$$

n_R = number of measurements in a particular cell range. The spread in ΔT does not follow a statistical error distribution. At 1500 LT more than 80% and at 0300 LT more than 90% of all measurements are within the 1σ range (where as for a Gaussian distribution one expects only 60%). Most of the runaways are probably caused by extreme geophysical conditions and restricted to a particular area. Temperatures which differ more than 1σ from the model are considered to be outside of quiet or usual conditions.

The temperatures outside the σ_M boundary are used to look for areas, where data spread may occur. To this end, we determined the percentage of spread values relative to all measurements as a function on geomagnetic latitude and longitude. Figure 10 shows small peaks with more than 20% of spread data at about 20° latitude of both sides of the equator at 1500 LT. At 0300 LT all the spread temperatures occur in an area around 60° latitude.

A corresponding analysis of the spread data for different geomagnetic longitudes indicate peaks at 60° and 170° longitude during daytime. The peaks seem to be higher than the expected random noise at 1500 LT.

Summary and Conclusions

The empirical model based on the AEROS-A data predicts electron temperatures for 0300 LT and 1500 LT. Temperatures can be calculated worldwide either with geographic or geomagnetic coordinates between 300 and 700 km altitude. The model temperature is in quite good agreement with ground observations done by incoherent radar scatter technique. This is demonstrated in a second paper (Spenner et al., 1979). The model temperatures are considerably lower than the values predicted by the ESRO 4 model (Dorling and Raitt, 1976), though both measurements were taken at the approximately same time period. The observed longitudinal variations show maxima of the electron temperature in areas where the density of neutral Argon observed by the ESRO 4 gas analyzer (von Zahn and Fricke, 1978) exhibits a minimum in the northern hemisphere ($\approx 120°$, geomag. long.) and southern hemisphere ($\approx 320°$ geomag. long.). The temperature minima are located as well where the neutral density maximum of Argon occurs with the exception of the South Atlantic anomaly region.

Appendix

1. Spherical Harmonics

The oscillation of a sphere can be described by the differential equation

$$\Delta U = \frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} \tag{I}$$

where U is the wave function, c is the wave velocity, and t the time. U may be written in the form

$$U = F(r, \theta, \phi) \cdot T(t)$$

where r is the radius of the sphere, θ the latitude and ϕ the longitude. We are interested only in the part of a solution of Eq. (I) depending on θ and ϕ. It can be shown that this part follows the differential equation.

$$\frac{1}{\cos\theta} \frac{\partial}{\partial\theta} \left(\cos\theta \frac{\partial Y}{\partial\theta}\right) + \frac{1}{\cos^2\theta} \frac{\partial^2 Y}{\partial\phi^2} + \alpha Y = 0 \tag{II}$$

where Y depends only on θ and ϕ, α is constant in θ, ϕ.

All the algebraic (in $\cos\theta$, $\cos\phi$) solutions of Eq. (II) are given by the spherical harmonics Y_{nm} of the form

$$Y_{n,m}(\theta, \phi) = \sin^{|m|}\theta \frac{d^{|m|} P_n(\cos\theta)}{(d\cos\theta)^{|m|}} \cdot e^{\pm im\phi} \cdot C_{nm} \tag{III}$$

where P_n is the Legendre polynomial of order n and C_{nm} is a normalizing factor. Replacing the exponential term by sin and cos functions in (III) and normalizing the spherical harmonics so that they are 1 at the north pole (Bronstein-Semendjajew, 1956) we get

$$Y_{nm}(\theta, \phi) = C_{nm} \sin^{|m|}\phi \cdot \frac{d^{|m|}}{d(\cos\theta)^{|m|}} P_n(\cos\theta) \cdot R_m(\phi)$$

where

$$R_m(\phi)=(-1)^m \cdot \begin{cases} \sin m\phi & \text{for } m<0 \\ \cos m\phi & \text{for } m\geqq 0, \end{cases}$$

$$C_{nm}^2=\frac{(n-|m|)!}{(n+|m|)!}\cdot\frac{1}{4\pi} \tag{IV}$$

The integral over the volume Ω becomes

$$\oint Y_{nm}\cdot Y_{n'm'}\,d\Omega=\delta_{nn'}\cdot\delta_{mm'} \tag{V}$$

because the Eigensolutions of differential equations are orthogonal.

2. *Functions in Height*

The temperature can be described as

$$T(h,\theta,\phi)=\sum_{k=0}^{K}\sum_{n=0}^{N}\sum_{m=-n}^{n} a_{knm}\,Y_{nm}(\theta,\phi)\cdot F_k(h) \tag{VI}$$

where Y_{nm} are the spherical harmonics as defined by Eq. (IV) and F_k are polynomials in h of degree k. Any function T can be expressed in such a series expansion.

For easy computation of the a_{knm} we desire that (see later, computation of the a_{knm})

$$I=\int Y_{nm}F_k\cdot Y_{n'm'}F_{k'}\,d\tau=\delta_{mm'}\delta_{nn'}\delta_{kk'}. \tag{VII}$$

The integral becomes with Eq. (V)

$$I=\oint_{\Omega,h} Y_{nm}Y_{n'm'}\cdot F_k\cdot F_{k'}h^2\,dh\,d\Omega$$

$$=\delta_{nn'}\delta_{mm'}\int_{h_1}^{h_2} h^2F_kF_{k'}\,dh$$

and with (VII) we get

$$\int_{h_1}^{h_2} h^2F_kF_{k'}\,dh=\delta_{kk'} \tag{VIII}$$

where h_2 and h_1 are the boundaries of the considered height interval.

The expansion of the F_k is

$$F_0=a_{00}$$

$$F_1=a_{10}+a_{11}h.$$

$$\ldots$$

$$F_n=a_{n0}+a_{n1}h+\ldots a_{nn}h^n.$$

Eq. (VII) yields for $k=k'=0$

$$a_{00}=\frac{1}{\sqrt{\int_{h_1}^{h_2} h^2\,dh}}$$

and

$$a_{00}\cdot\int_{h_1}^{h_2}(a_{10}+a_{11}h)h^2\,dh=0 \quad \text{for } k'=0,\ k=1$$

$$\int_{h_1}^{h_2}(a_{10}+a_{11}h)^2h^2\,dh=1 \quad \text{for } k=k'=1$$

Table 1

h	n	m	a_{knm} (1500 LT)	a_{knm} (0300 LT)
0	0	0	9290 E+1	4976 E+1
1	0	0	6506 E+0	1651 E+0
2	0	0	2826 E−1	6561 E−1
3	0	0	−1044 E+0	−2054 E−1
0	1	−1	5462 E−1	−3974 E−1
1	1	−1	−2884 E−1	−2404 E−1
2	1	−1	8509 E−2	−6016 E−2
3	1	−1	−8540 E−2	−4226 E−2
0	1	0	−1087 E+0	−8254 E−2
1	1	0	−1080 E−1	−6363 E−1
2	1	0	−1542 E−1	−6671 E−1
3	1	0	3420 E−1	−2064 E−1
0	1	1	−3246 E−1	3363 E−1
1	1	1	1006 E−1	−6907 E−2
2	1	1	3788 E−1	−1126 E−1
3	1	1	−8795 E−2	3128 E−2
0	2	−2	−4113 E−1	−4121 E−1
1	2	−2	−5864 E−2	5811 E−2
2	2	−2	1255 E−1	1958 E−1
3	2	−2	−5917 E−2	1178 E−1
0	2	−1	−2183 E+0	−2973 E+0
1	2	−1	−5367 E−2	6294 E−2
2	2	−1	2327 E−1	3924 E−1
3	2	−1	8651 E−2	−4952 E−2
0	2	0	1735 E+1	1684 E+1
1	2	0	1680 E+0	2337 E+0
2	2	0	−1453 E+0	6999 E−1
3	2	0	6019 E−1	−2839 E−3
0	2	1	3356 E+0	1802 E+0
1	2	1	5824 E−1	9037 E−2
2	2	1	−6681 E−1	2208 E−1
3	2	1	4200 E−2	−4697 E−2
0	2	2	5769 E−1	−2481 E−1
1	2	2	1343 E−1	−1254 E−1
2	2	2	−1604 E−1	6481 E−2
3	2	2	−1403 E−1	−2056 E−2

This is a linear homogeneous equation which yields a linear relation between a_{10} and a_{11}, and a quadratic equation for a_{10}. In analogous manner one can successively compute the coefficients a_k for higher degrees.

In the case of $h_1=3$ and $h_2=7$ (normalized to 100 km) it becomes

$$F_0=0.86066\cdot 10^{-1}$$

$$F_1=-0.3965+0.6910\cdot 10^{-1}h$$

$$F_2=1.5203-0.6105h+0.575\cdot 10^{-1}h^2$$

$$F_3=-6.6602+4.254h-0.8616h^2+0.55783\cdot 10^{-1}h^3 \qquad \text{(IX)}$$

3. Computation of a_{knm}

The derived orthogonal function Y_{nm} and F_k allows us to derive easily the series expansion parameters a_{knm}.

Equation (VI) multiplied by $Y_{n'm'} F_{k'}$ becomes

$$T Y_{n'm'} F_{k'} = Y_{n'm'} F_{k'} \sum_{k,n,m} a_{knm} \cdot F_k .$$

Integration over $d\tau$ yields

$$\int T Y_{n'm'} F_{k'} \, d\tau = \int \sum_{k,n,m} a_{knm} Y_{nm} Y_{n'm'} F_k F_{k'} \, d\tau$$

$$= \sum_{k,n,m} a_{knm} \int Y_{nm} \cdot Y_{n'm'} \, d\Omega \int h^2 F_k \cdot F_{k'} \, dh$$

It becomes with use of Eqs. (V) and (VIII)

$$\int T Y_{n'm'} F_{k'} \, d\tau = a_{k'n'm'} \tag{X}$$

The a_{knm} values from the AEROS-A measurements determined by Eq. (X) and using the mean temperature data as described earlier are listed in Table 1. The table presents the a_{knm} coefficients up to the second order in N, M and the third order in K. A better approximation can be achieved by higher order in N, M as demonstrated in Fig. 1.

Acknowledgement. We are indebted to the BMFT and the DFVLR/BPT for sponsoring the Project AEROS.

References

Brace, L.H.: The global structure of ionosphere temperature. Space Res. **10**, 633–651, 1970

Bronstein-Semendjajew: Taschenbuch der Mathematik, S. 401. Moskau: Staatlicher Verlag für technisch-theoretische Literatur 1956

Dorling, E.B., Raitt, W.J.: Electron temperature models of the thermosphere to 1,100 km based on ESRO-4 measurements. Planet. Space Sci. **24**, 739–747 (1976)

Hedin, A.E., Mayr, H.G., Reber, C.A., Spencer, N.W., Carignan, G.R.: Empirical model of global thermospheric temperature and composition based on data from the Ogo 6 quadrupole mass spectrometer. J. Geophys. Res. **79**, 215–225 (1974)

Schunk, R.W., Nagy, A.F.: Electron temperatures in the *F*-region of the ionosphere: Theory and observations. Rev. Geophys. Space Phys. **16**, 355–399 (1978)

Spenner, K.: Quiet and disturbed electron temperature and density at different latitudes during daytime. Space Res. **15**, 363–368 (1975)

Spenner, K., Bilitza, D., Plugge, R.: Intercomparison between AEROS electron temperature model and mean temperature profiles of different incoherent scatter radar stations. J. Geophys. **46**, 57–61 (1979)

Spenner, K., Dumbs, A.: The retarding potential analyzer on AEROS-B. J. Geophys. **40**, 585–592 (1974)

Spenner, K., Dumbs, A., Lotze, W., Wolf, H.: Correlation between daytime electron temperature and density variations at low latitudes. Space Res. **14**, 259–263 (1974)

Spenner, K., Plugg, R.: Electron temperature model derived from AEROS-A. Space Res. **18**, 241–244 (1978)

Zahn, U. von, Fricke, K.H.: Empirical models of global thermospheric composition and temperature during geomagnetically quiet times compared with ESRO 4 gas analyzer data. Rev. Geophys. Space Phys. **16**, 169–175 (1978)

Received October 31, 1978; Revised Version February 7, 1979

J. Geophys. 46, 57–61, 1979

Journal of Geophysics

Intercomparison Between Aeros Electron Temperature Model and Mean Temperature Profiles of Different Incoherent Scatter Radar Stations*

K. Spenner, D. Bilitza, and R. Plugge

Institut für physikalische Weltraumforschung der Fraunhofer-Gesellschaft, Heidenhofstraße 8, D-7800 Freiburg, Federal Republic of Germany

Abstract. The *F*-region temperature model based on AEROS-A measurements is compared with mean height profiles of incoherent scatter radar observations at middle and low latitudes and at the magnetic equator. The mean values of the ground station are in good agreement with the model temperature confirming that the generated model is representative for the electron temperature within the natural spread produced by geophysical events.

Key words: Ionosphere – *F*-region – Electron temperature model – Aeros-A measurements – Electron temperature height profiles – Incoherent scatter observations.

Introduction

In the last decade a large amount of electron temperature data became available observed mainly by satellites and ground-based Incoherent Scatter Radar (ISR) Stations. The satellite measurements can determine the world-wide temperature distribution in the *F*-region. The satellites cover large areas within relatively short time periods. They are considered an effective tool to obtain global temperature patterns. Along a satellite orbit many parameters change simultaneously and, therefore, it becomes uncertain quite often whether particular observations are modified more by geographic, temporal or time variations-mentioning only a few of possible sources. Therefore, significant global temperature patterns based on satellite data can only be obtained by a very large amount of data on statistical basis and even then the separation of different influences is not always unequivocal. In particular, height profiles constructed from a large number of in situ observations can be strongly affected by temporal changes when the data are collected during a long time period. The ground-based radar stations are better able to determine temperature profiles at a fixed location

* Dedicated to Professor Dr. K. Rawer on the occasion of his 65th birthday

0340-062X/79/0046/0057/$01.00

in a short time. They produce the most reliable height profiles and are an ideal test of the satellite observations.

The AEROS temperature model (Spenner and Plugge, 1979) based on more than 10,000 measurements describes the global distribution in the height interval 300 to 700 km for the first 7 months of 1973. A crucial test of the model can be obtained by comparing height profiles of the model with those of different ISR stations.

As the number of passes near active ISR stations is rather scarce, it seems reasonable to check the model with mean profiles of the available ground observations for equatorial, low and middle latitudes.

Even when particular in situ measurements were in good agreement with the ISR ground observations (Spenner and Rawer, 1978), it has to be demonstrated that the modeling functions represent the mean temperature pattern. The aim of this paper is to check the AEROS temperature model by available ground observations and to compare results of the model with other observations. A large amount of such different results is summarized by Schunk and Nagy, 1978.

ISR-Data Base

Measurements of the three ISR stations, Millstone Hill/USA, Arecibo/Puerto Rico and Jicamarca/Peru have been used for our comparison. They are considered as representative for middle, low and equatorial latitudes.

Usually the stations have measured periods of 1–3 days each month and measuring times 10–30 min per profile. The processed data are collected on magnetic tapes for up to 4 years. The tapes used for our comparison are summarized in Table 1, giving the periods of measurements, the 12 months running mean of the Zurich Sunspot Number R_{12} and the location of the ISR-station. Our mean profiles have been obtained by averaging over the indicated periods (Table 1). For comparison with the AEROS model we used only ISR-measurements within the local time ranges 14–16 LT and 2–4 LT. Even though a seasonal

Table 1. Location and time of the used data

Tape name	Station	Location		time	R_{12}	Number of used profiles
		geog.	geom.			
N35	Arecibo Puerto Rico	18N 293E	30N 2E	12/71–12/72	58– 73	50
W287	Jicamarca Peru	12S 283E	1N 352E	11/66– 4/69	75–110	700
MH7	Millstone Hill USA	43N 289E	54N 357E	2/72–10/75	25– 71	100
			AEROS-A	12/72– 8/73	40– 50	

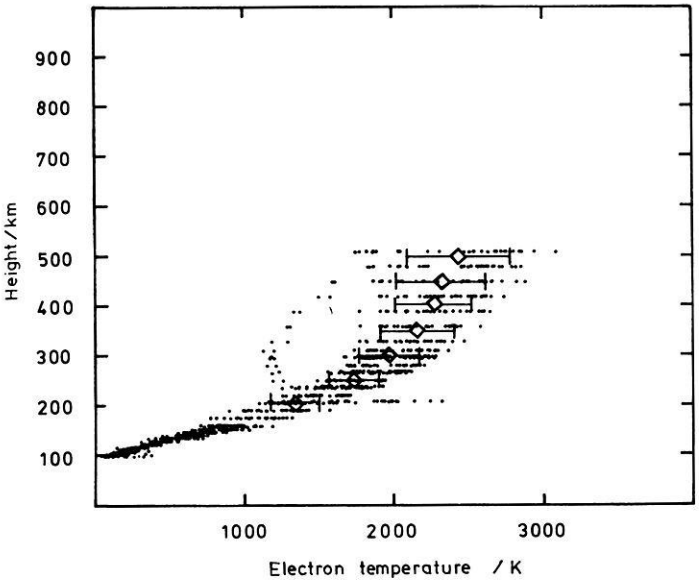


Fig. 1. Millstone Hill ISR-data for winter midday from 1972 to 1973 and median value with mean absolute deviation

variation of the electron temperature is not reproduced by the AEROS-data we restrict ourselves to ISR-data for the first three months of every year, in analogy to the first three months of 1973 which had been used for the AEROS model. As is seen in the last column of Table 1 even with these restrictions excluding annual and diurnal meaning effects there remain enough profiles to given representative mean profiles for the indicated solar activity ranges. An example of our averaging is seen in Fig. 1 showing the increasing variability of the *F*-region electron temperature with increasing height. The observed variability with height is believed to be real and not mainly due to errors in the measurements. This assumption is supported by the small variance during nighttime and the regularity of most of the individual profiles.

Results

The ISR-standard-profiles together with the mean absolute deviations for the three locations and for day and night are given in Figs. 2 and 4. The mean spread of time middle latitude data (Fig. 2) for daytime increases with height and gets as large as 15%. In equatorial latitudes (Fig. 4) the highest data scatter is observed in the height range 200–300 km which is the region of the local temperature maximum. The Arecibo data scatter (Fig. 3) behaves somewhere in between. For all three locations the mean deviation does not exceed 15%.

The described feature shows that the smaller the difference between neutral and electron temperature, and the better the heat contact of the electron gas to the neutrals is, the smaller is the electron temperature scatter. The same

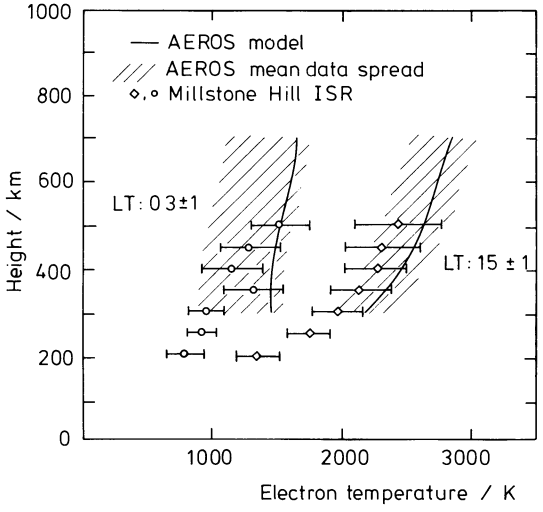


Fig. 2. Comparison between Millstone Hill data and AEROS model

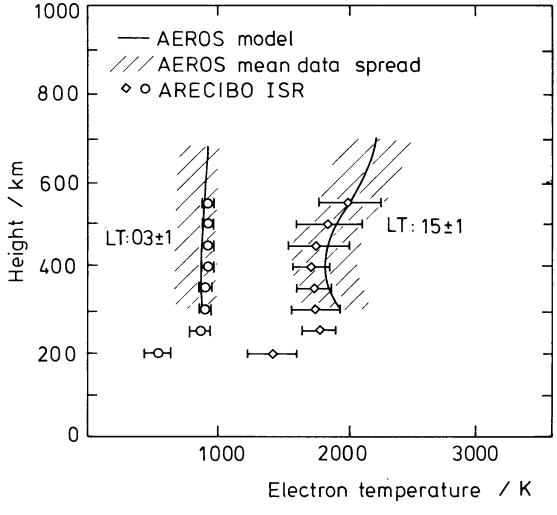


Fig. 3. Comparison between Arecibo data and AEROS model

holds for nighttime, where only the Millstone Hill data show large mean deviations increasing with height.

The solid lines in Figs. 2–4 represent the AEROS-model profiles for day and night in the height range 300–700 km covered by the satellite AEROS-A. The hatched areas indicate the mean deviation of the AEROS-data, as had been described in the foregoing article. For all three stations the model values lie within the expected failure zone. For Millstone Hill there is a general tendency of the model to exceed the ISR mean profile by about 10%. The same feature has been reported by Benson et al. (1977). Their extensive comparison between Millstone Hill measurements and satellite data of AE-C show a general excess of 11% in electron temperature, while the ion temperature agreed well and the discrepancy for other stations was less than 3%.

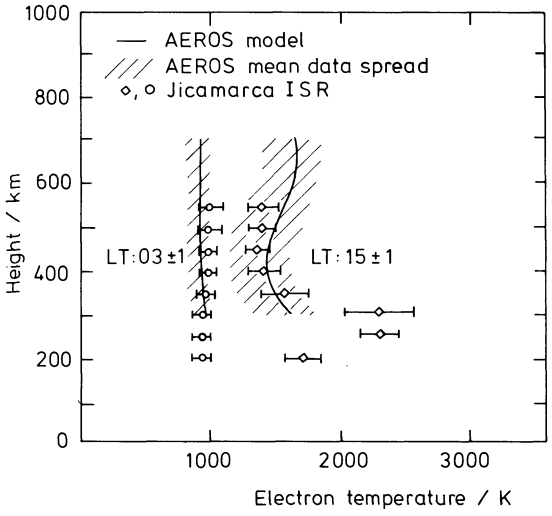


Fig. 4. Comparison between Jicamarca data and AEROS model

In analogy to this, we find good agreement of the AEROS model with the other stations, Arecibo and Jicamarca. Even though the solar activity range is not the same as that covered by the AEROS-A data used for the model, namely $R_{12}=40$–50. As was reviewed by Schunk and Nagy (1978) the electron temperature in the high electron density region 200–400 km seems to become higher with decreasing solar activity. So the few percent higher model temperature at 1500 LT may represent even better the actual temperature profile during the AEROS mission than the ISR data obtained at a higher solar activity (Table 1). The result of the comparisons shown here confirms that the AEROS model temperatures are quite representative within the natural spread generated by different geophysical conditions.

Acknowledgement. The Arecibo and Jicamarca tapes were kindly sent to us by the World Data Center A in Boulder, Colorado/USA. For the Millstone Hill data, we are indepted to W. Oviver who compiled the tape with one of the authors during whose stay at Millstone Hill.

References

Benson, R.F., Bauer, P., Brace, L.H., Carlson, H.C., Hagen, S., Hanson, W.B., Hoegy, W.R., Torr, M.R., Wand, R.H., Wickwar, V.B.: Electron and ion temperatures – a comparison of ground-based incoherent scatter and AE-C satellite measurements. J. Geophys. Res. **82**, 36–42, 1977

Schunk, R.W., Nagy, A.F.: Electron temperatures in the *F*-region of the ionosphere theory and observations. Rev. Geophys. Space Phys. **16**, 355–399, 1978

Spenner, K., Rawer, K.: *F*-region temperatures by AEROS-A satellite compared with incoherent scatter radar. J. Atmos Terr. Phys. **40**, 969–927, 1978

Spenner, K., Plugge, R.: Empirical model of electron temperature distribution between 300 and 700 km based on data of AEROS-A. J. Geophys. **45**, 1979

Received October 31, 1978; Revised Version February 7, 1979; Accepted May 2, 1979

J. Geophys. 46, 63–75, 1979

Journal of Geophysics

Simultaneous Smoothed Variations of Signal Amplitude and Mean Doppler Shift in 42 MHz Auroral Backscatter

C.I. Haldoupis*, and G.J. Sofko

Institute of Space and Atmospheric Studies, Department of Physics, University of Saskatchewan, Saskatoon, Saskatchewan, S7N OWO Canada

Abstract. A recent detailed comparison of the amplitude of 50 MHz radio auroral backscatter at Anchorage with Chatanika incoherent radar data (Siren et al., 1977) has shown a good correlation between the echo amplitude and both the electrojet current and electric field strength. In this paper, smoothed backscatter data (the smoothing limits frequency variations to $\leqq 0.1$ Hz) from the 42 MHz CW Saskatoon Doppler system are used to make a detailed comparison between the amplitude and mean Doppler shift variations during selected portions of the strong April 18/19, 1977 event. A close relationship between these two variations is shown to exist, indicating that a common parameter (either the electrojet current or the electric field) 'modulates' both the mean drift velocity (related to the mean frequency shift) and the scattering cross section (related to the echo amplitude) of the echoing irregularities. However, the amplitude-mean Doppler shift relationship is more complicated than the linear current-amplitude relation in the Chatanika-Anchorage results, and is shown to change with the type of scattering mechanism present.

Key words: Auroral backscatter – Mean Doppler shift – Echo amplitude – Auroral electrojet.

Introduction

When radio waves of constant amplitude and frequency are transmitted towards the auroral ionosphere, often part of the signal is backscattered from the auroral medium if the frequency is in the appropriate range and certain geometric conditions are met (e.g., proper aspect sensitivity). The backscatter signal possesses, in general, both amplitude and frequency variations, because the temporal and spatial fluctuations in the auroral ionization act as 'natural modulators'.

Present address: Physics Department, University of Crete, Iraklion, Crete, Greece

0340-062X/79/0046/0063/$02.60

The variations in signal amplitude are presumably due to changes in both the population and strength of the scattering irregularities, while the fluctuations in frequency are primarily due to motions of the scattering ionization. The 'messages' contained in the amplitude (AM) and frequency (FM) modulation imposed by the scatterers are expected to contain important clues concerning the nature of the radio aurora.

Recent studies (Czechowsky and Lange-Hesse, 1970; Greenwald et al., 1973, 1975; Ecklund et al., 1974; McDiarmid et al., 1976; Siren et al., 1977) have shown that the radio aurora is closely associated with auroral electrojets. The spatial relationship between radio aurora and electrojet currents supports the idea that the scattering irregularities are generated through various current-associated plasma instabilities, and that the electrojet parameters play a major role in determining both the scattering cross-section and the motion of the scatterers.

Greenwald and Ecklund (1975), using a 50 MHz pulsed radar in Anchorage, Alaska, were the first to measure simultaneous range profiles of both the back-scatter amplitude and mean Doppler shift in order to study the temporal and spatial structure of the radar aurora. The temporal variations in both amplitude and frequency of the received signal can also be obtained with a continuous wave (CW) Doppler system. Although a CW system does not give range information, it is useful in providing continuous information over the very short time intervals (5–200 ms) over which individual signal bursts occur (e.g., Haldoupis and Sofko, 1978a). In addition, through the use of smoothing, it is well suited to the study of long term variations of the mean signal characteristics which correspond to the changing average conditions in the total, aspect-sensitive, scattering volume.

In a previous paper (Haldoupis and Sofko, 1978a) digital demodulation techniques were applied to short (0.4 s) time sequences to extract concurrent time variations in amplitude and frequency in order to study the short term characteristics of ion-acoustic type echoes (these are the echoes whose Doppler spectrum exhibits a strong narrow peak at a frequency shift range which corresponds to the ion-acoustic velocity range in the auroral medium) for which the effect of interference between coexisting independent spectral components was not very serious. The purpose of the present investigation is to apply smoothing (only frequencies $\leqq 0.1$ Hz are considered) to auroral data recorded in analogue form from two linear demodulation units (an envelope and a frequency detector) in order to study simultaneous time sequences of mean Doppler shift (or mean signal frequency) and mean signal strength. Comparison of these sequences is found to be quite useful in studying the 'modulation' mechanism responsible for the observed fluctuations in both the signal strength (or scattering cross-section) and the mean line-of-sight motion of the scattering irregularities.

Experimental Details and Method of Analysis

The auroral backscatter data are obtained by means of a CW bistatic radio Doppler system located near Saskatoon, Canada (geographic coordinates: 52° N,

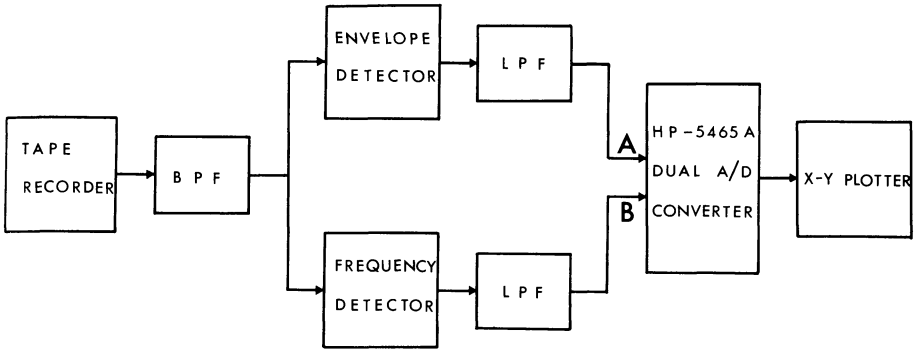


Fig. 1. Block diagram of the system used for studying simultaneous mean Doppler shift and signal strength variations. *BPF* Band-pass filter; *LPF* Low-pass filter

106° W; geomagnetic coordinates: 60° N, 49° W). The transmitter, which generates a CW signal at 42.1 MHz, is separated by 35 km from the receiving site, roughly along the east-west direction. The transmitting and receiving 5-element Yagi antennas are fixed in position pointing northward along the geomagnetic meridian. At the receiving site a weak ground wave received directly from the transmitter is passed through a VHF phase-locked-loop incorporating a stable 1562 Hz audio oscillator, thereby producing a stable reference signal shifted exactly 1,562 Hz from the transmitted wave. The reference signal is mixed with the backscatter to produce an audio output centered at 1,562 Hz. This output is recorded on an audio cassette recorder, and the tapes are analysed with a Hewlett-Packard Fourier Analyser System.

Two analogue demodulation units are employed to extract concurrent time variations in amplitude and frequency (or Doppler shift) from relatively long time intervals of tape-recorded data. Information about the signal strength is obtained with a portable linear envelope detector which is similar to those used for AM demodulation in heterodyne receivers and which consists of a full-wave precision rectifier and a low-pass filter. Continuous information about the instantaneous mean frequency (or mean Doppler shift) of the backscatter is obtained by means of a linear frequency demodulator employing a Signetics 565 phase-locked-loop. Tests of the demodulator performance, by comparison with independent spectral analysis of tape-recorded data, have shown that this simple unit provides a reliable estimate of the mean Doppler shift variations with time. More details about the experimental system and the characteristics of the above demodulation units are given by Haldoupis et al., 1978.

Figure 1 illustrates the technique used for studying concurrent mean Doppler shift and signal strength variations of the recorded echoes. The analogue output of the tape recorder passes through a band-pass filter (which removes any spurious frequencies below 1,000 Hz and above 2,000 Hz) and is then divided into two paths, one of which goes into the envelope detector and the other into the frequency demodulator. Both the frequency and amplitude demodulated outputs are lowpass filtered to attenuate high frequency variations of the incoming signal so that the sampled demodulated sequences are relatively smooth

(the cutoff frequency was 0.1 Hz). The two sequences are sampled simultaneously by the HP-5465A dual A/D converter (on which the number of points, N, and the sampling interval, Δt, can be adjusted), stored in different memory blocks, and then plotted separately using an X-Y plotter. In the following, a few typical examples of relatively long period concurrent variations of mean Doppler shift and relative signal strength will be presented and discussed. All the examples presented below are from the same strong event of April 18–19, 1977. The average K_p index for the times treated in this paper was 6.

Experimental Results

Figure 2 shows an example of simultaneous mean amplitude and Doppler shift time sequences corresponding to $\sim$17 min of continuous signal. A brief inspection shows that there is good correlation between the two records. The long term variations in amplitude are generally accompanied by similar variations in the mean frequency. Usually a sudden change in amplitude occurs simultaneously with a change in the frequency of the signal. The similarity between the mean signal strength and the frequency shift long term fluctuations is a general feature and leads to the important conclusion that the same cause is responsible for the long term modulation of both the amplitude and frequency of the scattered signal, i.e., a common mechanism influences both the velocity and the scattering cross-section of the irregularities. However, the detailed effects of this common modulating mechanism on the strength and frequency of the signal differ and are rather complicated. Comparison of the two sequences indicates that the larger Doppler shifts do not necessarily correspond to the stronger signal. For example, the strongest signal in Fig. 2 is observed in the 900–1,000 s time interval but the largest Doppler shift occurs between 300 and 450 s; also the signal strength in this 300–450 s time interval is as strong as that in the 650–850 s interval but the corresponding frequency shift in the latter is close to zero. Greenwald and Ecklund (1975) reported similar results when comparing simultaneous range profiles of the backscatter amplitude and mean Doppler shift velocity obtained by using a 50 MHz radar pointing towards geomagnetic north.

In an attempt to gain more information about the above 'modulation' process, sequential power spectra (averages of 10 individual 0.41 s spectra sampled evenly over $\sim$55 s intervals of data) were computed by using the fast Fourier transform technique. Samples of the computed averaged spectra are shown in Fig. 3; each is marked with a capital letter and corresponds to a 55 s segment of mean amplitude and frequency shift designated with the same letter in Fig. 2. Examination of the power spectra shows that most of the long ($\sim$17 min) data sequence under discussion is associated with diffuse broad spectra centered close to zero shift position (e.g. Fig. 3A, D and E). However for the interval from about 300–450 s (spectra B and C in Fig. 3) a strong group of spectral peaks, confined to a relatively narrow frequency band of about 40 Hz and centered at about 125 Hz (450 m/s radial speed), dominates the spectrum. The

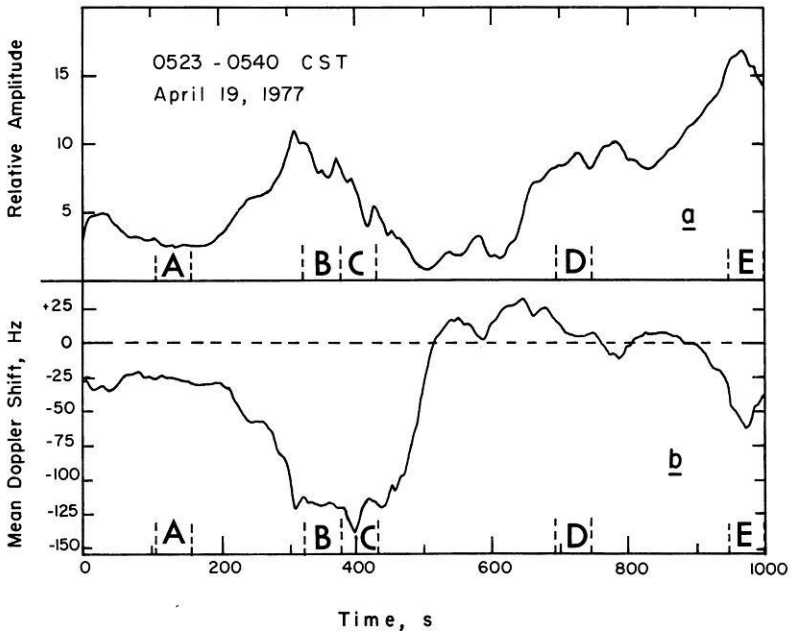


Fig. 2a and b. Concurrent long term variations of the relative amplitude and mean Doppler shift of 42 MHz auroral backscatter. The capital letters *A*, *B*, *C*, *D*, and *E* designate time segments whose power spectra are shown in Fig. 3. Note that Central Standard Time (*CST*) is 6 h behind UT, so the time interval is 1123–1140 UT

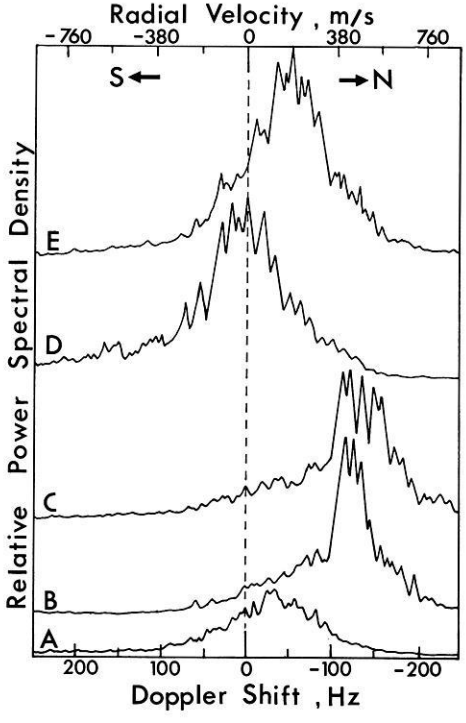


Fig. 3. Samples of Doppler power spectra from the data whose amplitude and Doppler shift are shown in Fig. 2. Each spectrum is marked with a capital letter and corresponds to the time interval designated by the same letter in Fig. 2

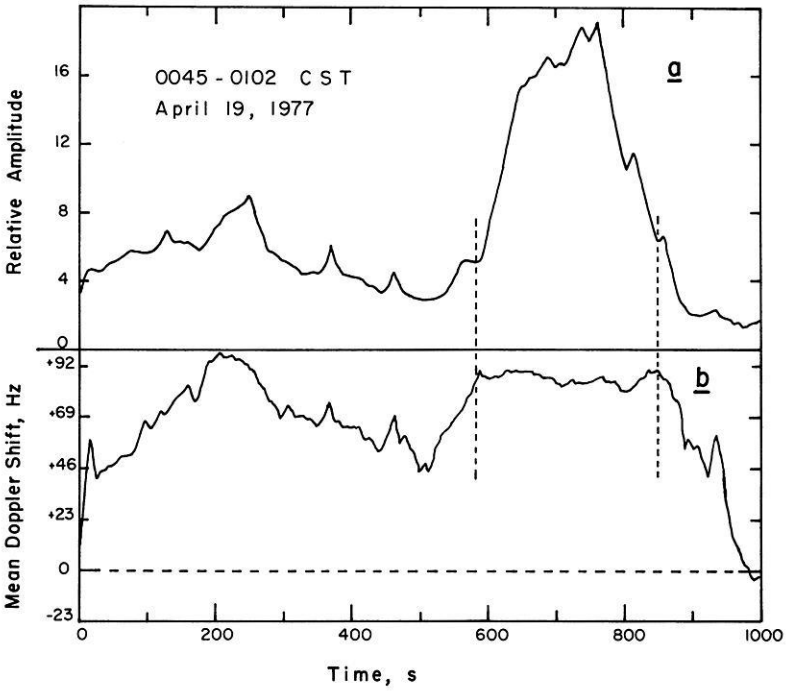


Fig. 4a and b. Plot of simultaneous amplitude and mean Doppler shift variations, showing good correlation from 0–600 s, then 'saturation' of the mean Doppler shift from 600–800 s

latter may well be due to a type of irregularity, such as primary ion acoustic plasma density waves (Haldoupis and Sofko, 1976; 1978b), differing from the irregularities corresponding to the rest of the data sequence. Presumably, ion acoustic plasma wave scatterers cause the almost constant Doppler shift observed in Fig. 2b during the 300–450 s time interval.

Figure 4 shows another example of concurrent amplitude and frequency sequences of 1,000 s duration. The two sequences are in good correlation between 0 and 600 s but this correlation breaks down completely for the time interval (enclosed in dashed lines in Fig. 4) from 600 to 850 s in which the frequency remains fairly constant but the amplitude varies. Detailed fading and spectral analysis of this event (Haldoupis, 1978) show: (1) that the signal during the latter time interval is dominated by strong specular components and exhibits slow and quasi-periodic fading as compared to the rapid fading observed for the former, and (2) the power spectrum associated with the 600–800 s part of the sequence is fairly narrow, with a small number of equally strong peaks grouped together in a narrow frequency band centered at about 85 Hz (300 m/s radial speed), in contrast to the broad type of spectrum corresponding to the initial time interval from 0 to 600 s. The above evidence suggests that a different type of irregularity may be responsible for the scattering in each time interval.

Usually, occurrences of constant frequency shift are observed only for relatively short intervals of time ($\leqq$5 min) associated with the ion-acoustic type

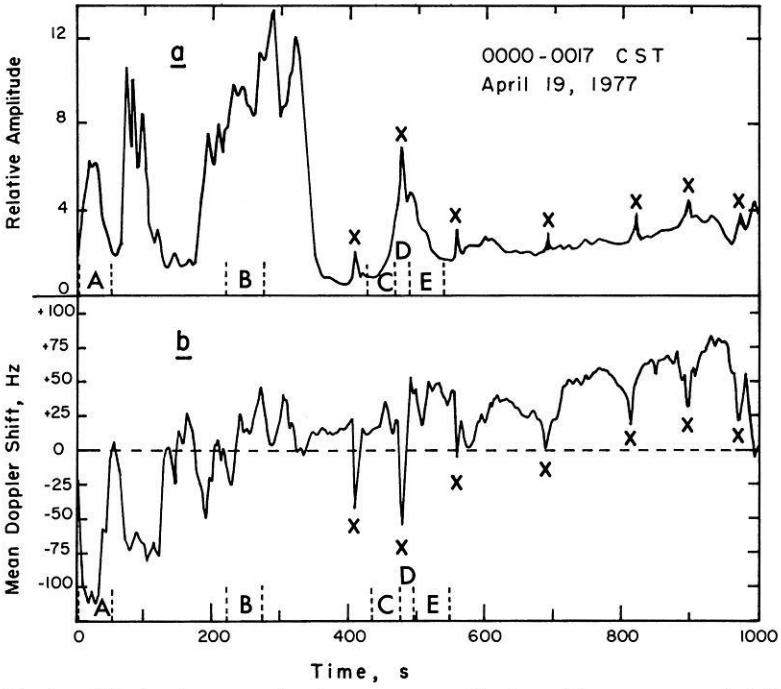


Fig. 5a and b. Another example of concurrent amplitude and frequency records. The peaks marked by 'X' are due to the growth of short-lived bursts of spectral power in the frequency range corresponding to northward-moving ion acoustic waves, as shown in Fig. 6D

of echoes. At such times, the Doppler shift remains almost constant (never exceeding ~140 Hz for the transmitted frequency of 42 MHz) even when the amplitude shows appreciable changes (Haldoupis and Sofko, 1978b). Reports of the occurrence of an almost constant velocity associated with ion-acoustic echoes, which are believed to be caused by longitudinal plasma density waves generated by the two-stream instability mechanism (Farley, 1963), have been made for both the auroral (Hofstee and Forsyth, 1969; Balsley and Ecklund, 1972; Haldoupis and Sofko, 1978b; Moorcroft and Tsunoda, 1978) and the equatorial (Farley and Balsley, 1973) ionospheres. Since this constant velocity is not predicted by the linearized theory, various non-linear theories have been developed (Sato, 1976 and references therein) to explain this saturation effect. The evidence presented in this paper suggests that the mechanism which imposes an upper limit to the velocity of this type of irregularities does not necessarily impose a similar limitation on the scattering cross-section.

A last example is presented in Fig. 5. The first third of the amplitude record consists of three sequential, relatively strong signal bursts of different strength and duration. These backscatter bursts presumably are due to the development of three different irregularity structures, as can be seen by comparing the amplitude and mean frequency variations. For example, the first burst (the weakest and shortest in duration) displays the largest Doppler shift and is due to scattering from irregularities moving northwards with speeds near the ion-acoustic

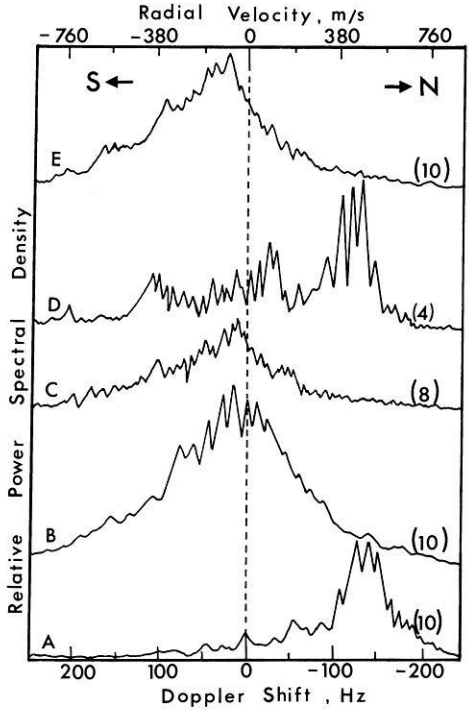


Fig. 6. Five spectra corresponding to the intervals labelled *A* to *E* in Fig. 5. The spectra *C*, *D*, and *E* illustrate the development of the spectral power in the ion-acoustic range, corresponding to the '*X*' peaks in Fig. 5. The bracketed numbers indicate the number of individual spectra averaged

speed in the medium. On the other hand, the third burst (the strongest and longest lived) displays relatively small deviations in frequency about zero shift level. It should be noted that, in all these bursts, sudden changes in amplitude are always accompanied by variations in the mean frequency shift.

The last two-thirds (from 380–1,000 s) of the sequence shown in Fig. 5a would represent a continuous smooth amplitude variation if the superimposed short lived narrow peaks (marked with 'X' in Fig. 5) were not present. Inspection of the corresponding Doppler shift sequence in Fig. 5b shows that the mean frequency shift changes slowly with time (presumably due to a slowly increasing southward drift) except when the short lived echoes dominate the scatter. These echoes are associated with sudden northward relative motions likely caused by a different type of irregularity.

Examination of averaged Doppler spectra associated with different portions of the time sequences in Fig. 5 reveals several different Doppler spectral types. Samples of these Doppler spectra are plotted in Fig. 6; each averaged spectrum is marked with the same capital letter that corresponds to a segment of mean amplitude and frequency shift in Fig. 5. Spectrum A in Fig. 6 shows that the first signal burst in Fig. 5a has a relatively narrow spectrum confined to a frequency shift range which corresponds to the ion-acoustic velocity range in the auroral plasma; on the other hand, the Doppler spectrum B in Fig. 6 associated with part of the strongest signal burst in Fig. 5a is very broad and centered close to zero shift. Comparison of the sequential spectra C, D and

E in Fig. 6 with Fig. 5 reinforces the suggestion, stated in the previous paragraph, that the shortlived peaks (marked with 'X') are caused by a type of scattering irregularity which differs from the one before and after these peaks. Figure 6D indicates clearly that for a short period of time ($\sim$20–30 s) a northward-moving group of irregularities develops with velocities confined to a narrow range ($\sim$350–500 m/s) near the ion-acoustic velocity range in the medium. These irregularities cause the sudden increase observed in the signal strength as well as the abrupt change in the mean Doppler shift; in Fig. 6C and D the broad spectral background centered close to zero shift position likely is caused by a different type of scatterers.

Discussion

The radio aurora is closely related to the magnetic substorms and in particular to the auroral electrojet current system (Tsunoda et al., 1976). Good spatial relationship between the radio aurora and the electrojet currents has been verified by experimental results, based on simultaneous radar scatter and ground magnetograms, obtained by various researchers (Greenwald et al., 1973; Ecklund et al., 1974; McDiarmid et al., 1976). This is in agreement with growing evidence that the auroral electrojet current represents the primary source of energy which sustains, through turbulent mixing and/or plasma instability mechanisms, the auroral scatterers. Studies by Gray and Ecklund (1974), Greenwald et al. (1975), and Czechowsky and Lange-Hesse (1970) have shown that at times the scatter amplitude is linearly related to the electrojet current intensity as deduced from ground-based magnetometers. A rough comparison between the chart-recorded signal strength records of our backscatter system and Saskatoon magnetograms shows that the radio backscatter is always accompanied by geomagnetic activity and that usually, but not always, there is a direct relationship between the strength and duration of the received echoes and those of the magnetic perturbations.

The most decisive recent experiment relating radio aurora to ionospheric parameters was the comparison by Siren et al. (1977) of the 50 MHz Anchorage coherent backscatter amplitude with electric fields and height-integrated current flow inferred from Chatanika incoherent scatter measurements of electron density and ion drift. The results indicate a linear relationship between the electrojet current magnitude and backscatter amplitude, the slope remaining constant whether the current is eastward or westward. A linear relationship was also found to hold between the north-south electric field strength and the echo amplitude, although threshold values of 25 mV/m northward and 10 mV/m southward had to be exceeded before the onset of instabilities in the eastward and westward electrojets respectively. Assuming that these results apply also to the 42 MHz Saskatoon Doppler system (which receives echoes from the L-value range 5.6–6.5 whereas Chatanika is at $L \sim 5.7$), the present results indicating a close relationship between amplitude and mean Doppler shift imply that the mean drift motion of the scatterers is determined by the electrojet current and/or the electric field.

At the present time it is believed that most of the radio auroral irregularities are generated by plasma instability mechanisms (basically the two-stream and the gradient-drift instabilities) which are current-associated since they require a certain relative electron-ion drift to exist in order for the plasma density waves to develop. The existing instability theories predict that the generated primary plasma waves propagate in the direction of the electron drift with a phase velocity close to the electron drift velocity, which itself is associated with the existing current system presumably present in the scattering region. Sudan et al. (1973a) used a nonlinear approach to show that when primary plasma wave irregularities (predicted by linear theory) are generated along the main electron drift, smaller scale secondary plasma waves may become unstable and propagate in directions centered about the perpendicular to the primary plasma waves. The phase velocity of the secondary waves depends again on the electron-ion drift velocity and the amplitude $A(A=\Delta N/N)$ of the primary wave. The plasma wave scatterers are strongly anisotropic spatially periodic structures of ionization, and cause coherent weak scattering of the incident radio waves if the plasma waves propagating along the radio wave line-of-sight have a wavelength equal to half the radio wavelength (Flood, 1967). Generally, for the case of plasma wave scatterers, it is believed that the relative electron-ion drift velocity is an important factor in determining not only the Doppler shift in frequency but also the strength of the scatterers as well (Farley, 1963; Unwin and Knox, 1971).

Very clear evidence of ion-acoustic waves has been presented in this paper. The results of Figs. 2 and 3 in the 300–500 s period ($\sim$0530 CST) labelled 'B' and 'C', and the 0–55 s interval labelled 'A' in Figs. 5 and 6 ($\sim$0000 CST) clearly show evidence of ion-acoustic echoes associated with northward motions, while the 600–800 s interval in Fig. 4 ($\sim$0056 CST) shows an ion-acoustic event associated with southward motion. The latter case is interesting since it suggests that the mean Doppler shift exhibits saturation but the amplitude does not. This agrees well with the observations by Siren et al. (1977) that no saturation is observed in echo amplitude with current or electric field strength, whereas many measurements (e.g., Hofstee and Forsyth, 1969; Haldoupis and Sofko, 1978b) have shown that the ion-acoustic velocity forms a limit to the two-stream instability echoes observed. The development of a brief northward moving ion-acoustic scatterer group in conjunction with a longer-lived broad spectrum is illustrated in the sections of Figs. 5 and 6 labelled C, D, and E.

Finally we mention briefly that some of the results presented suggest the possibility of irregularities other than those having a well defined wavelike structure (e.g., a random assembly of field-aligned short-lived blobs of ionization). In particular, the broad spectra of Fig. 3 associated with northward mean motions in the intervals A, D, and E in Fig. 2, and the broad spectra of Fig. 6 associated with southward mean motions in the intervals C, D, and E in Fig. 5, would seem to fall into this category. The possibility exists that a turbulent electrojet subject to inhomogeneous particle precipitation and rapidly changing electric fields may play an important role in generating and sustaining such an assembly of irregular scatterers. According to Sato (1968) the irregularities

in a highly turbulent electrojet cannot be recognized as having the form of plasma waves. For the above case a theoretical relationship between the electrojet intensity and the mean frequency and amplitude of the backscattered signal is not known (if there is such a relationship, it must be rather complicated). According to Holtet (1973), the mean drift velocity (mean Doppler shift) of an assembly of weak irregularities is determined mainly by the existing large scale DC electric field present; this mean drift is in the same direction as the electrojet current under the assumption that the Hall conductivity dominates over the Pedersen conductivity (which is a well observed fact for the auroral regions, particularly for the westward electrojet, e.g., Kamide and Brekke, 1977; Horwitz et al., 1978). Egeland (1973), in explaining the variation of mean Doppler shift with azimuth angle (for a steerable antenna) or time of occurrence (for a fixed antenna), suggested that the echoes are the result of irregular ionization blobs drifting along the electrojet so that the mean Doppler shift would be the drift velocity of the irregularities times the cosine of the angle between the radio beam and the direction of the electrojet. If this is the case the direction of the electrojet relative to the radio beam is very important in determining the observed Doppler shift of the received echoes.

The electrojet dimensions (or spatial extent) may well have a direct effect on the population of scatterers and the degree of spatial uniformity in their motion. Also, the location of the electrojet with respect to the radio beam and the earth's magnetic field lines is very important for the strength of the radio scatter since it may affect drastically the aspect sensitivity of the echoes. For example, it is a common observation of our system that the early evening period is characterized by few, if any, echoes and that these echoes are weak. Maximum echo occurrence frequency takes place in the early morning hours when also the strongest signals are received. This may well be because of the position of the electrojet with respect to the optimum aspect sensitivity location for our system. At a scatterer height of 110 km, the aspect angle varies from about 97° to 93.5° as the echo range goes from 400 to 700 km. At a scatterer height of 90 km, however, the aspect angle is somewhat improved, going from 94° to 92° over the range 400 to 750 km. A characteristic difference in ionospheric altitude between the eastward electrojet in the evening sector and the westward electrojet in the morning sector was measured by Kamide and Brekke (1977); using the Chatanika incoherent scatter facility, they found that the eastward electrojet in the evening sector and the predominantly westward current system in the morning sector are located at approximately 120 and 100 km respectively, with the latter system showing much more variability. Thus, morning echoes from lower heights are optimal for our system if the scatterers are aspect sensitive.

According to the above discussion one may argue that the auroral electrojet plays a basic role in the creation, strength and the motion of the scattering irregularities. Presumably changes in the intensity of the current system (mainly due to the fluctuations in electron density and/or local electric fields) and in its dimensions and location cause the modulation observed in the mean amplitude and frequency of the radio echoes. In addition to the electrojet, the type

of scatterer (e.g. a random assembly of numerous field aligned inhomogeneities or spatially periodic plasma wave scatterers) has an important effect on the amplitude-frequency relationship.

Conclusions

By using two linear demodulation units (an envelope and a frequency detector), simultaneous long period variations of the mean amplitude and frequency of the scattered signal have been investigated. The experimental evidence suggests that a common mechanism modulates both the average velocity (related to the mean Doppler shift) and the scattering cross-section (related to the signal strength) of the scattering irregularities. Although it was found that sudden changes in the signal strength are accompanied by well-defined changes in the mean Doppler shift, the overall relationship between the two variations is not a straightforward one. It was shown that the nature of the irregularities (i.e., the type of scattering mechanism) plays an important role in dictating the observed amplitude and frequency variations. For example, the evidence suggested that occasionally a physical process operates in the auroral plasma to impose an upper limit to the speed (and therefore the Doppler shift) of certain scatterers without limiting their scattering cross-section. Examination of concurrent averaged Doppler spectra indicated that this situation is observed usually with narrow ion-acoustic-type spectra. By assuming that the results of Siren et al. (1977) which show a linear relationship between 50 MHz backscatter amplitude and both the electrojet current and electric field (beyond the threshold values) over Chatanika ($L \sim 5.7$) are applicable to the Saskatoon 42 MHz Doppler system (echoes from $L = 5.6$ to 6.5), the close relationship between amplitude and mean Doppler shift for the (42 MHz) backscatter implies the common mechanism which 'modulates' both these quantities is either the electrojet current or the electric field or both. It was argued that changes in the strength, direction, dimensions and location of the flowing electrojet current system could be causing the long term variations observed in the amplitude and frequency of the auroral scatter.

Acknowledgements. The present research work was supported by the National Research Council of Canada.

References

Balsley, B.B., Ecklund, W.L.: VHF power spectra of the radar aurora. J. Geophys. Res. **77**, 4746–4760, 1972

Czechowsky, P., Lange-Hesse, G.: Substorm influence on VHF continuous wave auroral backscatter. In: Proc. of the 3rd ESLAB/ESRIN Symposium on Intercorrelated Satellite Observations Related to Solar Events, V. Manno and D.E. Page, eds.: pp. 268–275. Dordrecht, Holland: D. Reidel 1970

Ecklund, W.L., Carter, D.A., Keys, J.G., Unwin, R.S.: Conjugate auroral radar observations of a substorm. J. Geophys. Res. **79**, 3211–3213, 1974

Egeland, A.: Scattering of radio waves from auroral ionization. In: Cosmical Geophysics, A. Egeland, O. Holter, A. Omholt, eds.: pp. 311–325. Oslo: Universitetsforlaget 1973

Farley, D.T.: A plasma instability resulting in field-aligned irregularities in the ionosphere. J. Geophys. Res. **68**, 6083–6097, 1963

Farley, D.T., Balsley, B.B.: Instabilities in the equatorial electrojet. J. Geophys. Res. **78**, 227–239, 1973

Flood, W.A.: A brief review of auroral backscatter theory. In: Aurora and Airglow, B.M. McCormac, ed.: pp. 563–571. New York: Reinhold 1967

Gray, A.M., Ecklund, W.L.: The Anchorage, Alaska real-time auroral radar monitor: System description and some preliminary analyses. Tech. Rep. ERL 306-AL9. Boulder, Colo.: National Oceanographic and Atmospheric Administration, pp. 1–38

Greenwald, R.A., Ecklund, W.L.: A new look at radar auroral motions. J. Geophys. Res. **80**, 3642–3648, 1975

Greenwald, R.A., Ecklund, W.L., Balsley, B.B.: Auroral currents, irregularities, and luminosity. J. Geophys. Res. **78**, 8193–8203, 1973

Greenwald, R.A., Ecklund, W.L., Balsley, B.B.: Radar observations of auroral electrojet currents. J. Geophys. Res. **80**, 3635–3641, 1975

Haldoupis, C.: VHF Doppler investigations of the auroral ionosphere. PhD. Thesis, Univ. of Saskatchewan, Saskatoon, Canada. 1978

Haldoupis, C., Koehler, J.A., Sofko, G.J.: A VHF bistatic Doppler system for radio investigations of the auroral ionosphere. J. Phys. E. Sci. Instrum. **11**, 135–138, 1978

Haldoupis, C., Sofko, G.: Doppler spectrum of 42 MHz CW auroral backscatter. Can. J. Phys. **54**, 1571–1584, 1976

Haldoupis, C., Sofko, G.: Short term characteristics of ion-acoustic type radio auroral echoes. Can. J. Phys. **56**, 292–301, 1978a

Haldoupis, C., Sofko, G.: On ion-acoustic plasma waves in radio aurora and the two-stream instability. Can. J. Phys. **56**, 100–116, 1978b

Hofstee, J., Forsyth, P.A.: Ion-acoustic waves in the auroral plasma. Can. J. Phys. **47**, 2797–2803, 1969

Holtet, J.A.: Electric field microstructures in the auroral E-region. Geofys. Norv. **30**, No. 4, 1–88, 1973

Horwitz, J.L., Doupnik, J.R., Banks, P.M.: Chatanika radar observations of the latitudinal distributions of auroral zone electric fields, conductivities, and currents. J. Geophys. Res. **83**, 1463–1481, 1978

Kamide, Y.R., Brekke, A.: Altitude of eastward and westward auroral electrojet. J. Geophys. Res. **82**, 2851–2853, 1977

McDiarmid, D.R., Harris, F.R., McNamara, A.G.: Relationships between radio aurora, visual aurora and ionospheric currents during a sequence of magnetospheric substorms. Planet. Space Sci. **24**, 717–725, 1976

Moorcroft, D.R., Tsunoda, R.T.: Rapid-scan Doppler velocity maps of the UHF diffuse radar aurora. J. Geophys. Res. **83**, 1482–1492, 1978

Sato, T.: Brillouin scattering in plasmas as applied to ionospheric irregularities. J. Geophys. Res. **73**, 2941–2949, 1968

Sato, T.: On mechanisms governing the electrojet plasma instabilities. J. Geophys. Res. **81**, 539–546, 1976

Siren, J.C., Doupnik, J.R., Ecklund, W.L.: A comparison of auroral currents measured by the Chatanika radar with 50-MHz backscatter observed from Anchorage. J. Geophys. Res. **82**, 3577–3584, 1977

Sudan, R.N., Akinrimisi, J., Farley, D.T.: Generation of small scale irregularities in the equatorial electrojet. J. Geophys. Res. **78**, 240–248, 1973

Tsunoda, R.T., Presnell, R.I., Kamide, Y., Akasofu, S.I.: Relationship of radar aurora, visual aurora, and auroral electrojets in the evening sector. J. Geophys. Res. **81**, 6005–6015, 1976

Unwin, R.S., Knox, F.B.: Radio aurora and electric fields. Radio Sci. **6**, 1061–1077, 1971

Received July 3, 1978; Accepted June 5, 1979

J. Geophys. 46, 77–87, 1979

Journal of Geophysics

Overnight Statistical Variation of the North-South Movements of Radio Auroral Irregularities

C.I. Haldoupis* and G.J. Sofko

Institute of Space and Atmospheric Studies, Department of Physics, University of Saskatchewan, Saskatoon, Saskatchewan, S7N OWO, Canada

Abstract. The average overnight variation of the north-south mean movements of the auroral scatterers at 42 MHz has been determined by using mean Doppler shift chart recordings from 24 radio auroral events of different strength and duration. On the average, the scattering irregularities move equatorward before local magnetic midnight and poleward in the morning sector (after about 0200 CST). During the transition period when the motion reversal occurs, the mean Doppler shift values are less stable and can vary rapidly between positive and negative. For individual events the exact overnight variation and the time of motion reversal vary, possibly because of differences between events in the level of geomagnetic activity. It is argued that the overnight east-west component of the electric field variations in conjunction with both the meridional component of the electrojet current systems, in the evening and morning sectors, and in particular the Harang discontinuity, up until midnight, play the deciding role in determining the observed north-south movements of the irregularities.

Key words: Doppler shifts – Radio auroral motions – Auroral electrojets – Hall currents – Scatterers.

Introduction

Recent experimental studies using VHF backscatter radars and ground magnetograms from a number of stations beneath the scattering region show that there is good spatial correlation between the radio aurora and the auroral electrojets (Ecklund et al., 1977; Greenwald et al., 1973; McDiarmid et al., 1976). This observation is in agreement with growing evidence that the auroral irregularities which cause the scattering are closely associated with the auroral electrojets through either the plasma instabilities or turbulent mixing or particle precipita-

* *Present address*: Physics Department, University of Crete, Iraklion, Crete, Greece

0340-062X/79/0046/0077/$02.20

tion occurring in conjunction with the electrojets. According to Greenwald et al., 1975b, the irregularities causing the backscatter appear to be acting as tracers of the flowing currents. Other observations suggest that the radio aurora is closely associated with ambient DC electric fields which are believed to be of magnetospheric origin (Unwin and Knox, 1971; Tsunoda, 1975; Tsunoda and Presnell, 1976). The ambient DC electric field vector seems to be the main parameter which dictates both the particle velocity associated with the electrojets and the motion of the auroral irregularities (Holtet, 1973; Czechowsky, 1974; Ecklund et al., 1977).

A Doppler radio-wave system can measure the line-of-sight movements of the scattering irregularities and therefore may be used to deduce useful information about the auroral zone currents and ionospheric electric fields. A pulsed radar is better suited to the task than a continuous wave (CW) radio system, since the pulses provide information about both the echo intensity and the Doppler shift variations as a function of range. A CW Doppler system gives a Doppler spectrum which represents the average motions over the total, aspect sensitive, scattering volume.

During the last 20 years there have been several Doppler investigations of the radio auroral ionosphere which yielded considerable, although sometimes contradictory, information about the motions as well as the scattering mechanisms (for references before 1975 see Greenwald et al., 1975a; later references are: Tsunoda, 1975; Haldoupis and Sofko, 1976; Moorcroft and Tsunoda, 1978; Greenwald et al., 1978). Most of these studies were based on a limited number of events and limited sections of the auroral time sector. In the present paper, the systematic overnight behaviour of the drift motion of the auroral irregularities along the north-south direction is presented. According to Tsunoda and Fremouw (1976), poleward and equatorward movements are the most common characteristic observed in radar aurora. The results presented here are based on data from 24 auroral events of different strength and duration.

Experimental Details and Results

The observing system is a VHF bistatic radio Doppler system located near Saskatoon, Canada (geographic coordinates: 52° N, 106.5° W; geomagnetic coordinates: 60° N, 49° W). The transmitter, which generates a CW signal at 42.1 MHz, is separated by 35 km from the receiving site, roughly along the east-west direction. The transmitting and receiving 5-element Yagi antennas are fixed in position pointing northward along the geomagnetic meridian. Because of the low transmitted power ($\sim$40 W), the scatter observation period is limited to the period (usually from 2100 to 0800 CST or 0300 to 1400 UT) when the strongest radio auroral conditions are present. At the receiving site a weak ground wave received directly from the transmitter is passed through a VHF phase-locked-loop incorporating a stable 1562 Hz audio oscillator, thereby producing a stable reference signal shifted exactly 1,562 Hz from the transmitted wave. The reference signal is mixed with the backscatter to produce an audio output centered at 1,562 Hz. This output is recorded on an analog magnetic

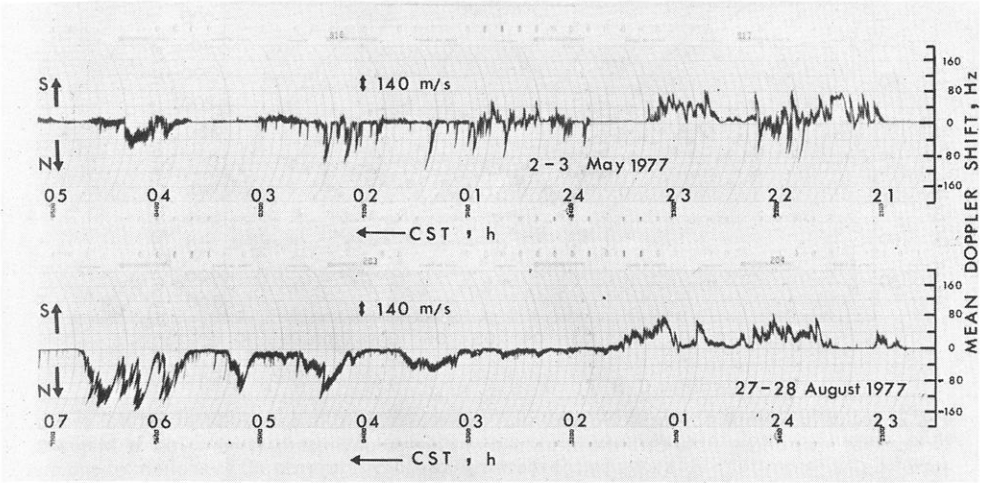


Fig. 1. Two typical examples of continuous chart recordings of mean Doppler shift variations of 42 MHz auroral backscatter due to the north-south movements of the scatterers. The horizontal axis represents local time (American Central Standard Time, CST)

tapedeck, and the tapes are analyzed using a Hewlett Packard Fourier Analyzer System.

The data used in this study was obtained by feeding the system output through a simple analog device, a frequency demodulator, to an Esterline Angus pen-recorder. This arrangement provided information on the mean Doppler shifts (or alternatively the mean radial motions) on a continuous 24-h basis. Tests of the demodulator performance, by comparison with independent spectral analysis of the simultaneously recorded tape data, have shown that this demodulator unit can provide a reliable estimate of the mean Doppler shift variations with time. More details about the experimental system and the characteristics of the frequency demodulator are given by Haldoupis et al. (1978).

The results presented in this paper are based on mean Doppler shift chart-recordings from 24 backscatter events of different strength and duration which occurred from April 15 to August 30, 1977. Two typical examples of mean Doppler shift pen-recordings corresponding to two different events are shown in Fig. 1. Since both the transmitting and receiving antennas were pointing towards geomagnetic north, the observed mean Doppler shifts in frequency reflect, approximately, the equatorward and poleward movements of the auroral ionization (notice that positive shifts in Fig. 1 are caused by southward motions). In sampling the records, a relatively large sampling interval, $\Delta t = 5$ min, was used throughout all recorded events. As a result, the 'fine structure' seen in the original pen-recordings (Fig. 1) and the strong short-lived spikes, usually due to the ion-acoustic type of echoes (Haldoupis and Sofko, 1978) lasting for about 1 to 3 min, are smoothed out in the results presented in Fig. 2.

Figure 2 illustrates the overnight variation of the mean frequency shift for four different events (the event of May 1–2, 1977 lasted for about 15 h and was the longest recorded in the present work). It can be seen from Fig. 2

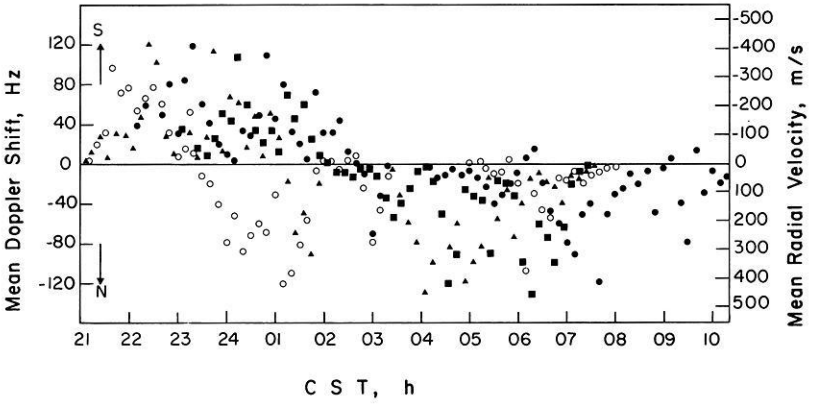


Fig. 2. Sequential samples of mean Doppler shift penrecordings from four different auroral events. Notice that after about 0200 CST the mean motions are predominantly northwards. ● May 1–2, 1977; ○ July 28–29, 1977; ▲ August 26–27, 1977; ■ August 28–29, 1977

that there is general agreement in the pattern of the overnight mean Doppler shift for all these events. It is clear that positive mean shifts (equatorward motions) are generally present before local midnight while negative shifts dominate the morning sector of the events. In addition to these two distinct periods, there is also a transition period (which can occur anywhere from about 2330 to 0330 CST) when the reversal of motion occurs. The exact time of reversal depends upon the particular auroral event (for example the reversal occurred at ~2330 CST for the July 28–29 event and at ~0230 CST for the May 1–2 event), but usually occurs between 0100 and 0300 CST. During the above transition period, the mean Doppler shift values are less stable and sometimes vary rapidly between positive and negative values. It has also been observed, by comparison with simultaneous signal strength chart recordings, that the echoes are weak, or disappear completely, during the transition of the mean Doppler shift from positive to negative values.

The morning sectors of the events are associated with radio echoes having, on the average, the most stable mean dynamic characteristics (e.g., mean Doppler shift and spectral width). Also it is during the morning period that the largest mean Doppler shifts (corresponding to about 700 m/s northward motions) are occasionally seen. From spectral analysis of the tape-recorded data (Haldoupis and Sofko, 1976; Haldoupis, 1978), it has been found that the spectral width is greater in the post-midnight period than before and around midnight. This is illustrated in Fig. 3 where samples of Doppler spectra received during the night of 4–5 August, 1975 are plotted in time sequence. From the increase in spectral breadth as the event proceeds in time, it can be inferred that the population of moving scatterers and the randomness in their motion increases especially after about local magnetic midnight (0200 CST). These observations, which agree with the result of Moorcroft and Tsunoda (1978), show that the parameters and dynamic characteristics of the auroral plasma are different in the morning sector than in the pre-midnight observation periods.

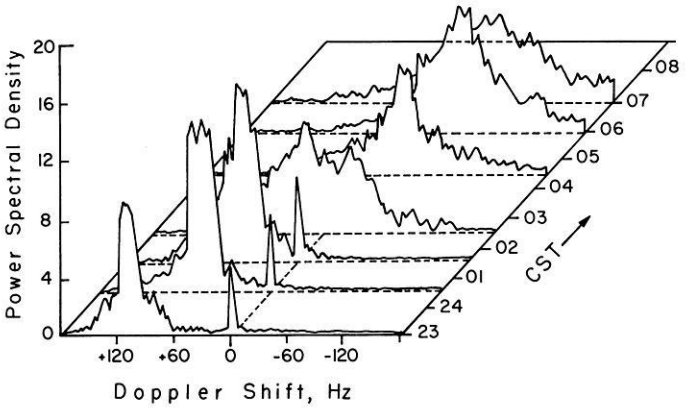


Fig. 3. A three dimensional plot (Doppler shift, local time, relative power spectral density) showing examples of typical auroral spectra from the night of 4–5 August 1975. Positive Doppler shifts are due to motions of the auroral irregularities towards the equator (1 Hz corresponds to 3.5 m/s). Notice that the character of the spectra changes drastically after about local magnetic midnight (~0200 CST)

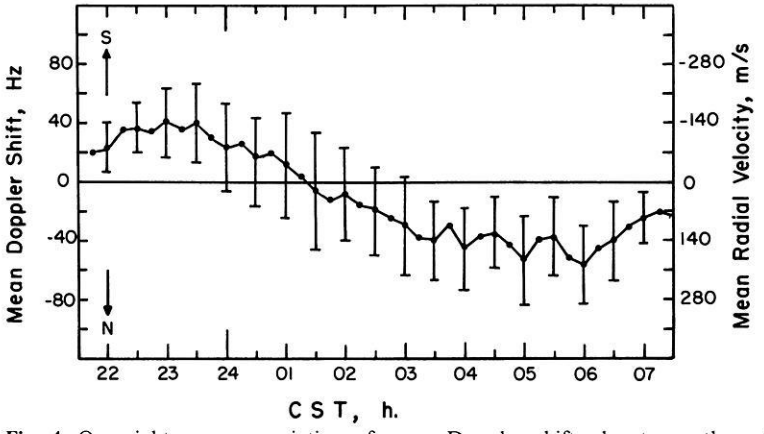


Fig. 4. Overnight average variation of mean Doppler shifts due to north-south movements of the auroral ionization. The bars denote the spread in the averaged values. This curve was based on 24 events of different strength and duration

The results of comprehensive statistical treatment of the data are presented in Figs. 4 and 5. Figure 4 illustrates the average overnight variation of the mean Doppler shifts based on data from all the events examined. The bars represent the standard deviation of the mean frequency shifts. This figure suggests that the time interval when the reversal of motion occurs extends from 00 to 03 h CST. Another interesting feature in Fig. 4 is the exact time when the reversal of the mean Doppler shifts occurs on the average; this is between 01 and 02 CST, close to the local magnetic midnight (~0140 CST at Saskatoon). Other points to notice in Fig. 4 are: (1) the relative spread is greater during the transition period (which means that the uncertainty of the mean Doppler

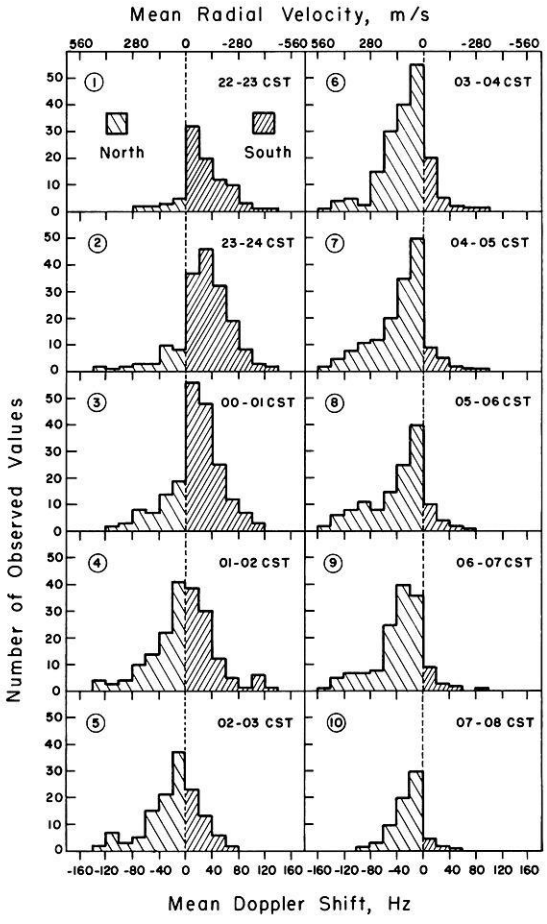


Fig. 5. Histograms (based on data from 24 events) showing the distribution of the mean Doppler shift in sequential one-hour intervals

shift position is larger), and (2) the mean frequency shifts are generally larger during the morning than the premidnight sector.

Figure 5 illustrates in a different form the statistical behavior of the mean frequency shift (or mean north-south motion of irregularities) for sequential one hour intervals of observation time. The radio scatter period (from 2200 to 0800 CST) is divided in 10 equal one-hour intervals and the mean Doppler shifts observed during these hour intervals, for all the events, are plotted in histogram form. Figure 5 shows clearly the increasing preference for negative shifts (northward motions) as the event proceeds in time from the premidnight to postmidnight sector. As can be seen from the histograms numbered 4 and 5, positive and negative shifts are equally probable during the transition time interval between 01 and 03 CST.

Interpretation and Discussion

The observed overnight variation of mean Doppler shifts due to the north-south motions of the auroral ionization cannot be explained from the present results

alone. Simultaneous information about the auroral electrojet currents (which can be inferred from magnetometer records at auroral latitudes), the electric fields, as well as the range distribution of the received records is required. Unfortunately in the present experiment none of these are available. For this reason the following discussion relies on observations, reported by other workers, which demonstrate a close relationship between radio aurora, auroral electrojets and electric fields.

If we accept the premise that the auroral irregularities which cause the scattering of VHF waves are weak scatterers (associated with electron density fluctuations, $\Delta N/N$, of a few percent), then the theory developed by Kato (1963) about the drift motion of an assembly of weak irregularities may apply. According to Kato, the mean propagation velocity of weak electron density irregularities in the ionospheric plasma, where an electric field $\mathbf{E}$ is present perpendicular to the earth's magnetic field $\mathbf{B}$, is given by

$$\mathbf{V} \simeq \frac{1}{1 + \nu_{en}\nu_{in}/\omega_{B_e}\omega_{B_i}} \frac{\mathbf{E} \times \mathbf{B}}{B^2}, \tag{1}$$

where ν_{en}, ν_{in} are the electron and ion collision frequencies and, ω_{Be}, ω_{Bi} are the electron and ion gyrofrequencies. Thus, the mean drift motion observed in the present results could be an $\mathbf{E} \times \mathbf{B}$ Hall drift motion due to an electric field in the east-west direction (presumably the antenna orientation does not allow the observation of east-west motions). It is worth mentioning that $\mathbf{V}$ in Eq. (1) is the same as the phase velocity of plasma wave irregularities (two-stream or gradient-drift) predicted by the linearized theory [Sudan et al., 1973, their Eq. (12)], under the assumption that the electron-ion drift velocity $\mathbf{V}_d$ is equal to $\mathbf{E} \times \mathbf{B}/B^2$ Hall motion.

Recent results reported by Ecklund et al. (1977) (based on simultaneous observations of plasma drifts and associated electric fields) support strongly the $\mathbf{E} \times \mathbf{B}$ motion of auroral irregularities along the electron drift motion. Czechowsky (1974) has explained his overnight Doppler shift observations along the north-south direction, made in northern Germany and Scandinavia, by considering $\mathbf{E} \times \mathbf{B}$ drift motion of field-aligned electron inhomogeneities; he also concluded that the electric field is the most important factor in producing the irregularity drifts observed in auroral regions.

According to the above line of reasoning, the variations of the observed mean Doppler shifts should follow closely any variations of the east-west component of the electric field. In other words, the southward motions seen before magnetic midnight are due to a relatively strong westward component of the electric field. This field component weakens as the event proceeds in time and the mean Doppler shift decreases. Eventually the westward component reverses into an eastward oriented field responsible for the observed northward motions in the morning sector. Presumably, the exact variation of the field (strength and time of reversal) depends upon the magnetic activity for that particular auroral event and geomagnetic location. It should be mentioned that the term 'electric fields' is taken to mean large scale DC auroral electric fields presumably of magnetospheric origin (Holtet, 1973).

The above explanation, based on the $\mathbf{E} \times \mathbf{B}$ irregularity motion, is supported by the experimental results of Mozer and Lucht (1974). They have found average diurnal variations of auroral electric fields by using electric field data obtained by the flight of 32 balloons at auroral latitude locations in western Canada and Alaska. The overnight behavior of the east-west component of the electric field in heir results follows a consistent variation, being westward before midnight and changing into eastward later in the morning. It was also found that the strength increases as the geomagnetic activity increases.

Another observation which seems to be closely related to the existence of an eastward electric field at auroral zone latitudes in the morning sector was made by Rostoker and Hron (1975). By using a meridian line of magnetometers they have pointed out that in the dawn sector there exists an equatorward electrojet causing positive bias in the D component of the magnetometer due to northward electron motion. Hughes and Rostoker (1977) have investigated further this equatorward Hall current in the morning sector and suggested that it flows through the lower latitude ionosphere and into the postnoon auroral oval where it diverges up the field lines. The fact that the average eastward electric field (observed by Mozer and Lucht, 1974), the equatorward-flowing Hall current (observed by Rostoker and Hron, 1975) and the average poleward drift motions observed in the present results occur about the same time (morning sector) suggests strongly an interrelation between these phenomena.

Kamide et al. [1976a) have shown, by using the incoherent scatter radar in Chatanika, Alaska, that the auroral current system has a northward component in the evening sector and a southward component in the morning sector. Zmuda and Armstrong (1974) have established that the electrojets are associated with a pair of field-aligned sheet currents. In the evening, the flow is into the ionosphere on the equatorward side of the auroral oval and out of the ionosphere on the poleward side; the current directions are reversed in the morning pair, while either pair can appear in the transitional sector around midnight. Also Kamide et al. (1976b) have observed, by using ground and satellite data, that the horizontal auroral electrojet in the morning sector flows southwestward so that the 'westward electrojet' (a common term used for the auroral current regime in the morning sector) has in reality a large southward component. Now, taking into consideration that: (1) the electron drift motion is nearly opposite to the electrojet direction, assuming the Hall conductivity dominates, and (2) that the motion of the scattering irregularities as observed from the ground is the ionospheric Hall electron drift (Ecklund et al., 1977), it is clear that the mean motions presented in this paper are consistent with the above observations of the meridional motion components at the electrojet current system. This suggests that the electrojets play an important role in generating the scattering irregularities and controlling their motion. This is consistent with the basic assumptions of the various plasma instability mechanisms which have been proposed for the auroral plasma (Greenwald et al., 1975b).

A strong possibility exists that the premagnetic midnight echoes associated with southward motions originate from the Harang discontinuity (HD). This is the dynamic transition boundary between the eastward electrojet in the evening

sector and the westward electrojet in the morning sector and may occur from 2000 MLT to magnetic midnight (Heppner, 1972) and may in fact 'jump" in longitude during a substorm (Horwitz et al., 1978). The HD, which is a highly turbulent region associated with a southward convective flow (Chen and Rostoker, 1974) may play an important role in the generation of southward moving scatterers. Another important observation which supports southward motions in this region is the westward orientation of the electric field (Maynard, 1974). Kamide et al. (1976b) suggest that the Hall conductivity is larger than the Pedersen conductivity at the HD, (as is true for the westward electrojet in general (Kamide and Brekke, 1977)) so that the northward ionospheric current (due to $\mathbf{E} \times \mathbf{B}$ southward electron drifts) prevails in the vicinity of the HD.

From the above, under the basic assumption that the dominant ionization movements are $\mathbf{E} \times \mathbf{B}$ Hall motions, it is clear that mean Doppler shift measurements of north-south auroral motions are useful in estimating the mean east-west component of the auroral electric field. Since the effect of neutral winds is negligible (Sudan et al., 1973) compared to the electric field motions, a simple equation for estimating the east-west mean electric field is given by

$$E_{my} \simeq 5.56 \times 10^{-2} V_{mx}, \text{ (mV/m)}, \tag{2}$$

where positive x and y are along the geomagnetic north and west respectively and V_{mx} is the north-south mean Doppler velocity [Eq. (2) is obtained from Eq. (1) by substituting $|B| \simeq 5.5 \times 10^4$ Wb/m^2 and $\nu_{en}\, \nu_{in}/\omega_{Be}\, \omega_{Bi} \simeq 10^{-2}$]. From Eq. (2) and Fig. 4 it can be shown that the average east-west component of the electric field takes values in the range of about ± 15 mV/m; this agrees with the agerage east-west electric field values observed by Mozer and Lucht (1974).

Conclusions

The results of this study show, on the average, a well-defined overnight variation of the mean motions of the auroral ionization along the north-south direction. Positive mean frequency shifts (due to equatorward motion) are generally present before local midnight while negative shifts (northward motions) dominate the morning sector of the events. The reversal of motion occurs during a transition period around local magnetic midnight. During this period, when the mean drift velocity along the line-of-sight goes to zero, the echoes are very weak or disappear completely; also the mean Doppler shift values are less stable and can vary rapidly between positive and negative.

Since there was not any direct information about other physical quantities (e.g., large scale electric fields, electrojet currents) which are believed to be related closely to radio auroral motions, interpretation of the above measurements was based on results reported by other investigations. The results suggest that the meridional motion components of the electrojet current system, in the evening and morning hours, and the Harang discontinuity, in the midnight-evening sector, play an important role in the generation and motion of the scattering irregularities. Based on the principle that the line-of-sight radio auroral Doppler velocities in the north-south direction are due to Hall drifts, such

velocities are a continuous source of information about the average east-west auroral electric fields. The present results show that this mean east-west field component varies between ~ 15 mV/m eastward a few hours before magnetic midnight and ~ 15 mV/m westward a few hours after magnetic midnight.

Acknowledgement. This research work was made possible by grants to G. Sofko by the National Research Council of Canada.

References

Chen, A.J., Rostoker, G.: Auroral-polar currents during periods of moderate magnetospheric activity. Planet. Space Sci. **22**, 1101–1115, 1974

Czechowsky, P.: Movement processes of auroral structures. J. Atmos. Terr. Phys. **36**, 61–77, 1974

Ecklund, W.L., Balsley, B.B., Carter, D.A.: A preliminary comparison of F-region plasma drifts and E-region irregularity drifts in the auroral zone. J. Geophys. Res. **82**, 195–197, 1977

Greenwald, R.A., Ecklund, W.L., Balsley, B.B.: Auroral currents, irregularities, and luminosity. J. Geophys. Res. **78**, 8193–8203, 1973

Greenwald, R.A., Ecklund, W.L., Balsley, B.B.: Diffuse radar aurora: spectral observations of non-two-stream irregularities. J. Geophys. Res. **80**, 131–139, 1975a

Greenwald, R.A., Ecklund, W.L., Balsley, B.B.: Radar observations of auroral electrojet currents. J. Geophys. Res. **80**, 3635–3641, 1975b

Greenwald, R.A., Weiss, W., Nielsen, E., Thomson, N.R.: STARE: A new radar auroral backscatter in Northern Scandinavia. Radio Sci. **13**, 1021–1039, 1978

Haldoupis, C.: VHF Doppler investigations of the auroral ionosphere. Saskatoon, Canada: PhD Thesis, University of Saskatchewan, 1978

Haldoupis, C., Koehler, J.A., Sofko, G.J.: A VHF bistatic Doppler system for radio investigations of the auroral ionosphere. J. Phys. E: **11**, 135–138, 1978

Haldoupis, C., Sofko, G.: Doppler spectrum of 42 MHz CW aurora backscatter. Can. J. Phys. **54**, 1571–1584, 1976

Haldoupis, C., Sofko, G.: On ion-acoustic plasma waves in radio aurora and the two-stream instability. Can. J. Phys. **56**, 100–116, 1978

Heppner, J.P.: The Harang discontinuity in auroral belt ionospheric currents. Geofys. Publ. **29**, 105–114, 1972

Holtet, J.A.: Electric Field microstructures in the auroral E-Region. Geofys. Norv. **30**, No. 4, 1–88, 1973

Horwitz, J.L., Doupnik, J.R., Banks, P.M.: Chatanika radar observations of the latitudinal distributions of auroral zone electric fields, conductivities, and currents. J. Geophys. Res. **83**, 1463–1481, 1978

Hughes, T.J., Rostoker, G.: Current flow in the magnetosphere and ionosphere during periods of moderate activity. J. Geophys. Res. **82**, 2271–2282, 1977

Kamide, Y., Akasofu, S.I., Brekke, A.: Ionospheric currents obtained from the Chatanika radar and ground magnetic perturbations at the auroral latitude. Planet. Space Sci. **24**, 193–201, 1976a

Kamide, Y., Yasuhara, F., Akasofu, S.I.: A model current system for the magnetospheric substorm. Planet. Space Sci. **24**, 215–222, 1976b

Kamide, Y.R., Brekke, A.: Altitude of the eastward and westward auroral electrojet. J. Geophys. Res. **82**, 2851–2853, 1977

Kato, S.: Theory of movement of irregularities in the upper atmosphere. Planet. Space Sci. **11**, 823–830, 1963

Maynard, N.C.: Electric field measurements across the Harang discontinuity. J. Geophys. Res. **79**, 4620–4631, 1974

McDiarmid, D.R., Harris, F.R., McNamara, A.G.: Relationships between radio aurora, visual aurora and ionospheric currents during a sequence of magnetospheric substorms. Planet. Space Sci. **24**, 717–725, 1976

Moorcroft, D.R., Tsunoda, R.T.: Rapid-scan Doppler velocity maps of the UHF diffuse radar aurora. J. Geophys. Res. **83**, 1482–1492, 1978

Mozer, F.S., Lucht, P.: The average auroral zone electric field. J. Geophys. Res. **79**, 1001–1006, 1974

Rostoker, G., Hron, M.: The eastward electrojet in the dawn sector. Planet. Space Sci. **23**, 1377–1389, 1975

Sudan, R.N., Akinrimisi, J., Farley, D.T.: Generation of small scale irregularities in the equatorial electrojet. J. Geophys. Res. **78**, 240–248, 1973

Tsunoda, R.T.: Electric field measurements above a radar scattering volume producing 'diffuse' auroral echoes. J. Geophys. Res. **80**, 4297–4306, 1975

Tsunoda, R.T., Fremouw, E.J.: Radar auroral substorm signatures. 1. Expansive and recovery phases. J. Geophys. Res. **81**, 6148–6158, 1976

Tsunoda, R.T., Presnell, R.I.: On a threshold electric field associated with 398-MHz diffuse radar aurora. J. Geophys. Res. **81**, 88–96, 1976

Unwin, R.S., Knox, F.B.: Radio aurora and electric fields. Radio Sci. **6**, 1061–1077, 1971

Zmuda, A.J., Armstrong, J.C.: The diurnal flow pattern of field-aligned currents. J. Geophys. Res. **79**, 4611–4619, 1974

Received June 13, 1978; Accepted June 6, 1979

J. Geophys. 46, 89–95, 1979

Journal of Geophysics

Q of Mode $_0S_0$*

L. Knopoff, W. Zürn[1], P.A. Rydelek, and T. Yogi[2]

Institute of Geophysics and Planetary Physics University of California, Los Angeles, USA

Abstract. Estimates of Q of mode $_0S_0$ as measured after the Chilean (1960) and Alaskan (1964) earthquakes showed large scatter. The Indonesian quake of August 19, 1977, has provided a new opportunity to determine the attenuation factor for $_0S_0$. Time lapse spectra of *a priori* selected high quality data were analyzed using a maximum-likelihood method and Ricean statistics. Data from South Pole and Los Angeles gave Q values of $6324(1\pm21\%)$ and $6859(1\pm17\%)$ respectively. Taken together the result is $6687(1\pm13\%)$.

Key words: Q-$_0S_0$-Time lapse spectra – Maximum Likelihood Method – Ricean statistics.

Introduction

The value of Q for the purely radial mode of free oscillation $_0S_0$ should be one of the most reliable of physical quantities for the earth. It is uncontaminated by the effects either of the earth's rotation or of asymmetries in the earth's internal constitution. Furthermore its value appears to be high, so its signal persists for an extremely long time. Despite these attractive features, it has only been observed infrequently and the values of Q reported for this mode have a wide variation.

Until recently, analysis of mode $_0S_0$ has focussed on recordings made from only two earthquakes, the great Chilean (1960) and Alaskan (1964) earthquakes.

---

* Publication no. 1977, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024, USA
Contribution No. 176, Geophysical Institute University of Karlsruhe, Hertzstr. 16, D-7500 Karlsruhe, Federal Republic of Germany

[1] *Present address:* Geowissenschaftliches Gemeinschaftsobservatorium, Universitäten Karlsruhe/Stuttgart, Heubach 206, D 7620 Wolfach, Federal Republic of Germany

[2] *Present address:* Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA

0340-062X/79/0046/0089/$01.40

Table 1. Values of Q for ${}_0S_0$ for Alaska and Chilean Earthquakes

Q	Author	Earthquake – Station
> 7,500	Ness et al. (1961)	Chile (1960) – UCLA
900	Smith (1961)	Chile (1960) – Isabella
> 25,000	Slichter et al. (1966)	Alaska (1964) – UCLA
13,000 ± 2,000	Slichter (1967)	Alaska (1964) – UCLA
4,229 ± 5%	Sailor and Dziewonski (1978)	Alaska (1964) – $UCLA_4$
3,996 ± 11%	Sailor and Dziewonski (1978)	Alaska (1964) – $UCLA_7$

Values of Q for this mode have been reported as listed in Table 1. Most of these analyses have been made from recordings made on the UCLA gravimeters/ultra-long period seismometers. For both earthquakes, analyses of these records made by measuring the spectral amplitude in each of several time-lapsed intervals, give values on the order of thousands. The much lower value by Smith was obtained from the spectral line shape, coincidentally from the only record in the list not taken at UCLA; we have performed a subsequent analysis which shows that the line will be broadened to Smith's value, if Q is of the order of thousands and account is taken of the $\sin x/x$ transform of the record window. The variation among the high values may be attributable to significant power input from even small amounts of aftershocks and from major glitches due to instrumental defects. Further, Sailor and Dziewonski (1978) have selectively rejected a number of estimates for some of the lagged points on *a posteriori* grounds; among the points rejected are those with the largest spectral amplitudes.

A new illustration of the variability of determinations of Q for this mode is given by Buland et al. (1979). Indicated Q values for five records of the Sumbawa earthquake of August 19, 1977, range from 2950 to 8090. The result of stacking all five records yielded a Q of 4100. The possibility continues to exist, in view of the large variation in values of Q, that all are compatible with each other within rather large error bounds. If this is the case, the problem of Q for ${}_0S_0$ is therefore reducible to one in which the identifiable errors are minimized; this, in turn, implies that we be able to compute the size of the residual error and test whether the values reported earlier are consistent.

II. Data and Analysis

As with Buland et al. (1979), the Indonesian earthquake of August 19, 1977, provided a rare opportunity to detect the radial mode of lowest order ${}_0S_0$ in a relatively undisturbed condition. Our data were collected with the ultra-long period seismometers (Nakanishi et al., 1976) in operation at the South Pole (SPA) and at UCLA (LMS). This earthquake, occurred at coordinates 11.1°S, 118.5°E., $h = 20$ km and origin time 06:08:55.2 UT and had a surface wave magnitude of 7.9, all data from PDE.

The output of the seismometer, as described in Nakanishi et al., (1979) is analog filtered in the pass band 270 s to 7200 s at the 3db points to suppress the diurnal and semi-diurnal tides at the long period end and to suppress large amplitude surface waves at periods around 20 to 30 s. The filtered output

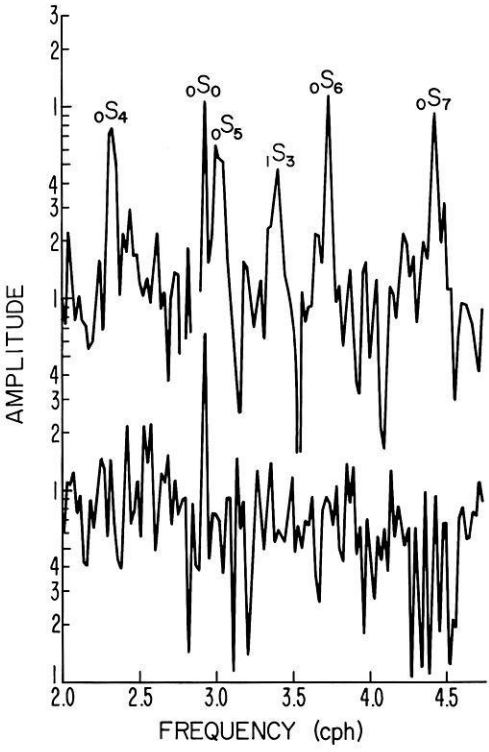


Fig. 1. Portion of spectrum at UCLA showing $_0S_0$ and other nearby modes. The upper curve is for the section starting at 40h. The lower is for the section starting at 386 2/3h. The modes other than $_0S_0$ are imbedded in the noise in the *lower curve*. The lower curve should be multiplied by 10

is sampled at 10 s intervals and recorded digitally. The records contain several large aftershocks and other large earthquakes as well as some unexplained glitches which are of instrumental origin. Only stretches of data without identifiable glitches or other interruptions were used in the data analysis. *No deglitching was applied to the data.* After this *a priori* selection of the raw seismographic data, no further selection has been applied. The preceding three sentences describe the basic differences between the present work and the preceding.

We have chosen a record length of 16,200 sample points of 45 h for each of the record segments to be analyzed. This number is appropriate for use with Singleton's Fast Fourier Transform since $16200 = 2^3 \times 3^4 \times 5^2$. Further, the effect of the periodicity of $_0S_0$ incommensurate with the record length is minimized. The nominal period of $_0S_0$ is 2.93246 cph or a period of 1227.64 s = 20.4606 min. The error in amplitudes due to the fact that 162000 s is 131.961 cycles of $_0S_0$ rather than an integral number of cycles is

$$1 - \frac{\sin 2\pi(0.039)}{2\pi(0.039)} = 1\,\%. \tag{1}$$

The record length is sufficiently long to resolve $_0S_0$ from $_0S_5$. The contribution of $_0S_5$ to amplitudes of $_0S_0$ from a similar $\sin x/x$ calculation is at most 2%. The rapid decay of $_0S_5$ compared with $_0S_0$ helps make this error small after two days (Fig. 1).

Table 2. Time lapse spectral estimates for ${}_0S_0$. (Note that start time is measured from 21 h 20 min on August 19, 1977, i.e., more than 15 h after the quake; each record length = 45 h)

Start time (hours)	${}_0S_0$ Amplitude (mV/cph)	Noise Level (mV/cph)	Start Time (hours)	${}_0S_0$ Amplitude (mV/cph)	Noise Level (mV/cph)
South Pole			UCLA		
0	9.1368	1.5	40.00	10.304	1.2
80.67	6.3314	1.0	122.67	9.6377	1.1
184.00	5.9886	0.9	194.67	8.1930	1.5
244.67	4.9114	1.0	314.67	6.6027	1.1
352.27	5.1269	1.2	386.67	6.5052	0.9
458.67	4.2927	0.9	462.67	5.6644	1.4
506.67	3.5833	1.2	507.75	5.5252	1.3
554.67	3.3947	1.1	578.67	5.9279	1.7
			626.67	4.4317	1.6

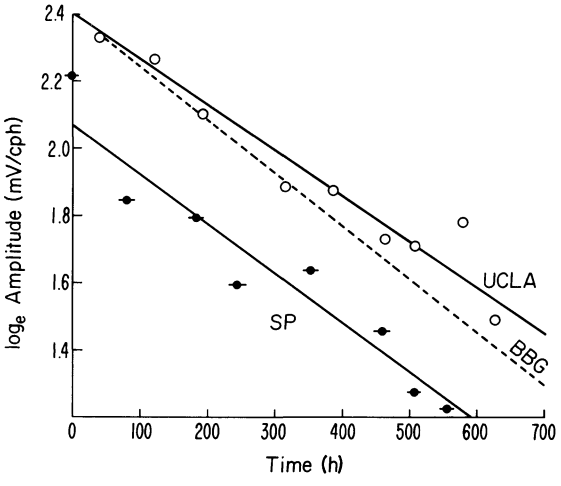


Fig. 2. Logarithmic spectral amplitudes for records at UCLA and South Pole. A straight line *(BBG)* with slope appropriate to $Q = 4100$ is indicated (Buland et al., 1979)

Spectra were taken of record segments in the frequency neighborhood of ${}_0S_0$ for the recordings at both SPA and LMS. As noted, the record segments are 45 h long and we exclude extraneous seismic events and other glitches. No overlap of record segments was allowed in either of the two station analyses. Eight time lags were taken for the South Pole record and nine for the UCLA record, a total of 275,400 points in all. For later use, the noise level in the neighborhood of ${}_0S_0$, is listed along with the results in Table 2. Time is measured from 21h 20m August 19, 1977. The amplitudes are given, in millivolts per cph at the output of the meters.

III. Maximum Likelihood Analysis

A linear least-squares fit through the points displayed in Fig. 2 is inappropriate because of the bias introduced by the logarithmic operator. Maximum likelihood

methods are more appropriate since they take into account the signal-to-noise ratio. However, the model for the noise must be considered carefully. We report in Table 2 and Fig. 2 spectral amplitudes without regard to phase. Because spectral amplitude is a scalar positive variable, its noise spectrum cannot be taken to be Gaussian; if it were Gaussian, this would imply that the tabulated values, which are the result of adding noise to the signal, could under some circumstances be negative.

The Ricean distribution function is appropriate for absolute spectral amplitudes, if the noise on the seismogram is Gaussian. Rice (1944, 1945, 1948; see also Middleton, 1968, Chapter 9) appears to have been the first to have shown that the probability of finding peak spectral amplitude estimates between A_i and and $A_i + dA_i$ is

$$P(A_i)\,dA_i = \frac{A_i}{N_i^2} \exp\{-\tfrac{1}{2}(A_i^2 + S_i^2)/N_i^2\} I_0\left(\frac{A_i S_i}{N_i^2}\right) dA_i \tag{2}$$

where S_i is the signal strength and N_i is the root mean square noise level of the system in the absence of the signal. We take the signal strength for the i^{th} time lapse t_i to be

$$S_i = A_0 \exp(-\pi t_i / Q T) \tag{3}$$

where T is the period of mode ${}_0S_0$.

For large S_i/N_i we find, from the first term of the asymptotic expansion of the modified Bessel function I_0, that

$$P(A_i)\,dA_i \approx \left(\frac{A_i}{2\pi S_i}\right)^{\frac{1}{2}} \frac{1}{N_i} \exp\{-\tfrac{1}{2}(A_i - S_i)^2/N_i^2\}. \tag{4}$$

This distribution is approximately Gaussian. If it were Gaussian we would have

$$P_i(A_i)\,dA_i = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \frac{1}{N_i} \exp\{-\tfrac{1}{2}(A_i - S_i)^2/N_i^2\}. \tag{5}$$

From Table 2, the ratios of A_i/N_i vary between 3.4 and 8.8. Let us take as a most unfavorable case, the values $S_i/N_i = 2.0$ and $A_i/S_i = 0.7$. We find that the functions in (2) and (4) differ by $5\frac{1}{2}\%$. Since this upper bound is small, we use (4) in our maximum likelihood estimation. The differences between (2) and (5) for these same conditions are about 13% and cannot be considered to be small. We have performed the maximum likelihood analysis on the values of Table 2 using the expression

$$\text{Min}\,\Phi = -\text{Max} \sum_{i}^{8 \text{ or } 9} \log P(A_i) \tag{6}$$

using the approximation (4) to Ricean statistics. The minimization is performed with regard to the variables Q and A_0 in (3). We get $Q_0 = 6324$ for ${}_0S_0$ measured at the South Pole and 6859 at UCLA.

With regard to estimates of error in the values of Q, we can write

$$\Phi \simeq \Phi_0 + {}^1/_2 \, (\delta Q)^2 / \sigma_Q{}^2 \tag{7}$$

since the function $\exp\Phi$ is normal in the vicinity of the extremum by the central limit theorem. In the above expression Φ_0 is the value of Φ at the extremum, $\sigma_Q{}^2$ is the variance we are seeking and δQ is a variation in Q from the value which yields the extremum. We obtain σ_Q by finding the value of δQ such that

$$\Phi - \Phi_0 = {}^1/_2 \tag{8}$$

This method was used in Knopoff (1961). We find the results

$$\begin{array}{ll} \text{South Pole} & Q = 6324 \; (1 \pm 21\%), \\ \text{UCLA} & Q = 6859 \; (1 \pm 17\%). \end{array} \tag{9}$$

The error is listed as a multiplicative percentage rather than as an additive percentage because of the exponential character of the operation involving the factor Q.

Since the amplitude of mode ${}_0S_0$ is uniform over the entire surface of the earth, we can combine the data of both stations after allowance for the different sensitivities of the two instruments. The instrument at South Pole is 1.38 times less sensitive than that at UCLA. The 17 record segments now overlap in time but still contain independent estimates. The analysis yields

$$Q = 6687 \; (1 \pm 13\%)$$

Our results are not in significant disagreement with that of Buland et al. (1979) for their longest record segment.

Acknowledgements. This work was supported by the National Science Foundation under grant DPP 76–17234 A01. Our special thanks go to the Office of Polar Programs for their invaluable help in the logistics of this research. We are grateful to Professor Karl Fuchs, Geophysical Institute, University of Karlsruhe, Germany for continuously supporting the participation of Walter Zürn in this program.

References

Buland, R., Berger, J., Gilbert, F.: Observations from the IDA network of attenuation and splitting during a recent earthquake. Nature **277**, 358–362, 1979

Knopoff, L.: Statistical accuracy of the fault-plane problem, Publications of the Dominion Observatory. **24**, 317–319, 1961

Middleton, D.: An introduction to statistical communication theory. New York: McGraw-Hill Book Co., 1968

Nakanishi, K.K., Knopoff, L., Slichter, L.B.: Observation of Rayleigh Wave dispersion at very long periods. J. Geophys. Res. **81**, 4417–4421, 1976

Ness, N.F., Harrison, J.C., Slichter, L.B.: Observations of the free oscillations of the earth. J. Geophys. Res. **66**, 621–629, 1961

Rice, S.O.: Mathematical analysis of random noise, I. Bell System Tech. J. **23**, 283–332, 1944

Rice, S.O.: Mathematical analysis of random noise, II. Bell Syst. Tech. J. **24**, 46–156, 1945

Rice, S.O.: Statistical properties of a sine wave plus random noise, Bell Syst. Tech. J. **27**, 109–157, 1948

Sailor, R.V., Dziewonski, A.M.: Measurements and interpretation of normal mode attenuation. Geophys. J. Roy. Astron. Soc. **53**, 559–581, 1978

Slichter, L.B.: Spherical Oscillations of the Earth, Geophys. J. R. Astron. Soc. **14**, 171–177, 1967

Slichter, L.B., MacDonald, G.J., Caputo, M. Hager, C.L.: Comparison of spectra for spheroidal modes excited by the Chilean and Alaskan quakes. Geophys. J. R. Astron. Soc. **11**, 256 1966

Smith, S.W.: An investigation of the earth's free oscillations. California Institute of Technology, Pasadena, California, USA: PhD Thesis 1961

Received March 15, 1979; Revised Version June 27, 1979; Accepted August 9, 1979

J. Geophys. 46, 97–121, 1979

Journal of Geophysics

Finite Element Convection Models: Comparison of Shallow and Deep Mantle Convection, and Temperatures in the Mantle

U. Kopitzke

Institut für Geophysik und Meteorologie der Technischen Universität Braunschweig, Mendelssohnstrasse 1A, D-3300 Braunschweig, Federal Republic of Germany

Abstract. A Finite Element Method for solving the convection problem in a fluid with position-dependent Newtonian viscosity is developed, using bicubic and biquadratic spline functions on a rectangular grid. Introducing weak (less viscous) zones at the active margins of the lithosphere, dynamical mantle convection models are established which have a nearly uniform surface (plate) velocity and a satisfactory heat flux profile. A comparison of upper and deep mantle convection shows:
– a moderate increase of viscosity with depth cannot confine the flow to the upper mantle;
– in shallow depth convection models the temperature is too low inside the cell, but deep mantle convection models yield satisfactory temperatures for the upper mantle.
For that reason deep (or whole) mantle convection should be the favored hypothesis.

Key words: Mantle convection – Finite element method – Mantle geotherm.

1. Introduction

There exist two major kinds of models which describe mantle convection or basic plate tectonic processes. One kind are simple kinematic models (for example: Schubert et al., 1976; Richter and McKenzie, 1978), making assumptions on a prescribed plate velocity (and a prescribed temperature at a reference depth). These models are easy to handle – they can often be performed by analytical methods. But they do not involve the driving mechanism –, i.e., thermally caused differences of density – and so it remains unclear whether the assumptions about plate velocity and temperature are justified and consistent with the chosen rheology.

The dynamical or intrinsic convection models on the other hand need only assumptions on the distribution of heat sources and the rheological behavior of

0340-062X/79/0046/0097/$05.00

the material. They yield informations about both the velocity and the temperature field. Such models were investigated, for example, by Torrance et al. (1973), McKenzie and Roberts (1974), Houston and DeBremaecker (1975) and DeBremaecker (1977a). These models were successful in showing that mantle convection is able to drive the plates, and that cells with a large aspect ratio (the ratio between length and depth of the cell) are possible under mantle conditions.

But so far the dynamical convection models failed to incorporate the lithosphere (defined in a rheological sense) in a reasonable manner. The upper boundary layer does not behave like a rigid plate – the surface velocity varies steadily, instead of being constant. In connection with this fact, the heat flux profile does not agree with the observed ones. DeBremaecker (1977a) discussed possible reasons for this disagreement. He argued that it might be caused by the fact that the surface is not free to move up and down in the model, or that it is assumed to be two dimensional or to be in steady state. However, it seems most likely that it is caused by a too simple rheological behaviour of the model lithosphere, which is assumed to have a high viscosity that is uniform over the whole length of the plate. At the active margins of a plate the behaviour can be quite different from that in the middle of the plate.

Another important point is – although not much attended by the authors – that the upper mantle temperature in the models is considerably lower than the expected one. This may be due to the fact that all these models assume the convection to be restricted to the upper mantle (to about 700 km depth).

The hypothesis that the lithospheric slab (and so the convection current, too) does not enter the lower mantle is mainly supported by the absence of earthquakes below 700 km and by the fact that the stress state in the lower parts of the subducted lithosphere is always compressive. So it is assumed that there is a barrier at about 700 km. Such a barrier could consist of a radial chemical heterogeneity (increased iron content) or of a very steep increase of viscosity at or below the 670 km – discontinuity. However, now there is an increasing opposition to this point of view (Davies, 1977; O'Connell, 1977). A change of iron content in the mantle cannot be excluded, but it is not required by recent mantle models, and the assumption that there is none seems to be favored (Ringwood, 1975, pp. 481–514). Sammis et al. (1977) pointed out that there should not be a large increase of viscosity in connection with the phase boundaries, and Jordan (1977) showed that the lithospheric slab might penetrate the lower mantle at the Kuril trench.

Until now only simple models of whole mantle convection exist, based on boundary layer theory (e.g., Elsasser et al., 1979) or on marginal stability analysis (e.g., Davies, 1977). It seems advisable to carry out fully dynamical models of deep convection and to compare them with shallow convection models. This may lead to further indications which could help decide which of the contradictory hypotheses is right.

So far always finite difference techniques have been used to model mantle convection. This method is well established and easy to implement for special problems. But its accuracy and reliability becomes doubtful when parameters and variables vary greatly from one point of the grid to the next like in the thin boundary layers of the convection cell. Finite element methods are more

accurate, because these quantities are not represented by single points but by continuous test functions.

I developed a finite element method for solving the equations of convection in a fluid with Newtonian position-dependent viscosity. Bicubic (for the stream function) and biquadratic (for the temperature) spline functions are used as test functions on a rectangular grid. Models of mantle convection are presented which have a nearly uniform surface velocity, and heat flux profiles which agree very well with the observed values. Dynamical model calculations are carried out for deep mantle convection and compared with upper mantle models. Special attention is payed to the thermal regime of the mantle and the mantle geotherm.

2. Mantle Temperatures

On the one hand convection models can give valuable indications on the temperature distribution in the mantle. On the other hand there are some observational constraints, which should be fulfilled by the model. Thus we have an important criterion to judge the model, since only little is known about the flow pattern beneath the plates and otherwise only surface features (plate velocity, heat flux, etc.) can be used to compare the model with the real earth. Therefore I shall first review shortly the recent observational knowledge about mantle temperatures.

For the lithosphere, temperatures can be derived from petrological data (pyroxene geothermometry). At the bottom of the oceanic lithosphere at 100 km, the temperature is 950°–1,200° C according to McGregor and Basu (1974) and 1,100–1,250° C according to Mercier and Carter (1975). No oceanic pyroxene data are available for greater depth. For the upper mantle, Tozer (1970) derived the temperature from electrical conductivity data. Assuming an olivine mantle with 10% iron content, he estimated the mean temperature of the uppermost 400 km to be below 1,290° C with a best fit of 1,150° C.

The olivine-spinel phase boundary gives another opportunity to calculate the temperature at its depth, because it is known from laboratory measurements, under which $P-T$-conditions the transition takes place. In this way Fusijawa (1968) derived a temperature of 1,340° ± 190° C for the depth level of about 370 km. Gebrande (1975) got 1,400° ± 100° C and Graham (1970) derived 1,450° ± 150° C, the latter used also information from a petrological model with the according temperature-dependence of the elastic moduli etc.

For the lower mantle only estimates from petrological models are available. These estimates are more insecure because a number of uncertain assumptions have to be made. Concerning the 670/700 km depth level, Graham and Dobrzykowski (1976) calculated the temperature to be 1,600° ± 400° C, and Watt and O'Connell (1978) gave 1550° ± 150° C. The compositional model of Wang (1972) yields temperatures of 2,500° C for 1,300 km, and 3,000° C for 2,800 km depth with an error of ± 800° C.

A last constraint results from the fact that the outer core is liquid, so the temperature at the core-mantle boundary must be above the melting point of the core's iron alloy. The melting temperature depends on the light alloying

element which must be in the core; and because of the very high pressure, only rough estimates from thermodynamic principles can be made. For pure iron Higgins and Kennedy (1971) derived a melting point of 3,750° C at the core-mantle boundary, and Leppaluoto (1972) gave a value which is even 1,000 degrees higher. Silicon as an alloying element would not reduce the melting temperature very much, but sulphur can do so. For the iron-sulphur eutecticum ($\simeq Fe_2S$) Stacey (1977) estimated the melting point to be 2,600° C, and Tolland (1974) gives only 2,260° C. Because the temperature at the core-mantle boundary should be some hundred degrees above the melting point, a minimum temperature constraint of about 2,800° C exists. But - depending on the light element in the core - it might be that temperatures up to 5,000° C are required.

3. Description of the Model

Geometry and Hydrodynamic Conditions

Now the basic design of the model will be explained, and its main idealizations shall be discussed.

A convection cell in a rectangular enclosure - extended infinitely into the third dimension - is to be modelled. The cell shall be connected with an oceanic plate and be part of a periodic scheme of equal convection cells. These simple geometric conditions are required to keep the system practicable, but they imply several unrealistic simplifications:

- It is not likely that the flow pattern beneath the plates is two-dimensional or nearly two-dimensional.
- In the model the descending angle of the plate is 90°, while it is less in reality (15°–90°). Moreover, because of the periodicity, there are two oceanic plates converging against each other and *both* descending into the mantle. Such a behavior cannot be observed anywhere on the earth.
- The surface is not free to move vertically.
- The aspect ratio of the cell is prescribed and not free to adopt a 'natural value'.
- The curvature of the earth is not included. Concerning shallow convection models, this seems not to be so important, but for deep convection it could be of some consequences (see Chaps. 6.4 and 8).

The enclosure is filled with a Newtonian fluid. Its viscosity depends on position - mainly on depth - but it is not controlled by temperature and pressure. This, of course, is a serious simplification, because it is likely that the rheology of the earth's mantle is non-Newtonian and sure that it is strongly temperature-dependent. But the existing models show that the results do not become better if such a rheology is used. They seem to be even worse when the viscosity is not confined to a relative small maximum value for the lithosphere (Houston and DeBremaecker, 1975; DeBremaecker, 1977a). This may be due to a change in deformation mechanism when the temperature becomes too low ($T < 0.5\, T_{melt}$), from steady state creep (diffusion or dislocation creep) to plastic deformation, transient creep, fracture or elastic deformation. Especially in the

lithosphere – in the normal as in the subducted part – these deformation laws could be valid, and the effective viscosities estimated from steady state creep laws are too high. As a simple approximation it is sufficient to use an appropriate chosen depth dependent viscosity. Only for the lithosphere it seems reasonable to introduce also a length dependence, by taking different values for the middle of the plate and for the active margins (see Chap. 6.1).

The boundary conditions usually applied for the flow are 'free slip' at the sides (periodicity!) and at the top, and the no slip condition for the bottom. The latter could be justified by a very steep increase of viscosity at the bottom of the cell, which would also inhibit convection below that boundary. Concerning models of whole mantle convection, 'free slip' is the appropriate condition for the lower boundary, too.

Thermal Conditions

The cell shall be heated by homogeneously distributed (radioactive) inner sources, and by an uniform heat flux from below. That heat may be produced by radioactivity, too, or it may have other sources. Furthermore, frictional heating and the effects of adiabatic compression/decompression will be included as additional sources (or sinks) of heat. The thermal conductivity is mildly temperature dependent, therefore it shall be allowed to vary with depth in the model. The thermal boundary conditions are adiabatic at the sides (periodicity!), $T=0$ at the top, and a specified heat flux at the bottom. The flow is driven by buoyancy forces which are caused by lateral temperature differences via thermal expansion. The effects of phase transitions – both their driving (or hindering) force on the flow and their release of latent heat – will be ignored.

In the model only steady state convection cells are considered. Of course, the convecting mantle of the earth is not in steady state. But the time-dependence of real convection in the earth consists mainly in a change of size and dimensions of the plates (and cells), which cannot be simulated by such simple models at all. To get information about a hypothetical mean convection cell, it seems best to consider the steady state situation.

The model described so far contains a number of grave simplifications (two-dimensionality, steady state, etc.). For that reason its results should not be overestimated, for example by trying to explain peculiarities of special plates with them. But we can hope that they are useful in considering general aspects of plate motion, subplate flow, driving mechanism and especially the thermal regime of the mantle.

4. Equations to Be Solved

The governing equations of convection are (Andrews, 1972):

$$\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial z^2}\right)(\eta(\psi_{xx}-\psi_{zz}))+4\frac{\partial^2}{\partial x\,\partial z}(\eta\psi_{xz})=\frac{\partial}{\partial x}(\rho_0 g\alpha T)$$

(Hydrodynamic equation)

Table 1. List of symbols

α	Thermal expansivity	Q_{int}	Internal heat generation per unit volume
c_p	Specific heat	ρ_0	Density (mean or reference value)
η	Viscosity	T, T_{abs}	Temperature, absolute temperature
g	Gravity acceleration	$u=\psi_z$	Horizontal velocity
$\kappa=\frac{\lambda}{\rho c_p}$	Thermal diffusivity	$w=-\psi_x$	Vertical velocity
λ	Thermal conductivity	x	Horizontal coordinate
ψ	Stream function	z	Vertical coordinate, pointing upwards
Q_b	Heat flux through the bottom		

$$\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}+w\frac{\partial T}{\partial z}-\frac{1}{\rho_0 c_p}\left(\frac{\partial}{\partial x}(\lambda T_x)+\frac{\partial}{\partial z}(\lambda T_z)\right)+w\rho_0\alpha T_{\text{abs}}$$
$$=Q_{\text{int}}+\frac{1}{\rho_0 c_p}(\eta(\psi_{xx}-\psi_{zz})^2+4\eta\psi_{xz}^2).$$

(Energy equation).

These equations are valid for an incompressible fluid with negligible influence for inertia (Prandtl number is infinite). The Boussinesq approximation is made, that means that the density is taken to be constant, except to calculate the buoyancy term in the hydrodynamic equation where it is assumed to depend on temperature only via thermal expansion.

In order to prepare the equations for the numerical procedure, dimensionless variables are introduced, following Torrance et al. (1973). The quantities α, g, and c_p are assumed to be constant, λ is replaced by κ. Only the steady state situation shall be considered, so the term $\partial T/\partial t$ can be omitted. All quantities in the following equations are dimensionless, as defined in Table 2.

$$\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial z^2}\right)(\eta(\psi_{xx}-\psi_{zz}))+4\frac{\partial^2}{\partial x\,\partial z}(\eta\psi_{xz})=Ra\frac{\partial T}{\partial x} \tag{1}$$

$$-\frac{\partial}{\partial x}(\kappa T_x)-\frac{\partial}{\partial z}(\kappa T_z)+uT_x+wT_z+Di\cdot w\cdot T_{\text{abs}}$$
$$=Q_{\text{int}}+\frac{Di}{Ra}\cdot\eta\cdot((\psi_{xx}-\psi_{zz})^2+4\psi_{xz}^2). \tag{2}$$

The boundary conditions are

at the sides	$\psi_{xx}=0$			and	$T_x=0$,
at the top	$\psi_{zz}=0$			and	$T=0$,
at the bottom	$\psi_{zz}=0$	or	$\psi_z=0$	and	$T_z=-Q_b/\kappa$,

and $\psi=0$ on the whole enclosure.

Table 2. Dimensionless quantities. Dimensionless variables are signed by a "$\bar{}$", the physical variables have none. h, κ_0, T_0, and η_0 are properly chosen reference values

$\bar{x}$	$=x \cdot 1/h$	Coordinates
$\bar{u}$	$=u \cdot h/\kappa_0$	Velocity
$\bar{T}$	$=T \cdot 1/\Delta T_0$	Temperature
$\bar{\kappa}$	$=\kappa \cdot 1/\kappa_0$	Thermal diffusivity
$\bar{\eta}$	$=\eta \cdot 1/\eta_0$	Viscosity
$\bar{\psi}$	$=\psi \cdot 1/\kappa_0$	Stream function
$\bar{Q}_{\text{int}}$	$=Q_{\text{int}} \cdot h^2/(c_p \kappa_0 \Delta T_0 \rho_0)$	Rate of internal heating
$\bar{Q}_b$	$=Q_b \cdot h/(c_p \kappa_0 \Delta T_0 \rho_0)$	Bottom heat flux
Ra	$:=\dfrac{\alpha g \rho_0 \Delta T_0 h^3}{\eta_0 \kappa_0}$	Rayleigh number
Di	$:=\alpha g h/c_p$	Dissipation number

5. Numerical Method

The first attempt to solve the convection problem with a finite element method was made by Sato and Thompson (1976). Their calculation based on the original Navier-Stokes equation instead of the stream function formulation [Eq. (1)], and they used a 'classical' finite element method. The main advantage of the classical method is that it is possible to use an arbitrary irregular grid, this does not seem necessary for the model designed before. Therefore I prefered a spline function approach on a rectangular grid. It has the advantage that the required continuity of the test functions and their derivatives can be easily fulfilled.

The differential operator of the stream function Eq. (1)

$$\hat{L}=\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial z^2}\right)\eta\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial z^2}\right)+4\frac{\partial^2}{\partial x\,\partial z}\left(\eta\frac{\partial^2}{\partial x\,\partial z}\right)$$

is positive definite, so the Rayleigh-Ritz variational principle is applicable. This leads to a system of linear equations (see, for example, Prenter, 1975, pp. 201ff.):

$$A\,x=r \tag{3}$$

with the matrix elements being

$$a_{ij}=\int \psi_i \hat{L}\psi_j\,df \tag{4}$$

and the right hand side vector

$$r_i=Ra\int \psi_i \frac{\partial T}{\partial x}\,df.$$

ψ_i, ψ_j are the test functions and x_i the coefficients of the approximate solution

$$\psi=\sum x_i \psi_i.$$

The integral (4), which contains fourth order derivatives, can be transformed into an expression containing only second order derivatives:

$$\int \psi_i \hat{L} \psi_j \, df = \int \left(\frac{\partial^2}{\partial x^2} \psi_i - \frac{\partial^2}{\partial z^2} \psi_i \right) \eta \left(\frac{\partial^2}{\partial x^2} \psi_j - \frac{\partial^2}{\partial z^2} \psi_j \right) + 4 \frac{\partial^2}{\partial x \, \partial z} \psi_i \left(\eta \frac{\partial^2}{\partial x \, \partial z} \psi_j \right) df.$$

It is sufficient to use test functions which are continuous up to their second derivatives, for that reason bicubic splines with C^2 - smoothness (Prenter, 1975, pp. 131ff.) are used. The matrix A has band structure and is positive definite; therefore only one Cholesky transformation must be done, and then Eq. (3) can easily be solved for any arbitrary temperature field.

The differential operator of the energy Eq. (2) (for fixed values of the velocity field) is not positive definite, thus no variational principle exists. So I take a weighted residual process to solve the equation approximately. This means that the integrals

$$\int \theta_i (\hat{K} T - s) \, df \qquad i = 1, \ldots, n \tag{5}$$

are forced to become zero. $\hat{K}(u, w)$ is the differential operator of Eq. (2), s is the right hand side in (2), T is the approximate solution in spline space, and the θ_i are a set of weighting functions. Usually the weighting functions are chosen equal to the test functions (Galerkin method). The integrals (5), containing second order derivatives, can be transformed by partial integration, they will then contain only first order derivatives. For this reason it is sufficient to take test functions of lower order than for the hydrodynamic equation, and bi-quadratic splines with C^1-smoothness are applied.

In my calculations the Galerkin process proved to be successful only for small Rayleigh numbers, i.e., as long as the convective heat transport is not much more dominant than the conductive transport. But at high Rayleigh numbers ($>3 \cdot 10^4$), (spatial) oscillations occurred in the numerical solution which had the same periodicity as the finite element grid. Concerning the numerical treatment of convective heat transport, a special weighted residual method - similar to the well known 'upwind differencing scheme' in finite difference techniques - seems to be appropriate. As proposed by Zienkiewicz (1977, pp. 633ff.), an upwind finite element method was constructed by composing the weighting function of the normal Galerkin-(test-)function and a special antisymmetric function of higher order (Fig. 1). The antisymmetric function is weighted in a proper manner, dependent on the sign and magnitude of the local velocity field. This procedure made the oscillations disappear.

The numerical grid, which was used for shallow mantle convection models, is shown in Fig. 2. It is composed of 32 elements in length and 11 elements in depth, accordingly it contains 352 elements. In deep mantle models, 5 more elements are added in depth, the total number of elements is then 512. The size of the elements is variable, the vertical extension increases with depth by a factor

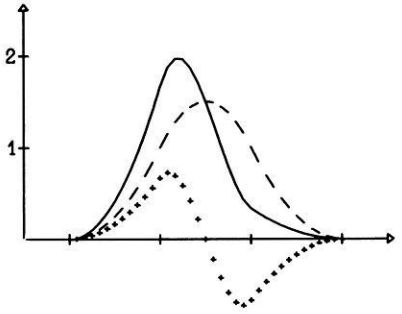


Fig. 1. Example of an one-dimensional weighting function on an equispaced grid for solving the convective heat transfer problem by an 'upwind' finite element method. ------: Normal or Galerkin weighting function (quadratic B-spline); + + + +: antisymmetric weighting function having 2/3 of the possible maximum amplitude; ——: the whole weighting function, being the sum of the other two functions

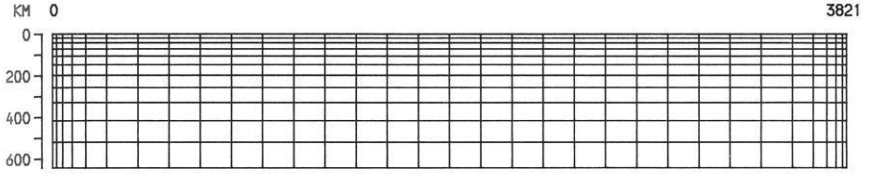


Fig. 2. Numerical grid used in shallow depth models. The mesh is finest in the upper and lateral boundary layers where the variations of stream function and temperature can be expected to be most striking

of 1.2 from one element to the following, starting with 20 km at the top of the lithosphere. The length of the elements remains constant in the middle (150 km), but at the margins it decreases exponentially down to 19.75 km for the last element.

To iterate into steady state, the easiest way is by starting from an initial temperature field, calculating the according velocity field by Eq. (1), and then evaluating by Eq. (2) a new temperature field. Since the $\partial T/\partial t$-term is omitted from Eq. (2), the calculated 'new' temperature field would result if the velocity field would be stationary. But as long as the velocity field has not reached the final state, the procedure has to be repeated: again the velocity field belonging to the last temperature distribution via Eq. (1) is calculated, and with it another improved temperature field, and so on.

But this simple method converged into steady state only when the Rayleigh number was small. At high Rayleigh numbers it was unstable, oscillations occurred which decreased very slowly or were even increasing. Therefore a modified method had to be applied: The calculated 'new' temperatures are replaced for the further calculation by a linear combination of 'old' and 'new' temperatures

$$T = (1-\beta)\, T_{\text{new}} + \beta\, T_{\text{old}}.$$

In deep mantle calculation, the factor β had to be as high as 0.9 to keep the procedure stable. Of course, a large β makes the rate of convergence slow. The need to use such a high 'reduction factor' may indicate that there is no true

stable steady state of convection at very high Rayleigh numbers. Of course, it is only a hope that this forced pseudo steady-state represents the mean situation of true time-dependent convection. However, the hope is encouraged by the observation that in the undamped solutions temperatures and stream function are oscillating just around their values of the damped solution.

The calculation was stopped when both the change in temperatures and in the stream function was at every point less than 0.1 % from one step to the next. Usually a much lower limit is required, this large limit was chosen in order to save computer time. When reducing the limit by one order of magnitude deviations of a few degrees can be observed in the temperature field. Thus steady state is not entirely reached, but it seems to be close enough, especially if one regards that steady state is only an idealization of the model with respect to the real mantle. The deviation between total heat input and surface heat flux, which is less than 1 % in the shallow and less than 3 % in the deep models, might be taken as another measure of the closeness to steady state. However, this may have different reasons, since conservation of energy is not guaranteed perfectly by the numerical procedure.

The reliability and high accuracy of the method was confirmed by comparing different test models with own finite difference calculations and with the results of Torrance et al. (1973) and Houston and DeBremaecker (1975).

6. Parameters of the Model

6.1. Viscosity

Both the linear diffusion creep and the nonlinear dislocation creep have a temperature-pressure dependence following the law

$$\eta = A \cdot T \exp((E^* + p V^*)/RT). \tag{6}$$

E^* = activation energy; V^* = activation volume;
A = proportionality constant.
(in the nonlinear case η must be considered as the effective viscosity and A would be stress-dependent)

In order to construct a viscosity - depth profile, proper values for E^* and V^* are chosen (100 kcal/mol; 10 cm^3/mol above 650 km and 9.5 cm^3/mol below 650 km) and a guess of the temperature - depth profile with regard to the discussion in Chap. 2 was made. The constant A was taken to be 100 Poise/K in order to make the viscosity minimum $3 \cdot 10^{21}$ Poise, which is in good accordance with data from postglacial uplift.

Applied to the lithosphere, Eq. (6) would give values that are too high, rising up to 10^{30} Poise and more above 50 km. For that reason it is only used to calculate the viscosity below 100 km. The profile is continued into the lithosphere in a reasonable manner, reaching a maximum value of $4 \cdot 10^{25}$ Poise at the top and having a mean value of about 10^{24} Poise (Walcott, 1970). The resulting profile is shown in Fig. 3a. Near the spreading center, the lithosphere becomes

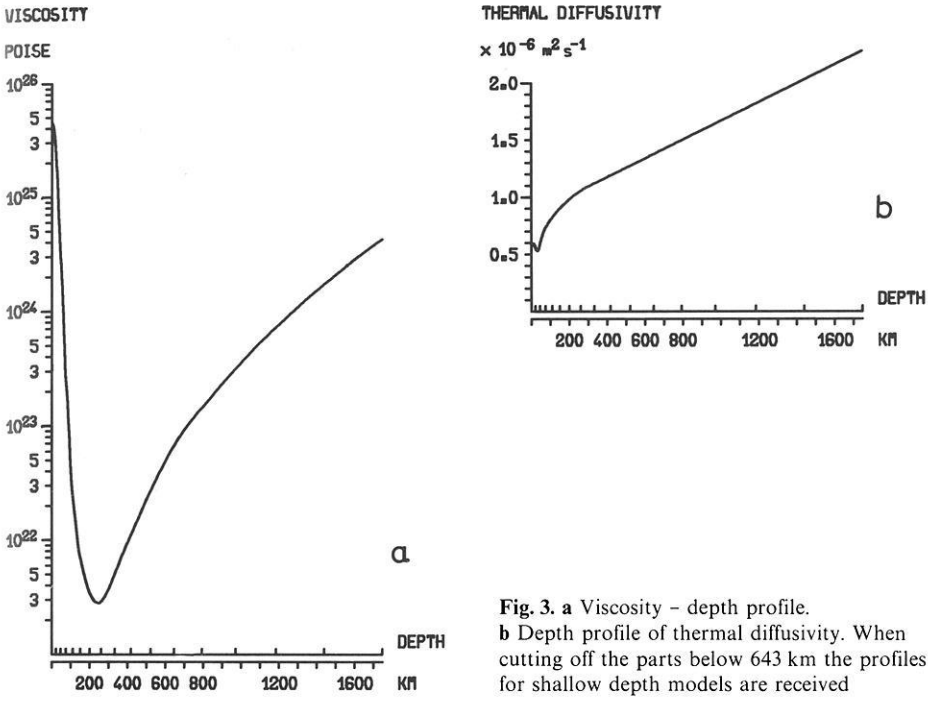


Fig. 3. a Viscosity - depth profile. **b** Depth profile of thermal diffusivity. When cutting off the parts below 643 km the profiles for shallow depth models are received

much thinner, and at the axis itself hot partially molten material rises up nearly to the surface. Therefore a horizontal layering of viscosity seems inappropriate in that region. A 'weak zone' is introduced at the margin, that means that the viscosity in the uppermost 100–150 km is reduced with respect to the normal lithosphere at the middle of the plate. At the spreading center, the asthenospheric minimum value of $3 \cdot 10^{21}$ Poise is continued up to the surface.

In the trench region, another deviation from horizontal layering occurs, because the cold highly viscous lithospheric slab descends at least several hundred kilometers into the mantle. On the other hand, there are some arguments which support the assumption that the effective viscosity is not so high here as in the middle part of the plate. The bending of the plate during the initial stage of subduction can be assumed to produce high stresses. The influence of that increased stress level is difficult to calculate but it is likely that it would lower the effective viscosity. If one assumes an elasto-plastic or brittle lithosphere, the plate would become weaker, since plastic flow, transient creep, or fracture would occur. If - on the other hand - a purely viscous model of the lithosphere with a nonlinear behavior following the law $\dot{\varepsilon} \sim \sigma^3$ is adopted, the effective viscosity is given by

$$\eta_{\text{eff}}(\sigma) = \sigma/\dot{\varepsilon} \sim 1/\sigma^2.$$

Thus it decreases rapidly if the stress level is increased. In the model viscosity distribution the trench region of the lithosphere has a reduced value of about

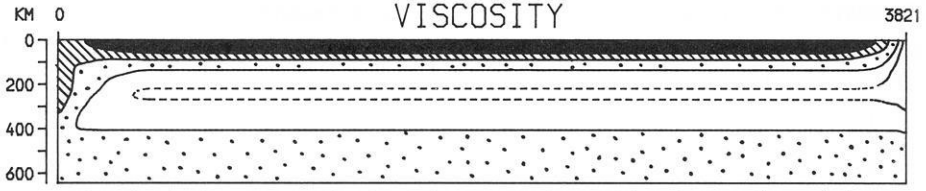


Fig. 4. Viscosity distribution used for the shallow convection models. The spreading axis is on the right and the subduction zone on the left. Viscosity is above 10^{24} Poise in the black area, between 10^{23} and 10^{24} in the hatched, between 10^{22} and 10^{23} in the dotted, and below 10^{22} Poise in the white area. Between the dashed lines in the white region it is below $3 \cdot 10^{21}$ Poise. In deep convection models the same distribution is adopted, completed by horizontally layered viscosity beneath 643 km according to the profile in Fig. 3a

10^{23} Poise. The whole viscosity distribution is shown in Fig. 4. We can hope, that this concept of weak zones leads to a more realistic behavior of the surface plate than in previous models with horizontally layered viscosity (DeBremaecker, 1977a).

6.2. Thermal Diffusivity

The depth dependence of the thermal diffusivity was calculated according to the formulae given by Schatz and Simmons (1972). In order to get better agreement with the experimental results of Kanamori et al. (1968), the term for the photon conductivity was slightly altered (e.g., by taking a higher threshold temperature). In this way we get slightly reduced values in the upper lithosphere and slightly higher values below than by using the original formula. For the oceanic crust, the diffusivity was chosen to have a constant value of $0.6 \cdot 10^{-6}$ m^2/s. The diffusivity profile is shown in Fig. 3b.

6.3. Heat Sources

The total amount of heat generation was determined in order to produce a mean surface heat flux of 1.5 HFU ($62\,mW \cdot m^{-2}$). In the shallow convection model, it is assumed that half of the heat is produced inside the cell and that the other half comes from below, thus assuming that the upper mantle is enriched with radioactive elements. For the deep mantle models, homogeneous distribution throughout the whole mantle is assumed, and the heat flux from the core is fixed to be 10% of the earth's total heat. Then 2/3 of the heat is produced in the cell and 1/3 comes from below. The absolute values are listed in Table 3.

6.4. Depth and Length of the Cell

The depth of the shallow convection cell is limited to 643 km. This is near the spinel-oxide phase boundary ($\simeq 670$ km) which might possibly confine the depth of upper mantle convection.

The depth in deep mantle convection models is 1,748 km which is about 60 % of the mantle depth. This arbitrary confinement ('intermediate convection model' in a medium position between upper mantle and whole mantle convection) is mainly done in order to restrict the computational expense. It would be justified if the high viscosity in the lower mantle would cause that the return flow penetrates only intermediate depth or plays only a minor role in the lowermost mantle.

However, in some manner this model may nevertheless be considered as a model of whole mantle convection. That is because a 1,750-km-deep plane mantle has nearly the same volume as a 2,900-km-deep spherical mantle with the same surface area. In mantle convection the heating of the cell is mainly a volume process, but the cooling is a surface process. If a flat mantle model is used instead of a spherical mantle, it seems best from the thermal point of view to choose the surface and the volume equal to the spherical values.

Of course, in the spherical geometry the stream pattern would be changed which influences the temperatures as well. Therefore the interpretation of the model as whole mantle convection remains doubtful – some consequences are mentioned in the discussion part (Chap. 8). However, I assume that – if you want to avoid the several difficulties connected with a spherical model and use a plane model – it is better to adjust the volume to the true value and not the depth extent.

The length of the cell is fixed to be 3,821 km. This is about the intermediate range of the earth's plates, for example the Nazca plate is approximately of this length.

6.5. Velocity of the Plate

The mean plate (or surface) velocity is, of course, not an input parameter of the model. But to make the different models comparable, they should have a similar plate velocity. Furthermore, to make reasonable statements about the mantle temperatures, the plate velocity should have a value which may be considered as the mean value of the earth's plate system.

This mean value is determined as follows: The annual rate of sea-floor spreading over the whole earth is about 2.0–3.5 km^2/a, Dickinson and Luth (1971) assume 1.7–2.3 km^2/a, Bickle (1978) gives 2.7–3.3 km^2/a and Garfunkel (1975) calculated 3.15 km^2/a. If a value of 2.9 km^2/a is taken and the earth's surface area is divided by it, we get as mean overturn time of a lithospheric plate 176 m.y. The 3,821-km-long model plate shall have the same overturn time and thus a velocity of 2.17 cm/a.

The Rayleigh number in my model calculations was always fitted in such a way that a mean surface velocity between 2.1 and 2.2 cm/a resulted. This is allowed since the mean viscosity of the earth's (upper) mantle is uncertain in several orders of magnitude. In all the models the Rayleigh number had to be increased by a factor of about 4 compared with its standard value, corresponding to a reduction of the all-over-viscosity by a factor of 4 with respect to the profile shown in Fig. 3a.

Table 3. List of model parameters. If more than one value is given the first belongs to the shallow model and the second to the deep model. The appendix "*ref.*" means "reference value"

Depth	h	$= \frac{643}{1748}$ km	$= \frac{0.643}{1.748} \cdot 10^6$ m
Length	1	= 3821 km	$= 3.821 \cdot 10^6$ m
Aspect ratio	1/h	$= \frac{5.95}{2.19}$	
Viscosity, ref.	η_0	$= 10^{22}$ Poise	$= 10^{21}$ kg m^{-1} s^{-1}
Thermal diffusivity, ref.	κ_0	=	10^{-6} m^2 s^{-1}
Density, ref.	ρ_0	= 3.8 g/cm^3	$= 3.8 \cdot 10^3$ kg m^{-3}
Specific heat	c_p	= 0.287 cal/(g · grd)	$= 1.2 \cdot 10^3$ J kg^{-1} K^{-1}
Thermal expansivity	α	=	$3.0 \cdot 10^{-5}$ K^{-1}
Temperature difference, ref.	ΔT_0	=	2000 K
International heat production	Q_{int}	=	$\frac{4.88}{2.39} \cdot 10^{-8}$ W m^{-3}
Bottom heat flux	Q_b	$= \frac{0.75}{0.50}$ HFU	$= \frac{31.4}{20.9} \cdot 10^{-3}$ W m^{-2}
Rayleigh number, standard value	Ra_s	$= \frac{606155}{12195155}$	
Dissipation number	Di	$= \frac{0.1608}{0.4372}$	

6.6. Other Parameters

The other parameters of interest are less critical. In Table 3 all the parameters are listed.

7. Results

7.1. Shallow Mantle Convection (Model No. 20)

The results of the shallow mantle convection model is shown in Figs. 5–7, special values of interest are listed in Table 4 for all the models.

Surface Velocity and Heat Flux (Fig. 5). Except for the 260-km-wide marginal regions, the surface velocity is nearly uniform. It increases slightly towards the trench region, but only by 3%. The surface boundary layer in the model can be considered as a rigid plate in a very good approximation. Whether there is a slight deformation of real plates by travelling insignificantly faster near the trench cannot be resolved, but it does not seem impossible.

The heat flux profile is in good agreement with the 'reliable' heat flux averages for the North Pacific from Sclater and Crowe (1976) (rectangles in Fig. 5b).

Stream Pattern (Fig. 6a). The ascending and descending flow is restricted to narrow boundary layers; the ascending flow is even narrower and faster than the downgoing (maximum velocity 9.3 cm/a in the upwelling limb). Outside the

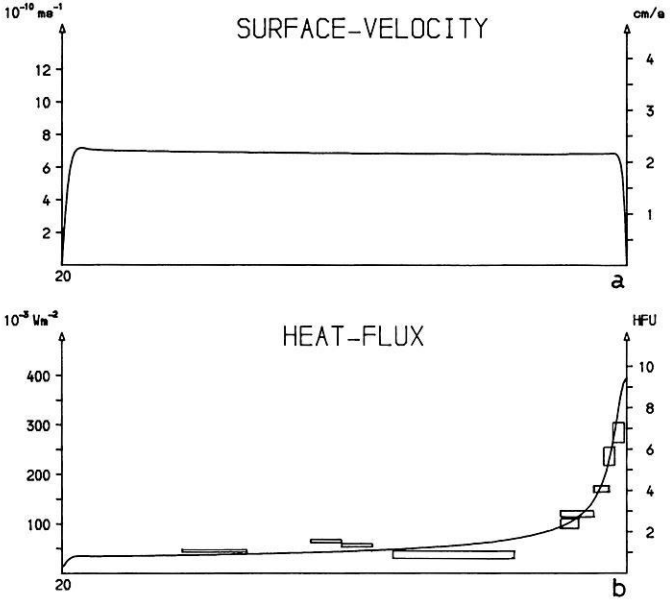


Fig. 5a and b. Model 20: Surface velocity **a** and surface heat flux **b** measured over the length of the plate. The spreading center is on the right. The rectangles in **b** are the 'reliable' heat flux averages from Sclater and Crowe (1976). Their original dependence on age was converted into length dependence using a constant plate velocity of 2.15 cm/year

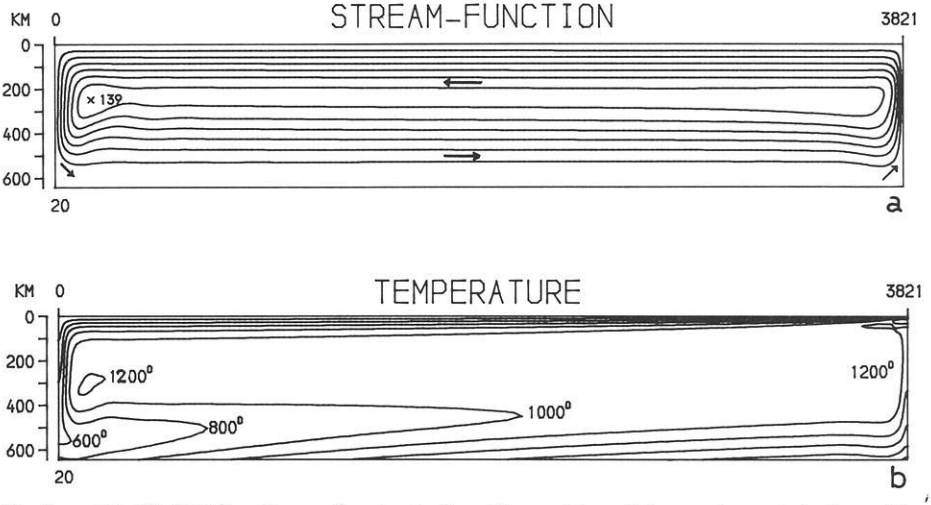


Fig. 6a and b. Model 20: **a** *Stream function isolines.* The position of the maximum is designated by an *x*. Lines are drawn every 20 (dimensionless) units. **b** *Temperature isolines.* Temperature is in °C, every 200° one line is drawn

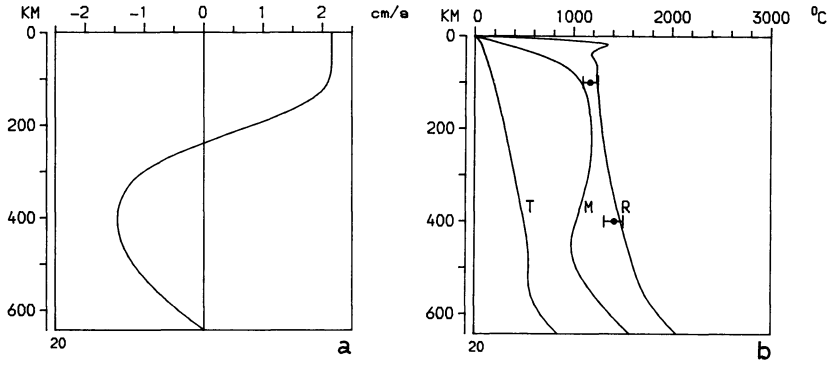


Fig. 7a and b. Model 20: **a** *Depth profile of horizontal velocity.* The profile is taken in the middle of the cell at half the length between spreading center and subduction zone. **b** *Temperature-depth profiles.* The profiles are taken in the trench region (*T*) at the length $l = 0$ km, in the middle of the cell (*M*) at $l = 1910$ km, and at the ridge (*R*) at $l = 3821$ km. Over most of the horizontal extent of the cell the profile is similar to the middle geotherm, only at the margins significant deviations occur. The following 'observational' values are included: 100 km: 1,100°–1,250° (Mercier and Carter, 1975) pyroxene-geothermometry; 400 km: 1,300°–1,500° (Gebrande, 1975) olivine-spinel-transform.

marginal regions, the flow is nearly horizontal, the vertical velocity remains below 0.1 cm/a almost everywhere.

In Fig. 7a the depth profile of the horizontal velocity in the middle of the cell is shown. The velocity remains constant in the lithosphere down to 100 km and then it begins to decrease. This contradicts the idea that the plate might be driven by an asthenospheric flow because in this case the velocity should first *increase* below the plate. The depth of reversal, where the plate - directed flow turns into a counter flow, is 240 km which is very near the depth of the minimum of viscosity (245 km).

Temperatures (Figs. 6b and 7b). The temperature field is characterized by a steep increase with depth in the lithosphere, an almost isothermal region from the bottom of the plate down to about 350 km over nearly the whole length of the cell, narrow boundary layers at the margins and a warm boundary layer at the bottom. A tongue of cold material extends horizontally from the descended lithosphere into the cell at a depth of 400–550 km.

The temperature at the bottom of the lithosphere (1,100° C at 100 km depth in the middle of the cell) is in good accordance with the values derived from pyroxene geothermometry (see Chap. 2). This may especially indicate that the values of thermal conductivity in the lithosphere are well chosen. But the temperatures at greater depths are too low: The mean temperature at 400 km is 1,040° compared with the most realistic value of 1,350°–1,450° C. An important feature of the temperature distribution is an inversion (negative temperature gradient with the z-axis pointing down) which appears between approximately 200 and 550 km depth over most of the cell's length.

Both the inversion of the temperature gradient and the low temperatures at the 400-km level are produced because the cold descending lithosphere does not

heat up enough before beginning the return flow. Such a temperature inversion has important consequences, because it would inhibit small scale convection in the asthenosphere as it is proposed by Richter (1973) and Richter and Parsons (1975). They assumed long Rayleigh-Bénard rolls which are oriented perpendicular to the main flow connected with the plate motion. But these convection cells require of course a superadiabatic gradient which does not occur in this model.

7.2. Alternative Shallow Convection Models (Nos. 21 and 22)

So far the shallow convection model is quite satisfactory in reproducing the main features of plate tectonics and upper mantle dynamics as they are known today, but its mantle temperatures are too low. The difference to realistic values is too great (300°–400°) to be tolerated. Therefore in the following two models attempts were made to raise the temperature by certain modifications.

The thermal conductivity in the mantle is slightly uncertain, especially the influence of photon conductivity. In model 21[1] a reduction of the conductivity by a factor of 0.8 was made ($\kappa_0 = 0.8 \cdot 10^{-6}\,m^2/s$). This size of reduction seems to be within a reasonable range. In the case of purely conductive heat transfer it would rise the temperature by 25% all over. But in the model, the increase was less, the temperature rose about 160° beneath the plate, to 1,260° at the bottom of the lithosphere and to 1,200° C at 400 km. The latter value is still too low and the first value seems even to exceed slightly the range of pyroxene temperatures. Neither the stream pattern nor the general features of the temperature field are changed significantly compared with model 20.

In model No. 22[1] the heat sources were redistributed, assuming a constant rate of heat production in the whole mantle. Then 32% of the total amount of heat is produced inside the cell and 68% is coming from below. The mean temperature at 400 km *sank* to 1,015° C while the temperature at 100 km depth rose by 10° to 1,110°. The result seems to be paradox, but it can be explained by considering the increased velocity of the ascending limb (see Table 4). The additional heat from the bottom is transported more efficiently to the surface or into the lithosphere. The lack of internal heating on the other hand slightly reduces the temperature at intermediate depth. It can be stated that neither a redistribution of heat sources according to uniform heat production nor a change of thermal conductivity improves the upper mantle convection model.

7.3. Deep Mantle Convection Models (Nos. 30 and 31)

In model 30 the depth of the cell is increased to 1,748 km, but for the bottom the no slip boundary condition is applied, considering that perhaps the increase of viscosity in the lower mantle would inhibit flow in the lowermost mantle.

---

[1] Without figure, for further results see Table 4

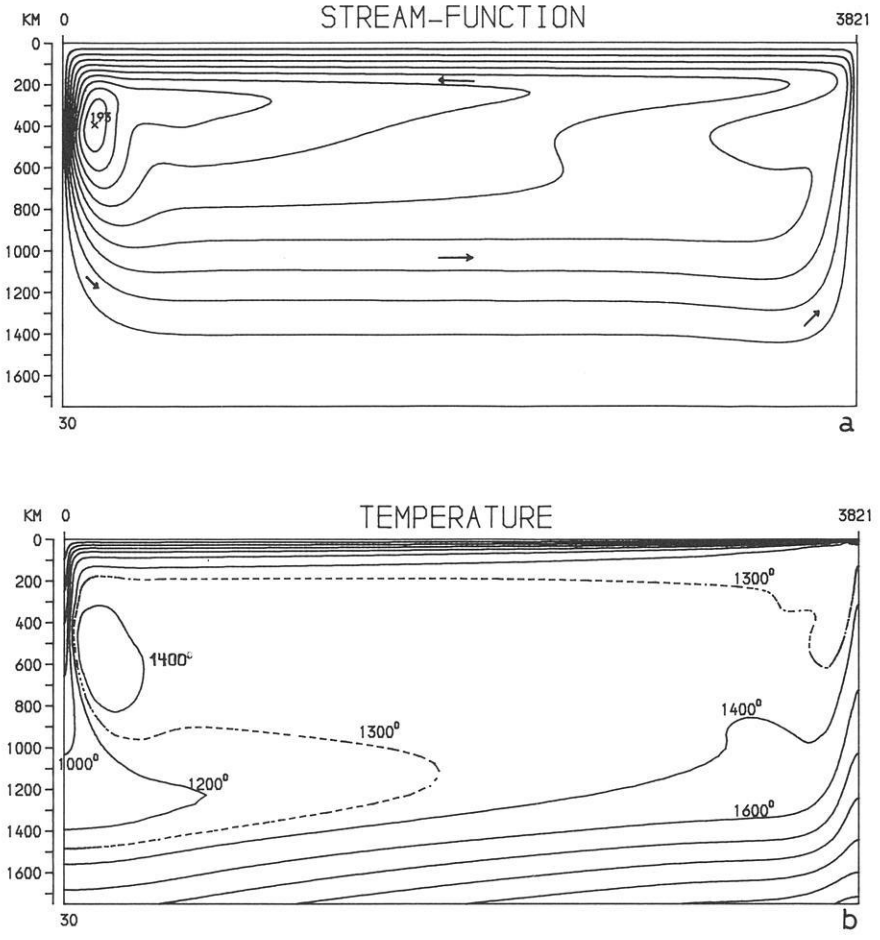


Fig. 8a and b. Model 30: **a** Stream pattern and **b** temperature field. One line is drawn every 20 units in **a** and every 200° in **b** respectively

The stream pattern (Fig. 8a) is now more complicated than in the shallow convection models. The maximum of the stream function is near the descending slab and much more striking, because an additional curl of flow is induced here. In the descending limb the flow is very rapid. The return flow is concentrated mainly between 800 and 1,500 km depth (Fig. 9a), while the reversal of flow (plate flow ↔ return flow) takes place between 200 and 300 km, which is nearly the same depth as in the shallow models. Between 400 and 800 km the return flow is quite slow and the flow tends to move upwards in this depth range even outside of the ridge region.

The surface velocity increases in this model by 11.5% over the length of the plate towards the trench. This value still seems to be tolerable for considering the surface layer as a rigid plate. This variation of velocity implies strong tensile stresses in the plate (see Fig. 10). Thus the plate should be mainly driven by the

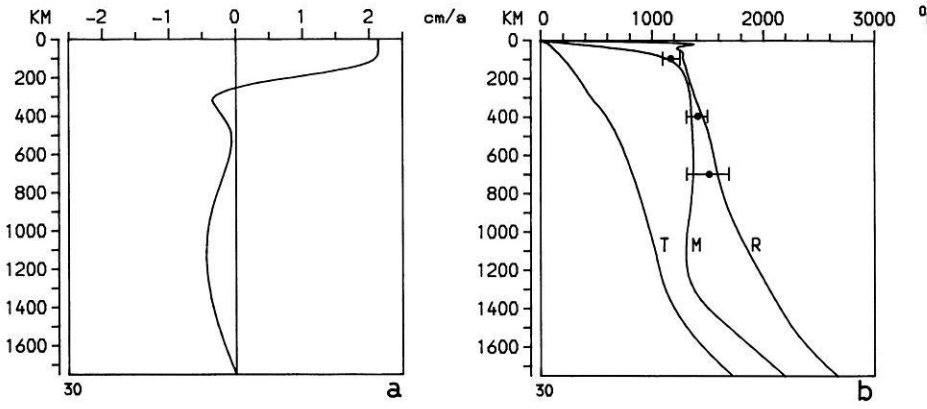


Fig. 9a and b. Model 30: **a** *Velocity-depth profile.* **b** *Temperature-depth profile,* T = trench region, M = middle, R = ridge region. Reference temperatures are included: 100 km: 1,100°–1,250° (Mercier and Carter, 1975) pyroxene-geothermometry; 400 km: 1,300°–1,500° (Gebrande, 1975) olivine-spinel-transformation; 700 km: 1,400°–1,700° (Watt and O'Connell, 1978) petrological model

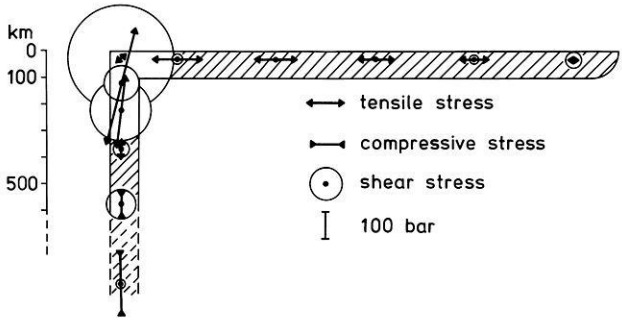


Fig. 10. Model 30: *Stress distribution in the lithosphere.* The plate is drawn schematically, the horizontal scale is reduced by a factor of 2. The normal stress in direction of the flow and the related shear stress are shown at 32 km depth (or distance from the edge of the subducted plate). It can be noted that the stress level is highest in the bending region of lithospheric subduction. This supports the argumentation for the reduction of viscosity in this region (see Chap. 6)

pull of its subducted part. Moreover, the stream pattern suggests that this pull is the main driving source of the flow in the whole cell. The heat flux profile differs only negligibly from the profiles of shallow mantle convection.

The temperature-depth profile is now much more satisfactory in the upper mantle (Fig. 9b). The mean temperature at 100 km depth is 1,180° and at 400 km 1,350° C, the mean value of the whole uppermost 400 km is 1,170° compared with the best value of 1,150° from electrical conductivity data (Tozer, 1970). These values are in good agreement with the data derived from observations, as discussed in Chap. 2. But in the lower mantle temperatures are below the classical estimates. Between 250 km and 1,225 km the temperature gradient is subadiabatic and in some parts even negative, and at 1,100 km depth there is a

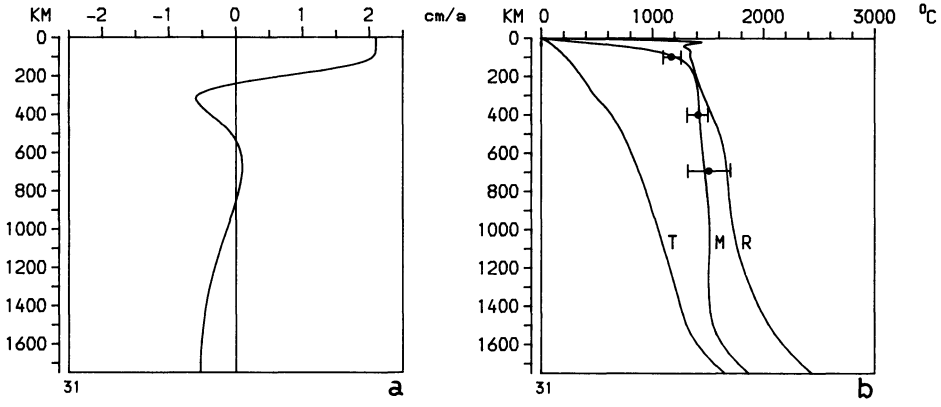


Fig. 11a and b. Model 31: **a** *Velocity-depth profile.* **b** *Temperature-depth profile,* T = trench region, M = middle, R = ridge. The same reference temperatures as in Fig. 9b are included. Compared with model 30 the boundary condition at the bottom is converted from 'no slip' to 'free slip'

minimum in the $T(z)$-profile of 1,300° C. At the bottom the mean temperature is 2,170° C.

It appears strange that there is a slight minimum of temperature in the upper mantle just beside the ascending flow beneath the ridge (Fig. 8b). It can be explained as follows: The cold tongue, extending from the subducted lithosphere on the left into the cell, does not heat up enough because of the very slow conductive heat transfer. When ascending again, it cools because of the adiabatic decompression, and it is then colder than the surrounding mantle. This cool lump near the spreading region seems to produce the strange stream pattern in the right side of the cell.

It was pointed out in Chap. 6.4 that a 1,750-km-deep convection model of a plane mantle may be considered to be similar to a 2,900-km-deep spherical mantle. For whole mantle convection, the boundary condition at the bottom shall be free slip. Thus in the next model everything remains unchanged, except a free slip condition at the bottom instead of 'no slip'. The stream pattern of model 31 is similar to the previous model, except for a greater depth-extent of the return flow which is concentrated between 1,000 km and the bottom. The general pattern of the temperature field is similar, too, except for a less developed lower boundary layer.

The maximum velocity of the return flow appears at the bottom (Fig. 11a). For that reason it is demonstrated that a viscosity of more than 10^{24} Poise in the lower mantle is not sufficient to confine convection to the upper mantle or even to concentrate the return flow at intermediate depth. Therefore the idea of 'intermediate depth mantle convection' (Chap. 6.4) seems to be very doubtful. From this point of view model 31 – considered to be similar to a whole mantle convection cell – seems to be more realistic than model 30 as an intermediate depth cell.

In model 31 the temperatures are slightly higher in most parts of the cell than in model 30 (Fig. 11b). At 400 km a mean temperature of 1,420° C is

Table 4. Model results

Model number	20	21	22	30	31
	Upper mantle convection			Deep mantle convection	
Pecularity	–	λ Reduced	Redistribut. heat sources	–	Free slip at the bottom
Ra/Ra_s	3.88	4.75	3.46	3.94	3.69
ψ_{max}	139	186 (149)	163	193	197
$\bar{v}_p$	2.17	2.18	2.15	2.15	2.11
v_d	−4.8	−5.7	−4.5	−8.3	−8.3
v_a	9.3	9.6	11.4	5.8	4.7
v_r	−1.46	−1.41	−1.44	−0.44	−0.53
Δv_p	3.1%	7.1%	0.7%	11.5%	12.1%
z_{rev}	239	238	238	254	242
T_{100}	1,100°	1,261°	1,109°	1,179°	1,247°
T_{400}	1,040°	1,200°	1,015°	1,350°	1,420°
T_{bot}	1,525°	1,695°	1,730°	2,170°	1,876°

Ra/Ra_s: Ratio between the actual Rayleigh number and its standard value
ψ_{max}: Maximum value of stream function $\bar{v}_p$ mean plate velocity in cm/year
v_a, v_d, v_r: Maximum velocity of ascending, descending and return flow in cm/year
Δv_p: Variation of the plate velocity over the plate length, the 260 km wide marginal regions are excluded
z_{rev}: Depth of reversal of flow direction in the middle of the cell (in km)
T_{100}, T_{400}, T_{bot}: Mean temperature at 100, 400 km depth and at the bottom (in °C)

reached, and in the lower mantle the gradient is slightly higher and the temperatures are about 200° above those of the previous model. But still the gradient is subadiabatic below 240 km down to about 1,450 km. The temperature at the bottom of the cell has a mean value of 1,880° which is by all means below the solidus of the core's material.

8. Conclusions and Discussion

Stream Pattern

The concept of weak zones at the active plate margins proved to be successful to make the highly viscous surface layer behave like a rigid plate. There are some plausible arguments to reduce the effective viscosity not only at the spreading center but also in the subduction region, but the weak zones have still to be justified by a more sophisticated model which regards the complicated temperature and stress controlled rheology of the lithosphere, which is not yet known in detail.

A viscosity of more than 10^{24} Poise in the lower mantle is not sufficient to inhibit deep mantle convection (this is already known from marginal stability analysis, too), but moreover, a moderate increase of viscosity with depth (by 3 orders of magnitude from the minimum to the bottom) cannot concentrate the

return flow to intermediate depth. Thus it seems very likely that the lower mantle is in convection if there are no other inhibiting factors, for example chemical heterogeneity.

The lack of earthquakes below 700 km may be caused by the transformation to a mixed-oxide or perovskite structure at about this depth. A changed mechanical behaviour or the elevated temperatures due to the transformation might remove the slab material out of its brittle region (Davies, 1977; O'Connell, 1977). The predominant compressive stress in the lower parts of the subducted plate must not indicate a barrier at 700 km. In fact the stress state in the descending flow is compressive below 450 km in the deep mantle models and below 350 km in the upper mantle model (see Fig. 10). However, reliable statements concerning this topic would require a model which simulates the descending behavior of the plate better than the present one.

Temperatures

The temperatures at the bottom of the lithosphere agree well with the data derived from pyroxene geothermometry for both the upper mantle and the deep convection model. In the upper mantle below the lithosphere differences occur: The deep convection models satisfy quite well the observational constraints (electrical conductivity, olivine-spinel transformation, petrological data), while in the shallow convection models the temperatures are too low, at the 400 km depth level by 300°–400°. From this point of view it seems very likely that convection is not confined to the upper mantle but is present throughout the whole mantle.

In all the models there is a subadiabatic temperature gradient at intermediate depth. In probably the best model (No. 31) the gradient is higher than in all the others, but it is still below the adiabatic gradient for depths greater than 250 km. If this is true in the mantle, there is only little foundation for small scale convection cells proposed by Richter (1973) and others. A type of small-scale convection initiated by boundary layer instability (Parsons and McKenzie, 1978) might still be possible, but it is doubtful whether it plays an important role, since it would hardly have a big depth-extent and since the Rayleigh number does not exceed its critical value very much.

Some difficulties still remain because of the cool lower mantle in the models. The liquidus of the outer core alloy must be exceeded. This is a strong constraint to the temperatures in the lowermost mantle. Even with the optimal composition (Fe_2S) the temperature at the core-mantle boundary should be at least 2,500° C, but it is more likely that the limit is even higher. Therefore the temperatures at the bottom of the deep mantle models are several hundreds or a thousand degrees below the required values. Several circumstances, which are not regarded in the models, may contribute to raise the temperature of the lower and lowermost mantle.

(a) The olivine-spinel phase boundary has a positive Clapeyron slope. This implies the release of latent heat when material crosses the boundary down-

wards and the lower mantle would be heated. A temperature rise in the order of 100° can be estimated.

(b) Higher temperatures in the lower mantle can be expected if the correct spherical geometry and the true depth of 2,900 km are included in the model. The rise of temperature with increasing depth of the convection cell is shown by comparison between shallow and deep convection models. But another special effect concerning the lower boundary layer results. A spherical mantle sector has only 30% of its surface area at the bottom while the top and bottom area are equal in the plane model. Therefore the same amount of heat coming from the core would imply a heat flux (per unit area) thrice the high in a spherical model, and the thermal gradient at the bottom would be thrice the gradient of the flat earth model. Thus a significant heat flux from the core is producing an important temperature rise in the lower thermal boundary layer of the cell, which might protect the outer core from the low temperatures in the mantle.

The D″-layer at the bottom of the mantle may be identified as this lower boundary layer (Jones, 1977). The other possible interpretation of the D′-layer as chemically different from the rest of the mantle (e.g., iron-enriched) also leads to a shielding of the core from low mantle temperatures because in this case it could not participate in convection. The core's heat must then be transported throughout it by conduction and also a steep temperature rise would be the consequence.

(c) The low model temperatures at the bottom can be caused by the symmetric lateral boundary conditions, which lead to double dipping slabs. Thus the two plates shield each other from the warm surrounding mantle and those parts which finally reach the bottom remain coldest. (One might argue that the whole temperature distribution, including subadiabatic gradients and low temperatures in the lower mantle, would be changed in favor of higher values, if there would be a single descending slab. Of course, the slab is then heated from both sides. However, it must be considered that the rate of subduction always equals the spreading rate and therefore the single slab has to have twice the descending velocity of the double slabs, and the increased heating would be compensated).

It does not seem likely that all these effects would raise the temperature by considerably more than a thousand degrees. Thus temperatures in the lower mantle would remain quite low, and at the core-mantle boundary about 3,000° C can be expected. Therefore the model results support the hypothesis of sulphur being the light alloying element in the core.

Another difficulty arising from such low temperatures in the mantle is the high viscosity which would result from Eq. 6. The increase of viscosity due to the rising pressure can not be compensated by rising temperature. But theoretical considerations suggest that the activation volume does not remain constant but decreases considerably with increasing pressure. According to Sammis et al. (1977), it falls from 11 cm^3/mol at the surface down to 2.5–5 cm^3/mol at the bottom of the mantle. Thus the effect of pressure could be less important and the viscosity could remain low enough to allow whole mantle convection despite the rather low temperature in the lower mantle.

Acknowledgement. I am grateful to the Deutsche Forschungsgemeinschaft, which is financing my work.

References

Andrews, D.J.: Numerical simulation of sea-floor spreading. J. Geophys. Res. **77**, 6470–6481, 1972

Bickle, M.J.: Heat loss from the earth: a constraint on Archean tectonics from the relation between geothermal gradients and the rate of plate production. Earth Planet. Sci. Lett. **40**, 301–315, 1978

Davies, G.F.: Whole-mantle convection and plate tectonics. Geophys. J. R. Astron. Soc. **49**, 459–486, 1977

DeBremaecker, J.-Cl.: Convection in the earth's mantle. Tectonophysics **41**, 195–208, 1977a

DeBremaecker, J.-Cl.: Is the oceanic lithosphere elastic or viscous? J. Geophys. Res. **82**, 2001–2004, 1977b

Dickinson, W.R., Luth, W.C.: A model for plate tectonic evolution of mantle layers. Science **174**, 400–404, 1971

Elsasser, W.M., Olson, P., Marsh, B.D.: The Depth of Mantle Convection. J. Geophys. Res. **84**, 147–155, 1979

Fusijawa, H.: Temperature and discontinuities in the transition layer within the earth's mantle: geophysical application of the olivine-spinel transition in the Mg_2SiO_4—Fe_2SiO_4 system. J. Geophys. Res. **73**, 3281–3294, 1968

Garfunkel, Z.: Growth, shrinking, and long-term evolution of plates and their implications for the flow pattern in the mantle. J. Geophys. Res. **80**, 4425–4432, 1975

Gebrande, H.: Ein Beitrag zur Theorie thermischer Konvektion im Erdmantel mit besonderer Berücksichtigung der Möglichkeit eines Nachweises mit Methoden der Seismologie. Universität München: Diss. Inst. für Allgemeine und Angewandte Geophysik, 1975

Graham, E.K.: Elasticity and composition of the upper mantle. Geophys. J. R. Astron. Soc. **20**, 285–302, 1970

Graham, E.K., Dobrzykowski, D.: Temperatures in the mantle as inferred from simple compositional models. Am. Mineral. **61**, 549–559, 1976

Higgins, G., Kennedy, G.C.: The adiabatic gradient and the melting point gradient in the core of the earth. J. Geophys. Res. **76**, 1870–1878, 1971

Houston, M.H., DeBremaecker, J.-Cl.: Numerical models of convection in the upper mantle. J. Geophys. Res. **80**, 742–751, 1975

Jones, G.M.: Thermal interaction of the core and the mantle and long-term behaviour of the geomagnetic field. J. Geophys. Res. **82**, 1703–1709, 1977

Jordan, T.H.: Lithospheric slab penetration into the lower mantle beneath the Sea of Okhotsk. J. Geophys. **43**, 473–496, 1977

Kanamori, H., Fujii, N., Mizutani, H.: Thermal diffusivity measurements of rock-forming minerals from 300° to 1100° K. J. Geophys. Res. **73**, 595–605, 1968

Leppaluoto, D.A.: Melting of iron by significant structure theory. Phys. Earth Planet. Inter. **6**, 175–181, 1972

McGregor, I.D., Basu, A.R.: Thermal structure of the lithosphere: a petrological model. Science **185**, 1007–1011, 1974

McKenzie, D.P., Roberts, J.M.: Convection in the earth's mantle: towards a numerical simulation. J. Fluid Mech. **62**, 465–538, 1974

Mercier, J.-Cl., Carter, N.L.: Pyroxene geotherms. J. Geophys. Res. **80**, 3349–3362, 1975

O'Connell, R.J.: On the scale of mantle convection. Tectonophysics **38**, 119–136, 1977

Parsons, B., McKenzie, D.: Mantle convection and the thermal structure of plates. J. Geophys. Res. **83**, 4419–4430, 1978

Prenter, P.M.: Splines and variational methods. New York: John Wiley and Sons 1975

Richter, F.M.: Convection and large scale circulation of the mantle. J. Geophys. Res. **78**, 8735–8745, 1973

Richter, F.M., McKenzie, D.: Simple plate models of mantle convection. J. Geophys. **44**, 441–471, 1978

Richter, F.M., Parsons, B.: On the interaction of two scales of convection in the mantle. J. Geophys. Res. **80**, 2529–2541, 1975

Ringwood, A.E.: Composition and petrology of the earth's mantle. New York: McGraw-Hill 1975
Sammis, C.G., Smith, J.C., Schubert, G., Yuen, D.A.: Viscosity-depth profile of the earth's mantle: effects of polymorphic phase transitions. J. Geophys. Res. **82**, 3747–3761, 1977
Sato, A., Thompson, E.G.: Finite element models for creeping convection. J. Comput. Phys. **22**, 229–244, 1976
Schatz, J.F., Simmons, G.: Thermal conductivity of earth materials at high temperatures. J. Geophys. Res. **77**, 6966–6983, 1972
Schubert, G., Froidevaux, C., Yuen, D.A.: Oceanic lithosphere and asthenosphere: thermal and mechanical structure. J. Geophys. Res. **81**, 3525–3540, 1976
Sclater, J.G., Crowe, J.: On the reliability of oceanic heat flow averages. J. Geophys. Res. **81**, 2997–3006, 1976
Stacey, F.D.: A thermal model of the earth. Phys. Earth Planet. Inter. **15**, 341–348, 1977
Tolland, H.G.: Thermal regime of the earth's core and lower mantle. Phys. Earth Planet. Inter. **8**, 282–286, 1974
Torrance, K.E., Turcotte, D.L., Hsui, A.T.: Convection in the earth's mantle. In: Methods in computational physics, Vol. 13, B.A. Bolt, ed.: pp. 431–454. New York, London: Academic Press 1973
Tozer, D.C.: Temperature, conductivity, composition and heat flow. J. Geomagn. Geoelectr. **22**, 35–51, 1970
Walcott, R.I.: Flexural rigidity, thickness, and viscosity of the lithosphere. J. Geophys. Res. **75**, 3941–3954, 1970
Wang, C.Y.: Temperature in the lower mantle. Geophys. J. R. Astron. Soc. **27**, 29–36, 1972
Watt, J.P., O'Connell, R.J.: Mixed-oxide and perovskite-structure model mantles from 700–1200 km. Geophys. J. R. Astron. Soc. **54**, 601–630, 1978
Zienkiewicz, O.C.: The finite element method, 3rd edn. London: McGraw-Hill 1977

Received February 19, 1979; Revised Version June 22, 1979; Accepted August 10, 1979

J. Geophys. 46, 123–133, 1979

Journal of Geophysics

Detection Probabilities for Weak Regional Seismic Events

A. Shapira[1], O. Kulhánek[2], and R. Wahlström[2]

[1] Department of Applied Mathematics, The Weizmann Institute of Sciences, Rehovot, Israel
[2] Seismological Institute, Box 517, S-75120 Uppsala, Sweden

Abstract. The direct estimation method for threshold magnitudes associated with teleseismic events is well-known and frequently used. A modification of the method is developed in order to make the approach applicable to events from regions with low seismicity. It is assumed that observations are supplied by a seismograph network located within the region. As an example, threshold magnitudes are given for the Uppsala station, Sweden. Reliability of the suggested approach is examined by means of a rockburst series.

Key words: Detection probability – Earthquakes in Sweden – Uppsala seismograph station – Weak regional Seismic events – Operational threshold magnitude.

1. Introduction

Capability of a network to detect weak regional events is essential for earthquake prediction and risk analysis. Besides engineering aspects, detectability estimates are of great importance for many seismological studies since they provide a measure of operational efficiency of the station and/or of the network *wrt* the given type of events.

Investigations considering teleseismic epicentral distances were carried out by Kelly and Lacoss (1969), Ringdal (1975), von Seggern and Blandford (1976), Pirhonen et al. (1976), Ringdal et al. (1977), among others. In the present work, we are concerned with the problem of evaluating the detection capability for weak regional events by means of a national seismograph station network. By regional events we mean events recorded merely by regional seismic networks. In general, we follow the direct estimation approach which compares the detection performance of a station in question with the detection performance of an independent reference system like e.g. ISC or NEIS. When considering events of low magnitudes, an immediate application of the direct method is not possible due to the lack of a reference system, because regional events are recorded by a very limited number of stations. Another problem emerges from the fact that the size of regional

0340-062X/79/0046/0123/$02.20

earthquakes is determined by means of a regional magnitude scale. Consequently, the detectability given in terms of magnitude thresholds is also dependent upon the procedures applied for the magnitude determination.

The station detectability of teleseismic events is expressed as a function of the reference system magnitude (hereafter, reference magnitude) and is related to a certain limited seismic region. However, in cases like, e.g., Swedish earthquakes any division of the region in a number of subregions is difficult. Moreover, for regions with low seismicity, additional grouping of events might jeopardize the statistical representation. Therefore, rather than relating the detectability of a station to a certain subregion we shall relate it to the epicentral distance.

Below, we present a modified direct estimation method adapted to Swedish regional conditions. As a quantitative example, threshold magnitudes are given for the seismograph station Uppsala (UPP). Results are examined by means of detectability estimates of rockbursts in Grängesberg (central Sweden). Although the observations treated in this work are limited to a specific region and to a specific network, the applicability of the method is rather general.

2. Modified Direct Estimation Method

Event detectability, of a particular seismograph station for a given seismic region, is expressed by means of the incremental detection probability as a function of magnitude. Ringdal (1975) and von Seggern and Blandford (1976) discuss the detection probability in greater detail. Here, we follow the approach of Ringdal.

2.1. Statistical Model

The statistical model used is based upon the following assumptions: Firstly, the reference magnitude, M_R, and the station magnitude, M_A, are distributed about the same true unknown magnitude, M, so that $M_R \sim N(M+C_R, \sigma_R^2)$, $M_A \sim N(M+C_A, \sigma_A^2)$. C and σ denote the magnitude bias and the magnitude variance, respectively. The threshold magnitude, M_t, is normally distributed with mean value μ and with variance σ^2, i.e., $M_t \sim N(\mu, \sigma^2)$. Secondly, M_A and M_R as well as M_A and M_t are mutually independent. Lastly, the number of detected earthquakes, N_c, exceeding a magnitude M_R, is expressible through the magnitude-frequency relationship

$$\log_e[N_c(M_R)] = \alpha - \beta M_R. \tag{1}$$

By accepting the above assumptions, the estimated probability of detection is (Ringdal, 1975)

$$P(M_R) = \Phi\left[\frac{M_R - b}{s}\right] \tag{2}$$

where Φ is the Gaussian cumulative distribution function and

$$\begin{aligned} b &= \mu + C_R - C_A + \beta\sigma_R^2 \\ s^2 &= \sigma^2 + \sigma_A^2 + \sigma_R^2. \end{aligned} \tag{3}$$

The observed detection curve, i.e., the percentage of events detected by the station, for different magnitude levels, is fitted to the $P(M_R)$-function in Eq. (2). The curve-fitting is carried out, e.g., by using the probit analysis technique (IBM Application Program, 1970) which yields the maximum likelihood estimate of b and s.

2.2. Operational and True Threshold Magnitudes

Following Eqs. (2) and (3) we may refer to three different threshold magnitudes:

(a) True threshold magnitudes, μ, which are more of theoretical importance but may also be used when comparing the detectability of various stations even from different networks. However, since our data are limited to regional observations, determination of C_R, and therefore also of μ in Eq. (3), is not possible (for details see Shapira et al., 1978).

(b) Biased or operational threshold magnitudes, $b \pm s$. These are directly obtained by fitting the observations to the function $P(M_R)$ in Eq. (2).

(c) Corrected operational threshold magnitudes, $\hat{b} \pm \hat{s}$, defined as

$$\begin{aligned} \hat{b} &= b + C_A = \mu + C_R + \beta \sigma_R^2 \\ \hat{s}^2 &= s^2 - \sigma_A^2 = \sigma^2 + \sigma_R^2. \end{aligned} \tag{4}$$

In this case we assume that $C_R + \beta \sigma_R^2$ remains constant, irrespective of the investigated station. Hence, $\hat{b} \pm \hat{s}$ can be used when comparing the detectability of stations within the network.

2.3. Reference Magnitude

For weak regional events we have to accept the regional network or part of it as the reference system. All magnitudes discussed throughout this work refer to regional magnitudes, M, defined for the region in question. Accordingly, we here define the reference magnitude of an event as

$$M_R = \frac{1}{K} \sum_{i=1}^{K} M_i \quad \text{for } i \neq A \tag{5}$$

where K is the number of available station magnitudes (investigated station A excluded) for the event and M_i is the magnitude of the event measured at the i-th station. The exclusion of M_A in Eq. (5) is due to the required independence between M_R and M_A.

2.4. Relocation of Earthquakes

Our intention is to study the detection probability of the station A as a function of M_R for a chosen epicentral distance. To introduce the distance dependence, we transform all available earthquakes to a common epicentral distance, say Δ_k. Let us

consider an earthquake of a known magnitude, M_R, at distance Δ from station A. We use the magnitude definition

$$M = f(D, T) + g(\Delta) \tag{6}$$

where D is the maximum ground amplitude, T is the corresponding period and $g(\Delta)$ is the calibrating function. A fictitious earthquake at distance Δ_k which causes a ground motion at A, with the same maximum amplitude, D, and period, T, will have, as follows from Eq. (6), the reference magnitude

$$M_R(\Delta_k) = M_R + g(\Delta_k) - g(\Delta). \tag{7}$$

Employment of Eq. (7) enables to determine magnitudes for fictitious events located at any chosen epicentral distance.

Throughout this work the decision 'detected' or 'not detected' is based upon the conserved parameters D and T. Thus, when the actual earthquake is detected by station A, it is assumed that all relocated earthquakes associated with this actual event are also detected. On the other hand, if the actual event is not detected, it means that none of the corresponding fictitious earthquakes are detected by A.

3. Detection Capability of UPP – A Quantitative Example

To demonstrate the applicability of the modified direct estimation method we determine the capability of UPP to detect Swedish events. UPP, the sole manned station within the Swedish seismograph station network (SSSN), is used for quick preliminary event locations and magnitude determinations. Thus, the knowledge of corresponding threshold magnitudes is of great practical importance.

3.1. Observational Material

The data used comprises 72 Swedish earthquakes detected during 1963–1972 (Kulhánek and Wahlström, 1977). By careful selection, man-made and possibly triggered seismic events (e.g., rockbursts) were excluded. All measurements, phase identifications and event discriminations have been carried out by one and the same seismologist (R.W.). In this respect, the data employed are homogeneous. It should be noted that a great majority of the 72 earthquakes were located merely by the SSSN. Only occasionally, closely located Finnish or Norwegian stations also contributed with readings. Epicentral and seismograph station (SSSN) locations are shown in Fig. 1.

3.2. Magnitude Bias and Variance

A regional magnitude scale for the Scandinavian area has been developed in Båth et al. (1976). The station-magnitude formula reads

$$M = \log_{10}(100\,\mathrm{D}) + F(\Delta, T) \tag{8}$$

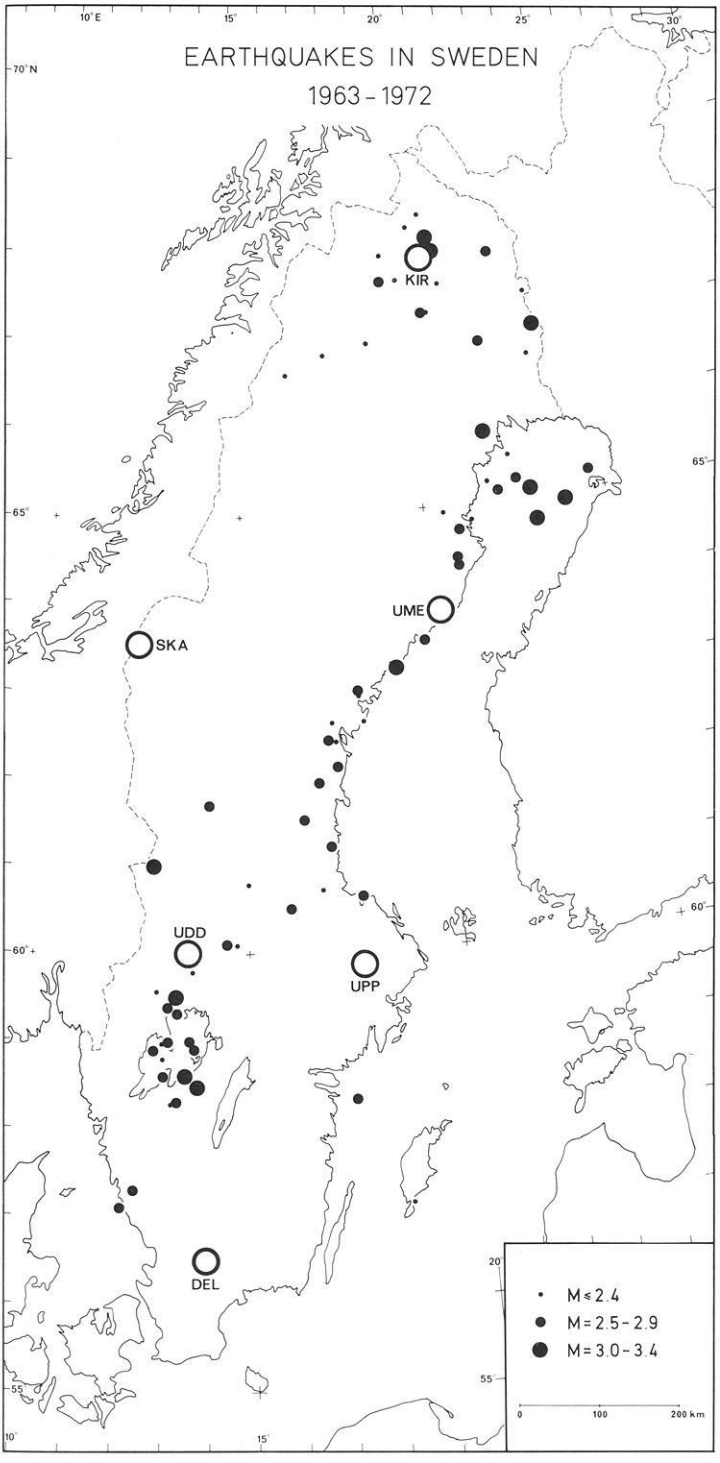


Fig. 1. Map showing the location of seismograph stations (*open circles*) and epicentres (*solid circles*) used in this paper

where D is the maximum ground amplitude of Sg in microns, measured on the short-period vertical-component seismogram. $F(\Delta, T)$ is the calibrating function available in tabular form (Båth et al., 1976) for $50\,\text{km} \leqq \Delta \leqq 1{,}500\,\text{km}$ and $0.3\,\text{s} \leqq T \leqq 1.4\,\text{s}$. As follows from Eq. (8), the calibrating function includes also the T parameter. Thus, Eq. (7) will be written as

$$M_R(\Delta_k) = M_R + F(\Delta_k, T) - F(\Delta, T). \tag{9}$$

Reference magnitudes are determined by applying Eq. (5), excluding UPP magnitude determinations. The station-magnitude bias, C_A, is determined by averaging the difference $M_A - M_R$ over all events with magnitudes reported by four or more stations. Then, $C_A = -0.06$ for $A = \text{UPP}$. The standard deviation of M_A *wrt* M_R is influenced by the number of stations, K, which contribute to estimations of M_R. Therefore, we define the variance as

$$s_R^2(K) = \frac{1}{N_K(K-1)} \sum_{j=1}^{N_K} \sum_{i=1}^{K} (M_{ij} - M_{Rj})^2 \tag{10}$$

where N_K is the number of events with magnitudes reported from K stations (UPP excluded), i is the station and j the event index, respectively. Numerical $s_R(K)$-values, for $K = 2, 3, 4, 5$ show that the relation $s_R^2(1) = K\, s_R^2(K)$ holds. The consistent $s_R(1)$-values are assumed to be good estimates of station magnitude variances, *wrt* the defined M_R, for any station within the SSSN. From our data we obtained $s_R(1) = 0.35$.

For those cases when the earthquake was not detected by UPP, the period, T, in Eq. (9) cannot be measured and has to be estimated. The present data do not confirm the expected pronounced correlation between the period and the magnitude and/or between the period and distance. The observations show that $T = (0.51 \pm 0.08)$ s. The mean value of 0.5s is therefore used as the period for events which were not detected by UPP. The error (two standard deviations) in period measurements of about ± 0.15s introduces inaccuracies into the $M_R(\Delta_k)$ determination which increase with distance. Using the $F(\Delta, T)$-table of Båth et al. (1976) and assuming $T = (0.50 \pm 0.15)$ s, we obtain the following $\sigma_A(\Delta)$ values

Δ	50–200 km	250–500 km	550–1000 km
$\sigma_A(\Delta)$	0.15	0.25	0.30.

We shall provide for the period effect by introducing a distance-dependent variance $\sigma_A^2 = (0.35)^2 + \sigma_A^2(\Delta)$, where here $\sigma_{\text{UPP}} = \sigma_A$.

3.3. Analysis and Results

Threshold magnitudes b_{50}, b_{90} and $\hat{b}_{50}$, $\hat{b}_{90}$ (indices denote probability level) are determined by the following procedure:

(a) All earthquakes are relocated to a common epicentral distance, Δ_k, from UPP. Magnitudes, $M_R(\Delta_k)$, of fictitious, i.e., relocated earthquakes are calculated according to Eq. (9).

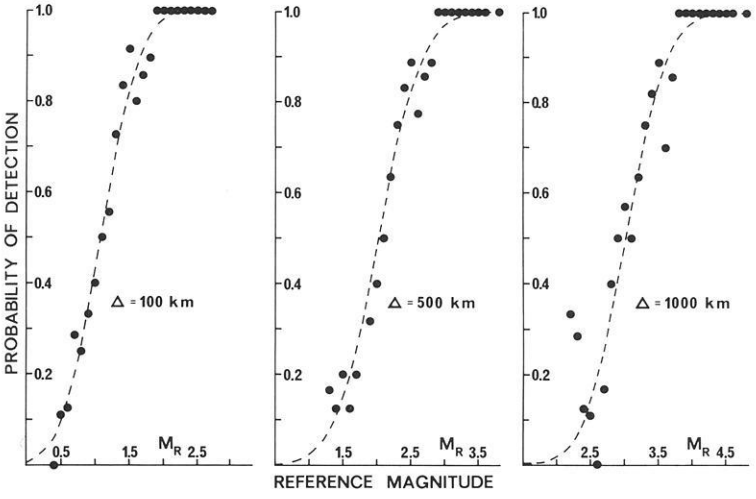


Fig. 2. Matched detection curves, $P(M_R)$, (*dashed lines*) and observed detection probabilities (*solid circles*) for UPP and epicentral distances of 100, 500, and 1000 km

(b) The observed detection probability for a chosen M_R is determined as the number of the detected events normalized by the total number of events. Magnitude intervals of 0.1 unit are used.

(c) Observed probabilities are matched with the probability distribution curve given by Eq. (2) to determine the operational threshold magnitude b_{50} and its corresponding variance s^2. From these values we calculate b_{90}, $\hat{b}_{50}$, and $\hat{b}_{90}$.

(d) We put $T=0.5$ s in cases when the earthquake was not detected, whereas for detected events the period is measured from the seismograms and, in general, deviates from its mean value of 0.5 s. If we could assume $T=0.5\,\text{s}=\text{const.}$, then the threshold estimates would be parallel to $F(\Delta, T=0.5\,\text{s})$. Since this is not the case, we repeat the calculations for epicentral distances from 100 to 1000 km with steps of 100 km. Good correlation for the threshold magnitudes, m, is found empirically for the form

$$m=a_0+a_1 \log_e \Delta+a_2 \Delta. \tag{11}$$

Table 1. Estimates of threshold magnitudes for UPP

Δ	b_{50}	b_{90}	$\hat{b}_{50}$	$\hat{b}_{90}$
100	1.1	1.7	1.0	1.3
200	1.5	2.1	1.4	1.7
300	1.7	2.4	1.7	1.9
400	1.9	2.6	1.9	2.2
500	2.0	2.7	2.0	2.3
600	2.2	2.9	2.2	2.5
700	2.5	3.2	2.4	2.8
800	2.6	3.3	2.6	2.9
900	2.8	3.5	2.8	3.1
1,000	3.0	3.6	3.0	3.2

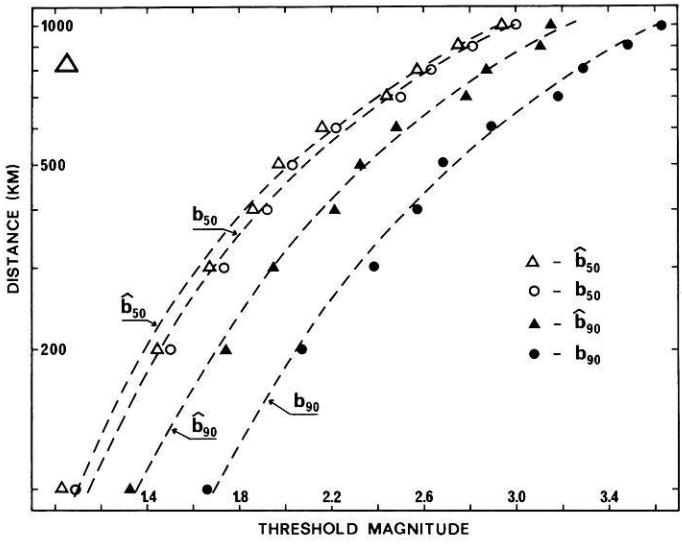


Fig. 3. Comparison of optimum fits (*dashed lines*) due to Eq. (11) and results obtained by the probit technique (*circles/triangles*), 50 % and 90 % levels as well as operational, b, and corrected operational, $\hat{b}$, threshold magnitudes are considered

As an example, Fig. 2 presents detectability curves for three chosen distances. Table 1 summarizes the estimates of the four threshold magnitudes for UPP. Coefficients a_0, a_1, a_2 are estimated in terms of the least-squares fits. Results are depicted in Fig. 3 for b_{50}, b_{90}, $\hat{b}_{50}$, and $\hat{b}_{90}$.

4. Examination of Results From a Sequence of Rockbursts

The mining area of Grängesberg (60.1°N, 15.0°E) in central Sweden, located approximately 150 km from UPP, is one of few seismic areas which may be used as an independent test of the estimated threshold magnitudes. During the period August 1974–October 1977 about 470 rockbursts, from the Grängesberg area, were recorded by at least one station from the SSSN (Båth and Wahlström, 1976 and Båth, 1977). Employing the available information on the rockburst series, we carried out the analysis as described by Ringdal (1975) or in other words, we applied the direct estimation method for UPP. Resulting values for UPP and $\Delta = 150$ km are: $b_{50} = 1.4$, $b_{90} = 1.8$, $\hat{b}_{50} = 1.4$, and $\hat{b}_{90} = 1.6$. These results compare favourably with estimates obtained from Eq. (11), differing by only 0.1 magnitude unit for all four threshold magnitudes (Shapira et al., 1978).

5. Discussion

The modification of the direct method enables one to investigate the detectability of weak events by means of seismograph stations located within the seismic region

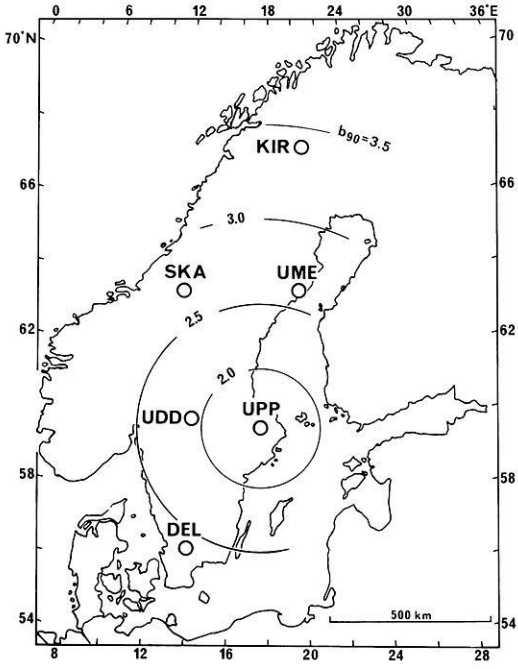


Fig. 4. Loci of constant operational threshold magnitudes, b_{90}, with respect to UPP

studied. As an example, loci of constant threshold magnitudes, b_{90}, *wrt* UPP are shown in Fig. 4. Certain practical considerations and limitations of the method proposed are discussed below.

Source Parameters. Generally speaking, Swedish earthquakes differ from each other by virtue of their different unknown source mechanisms, focal depths, etc. The relocation of events, which is the basic principle of the present approach, does not consider the possible influence of the source parameters. Varying source characteristics increase the standard deviations of estimated parameters. In spite of the undesired source influences, the described approach provides realistic estimates of threshold magnitudes as confirmed by the rockburst sequence.

True Magnitudes. Magnitudes considered in the present work are determined by means of a regional network consisting of six stations. In such a case, it becomes rather difficult to define the true magnitude of an event. Regarding Swedish recording conditions, the best estimate of the true (unknown) magnitude is provided by combined measurements from all the six stations.

The estimates of threshold magnitudes are obviously dependent upon the utilized magnitude scale. Therefore, direct comparison of threshold magnitude estimates resulting from different magnitude scales is not possible.

Frequency-Magnitude Relationship. An assumption of the random distribution of earthquakes with time has been made by present authors. Nevertheless, it is not impossible that the occurrence of Swedish earthquakes shows certain temporal variations. Caution has to be taken also when considering the frequency of the

Grängesberg rockbursts due to the possible time-depended triggering (Båth and Wahlström, 1976; Båth, 1977). However, the temporal effects are minimized by extending the period of observations. Thus, the potential, e.g., seasonal variations are considered to be smoothed out.

The parameter β in Eq. (1) is influenced by the relocation process. Nevertheless, Shapira et al. (1978) show that, for the present observational material, variations of β due to the event relocation are negligible.

Performance of UPP. The detection capability ascribed above to UPP is not necessarily the best one which may be achieved practically. Note, that the thresholds have been associated with short-period vertical-component Sg readings. It is true that, in general, Sg is the largest phase observed in seismograms of Swedish events recorded by the SSSN. However, horizontal amplitudes of Sg usually exceed the vertical amplitudes (Båth et al., 1976). Hence, inclusion of horizontal-component observations would very likely decrease, i.e., improve, the present threshold magnitudes.

6. Conclusions

Our main conclusions may be formulated as follows: (a) The applicability of the well-known direct estimation method has been extended to weak events recorded merely by regional seismic networks. (b) Threshold magnitudes are given numerically for the Uppsala station. (c) Reliability of the suggested approach has been confirmed by means of a rockburst series.

Acknowledgements. This research has been carried out at the Seismological Institute, Uppsala University, Sweden, with financial support of the Swedish Institute for Cultural Relations with Foreign Countries (for A.S.). Our sincere thanks are due to Professor J.R. Brown and Dr. E. Elvers for helpful discussions and critical reading of the manuscript. All computations have been made on the IBM 370/155 computer at the Uppsala University Data Center.

References

Båth, M.: A rockburst sequence at the Grängesberg iron ore mines in central Sweden. Part II. Seismological Institute, Uppsala, Rep. No. **5–77**, 32, 1977

Båth, M., Kulhánek, O., Eck, T. van, Wahlström, R.: Engineering analysis of ground motion in Sweden. Fourier and response spectra, attenuation, regional magnitude, energy, acceleration. Seismological Institute, Uppsala, Rep. No. **5–76**, 48, 1976

Båth, M., Wahlström, R.: A rockburst sequence at the Grängesberg iron ore mines in central Sweden. Seismological Institute, Uppsala, Rep. No. **6–76**, 48, 1976

IBM Application Program: Subroutine PROBT, System/360 Scientific Subroutine Package, Version III. 454 pp. New York: IBM Corporation, 1970

Kelly, E.J., Lacoss, R.T.: Statistical estimation of seismicity and detection probability. Mass. Inst. Tech., Lincoln Laboratory. Semiannual technical summary, Seismic discrimination, 30 June 1969, p. 17., 1969

Kulhánek, O., Wahlström, R.: Earthquakes of Sweden 1891–1957, 1963–1972. Technical report. Seismological Institute, Uppsala, 21, 1977

Pirhonen, S.E., Ringdal, F., Berteussen, K.A.: Event detectability of seismograph stations in Fennoscandia. Phys. Earth Planet. Inter. **12**, 329–342, 1976

Ringdal, F.: On the estimation of seismic detection thresholds. Bull. Seismol. Soc. Am. **65**, 1631–1642, 1975

Ringdal, F., Husebye, E.S., Fyen, J.: Earthquake detectability estimates for 478 globally distributed seismograph stations. Phys. Earth Planet. Inter. **15**, P24–P32, 1977

Seggern, D. von, Blandford, R.: Seismic threshold determination. Bull. Seismol. Soc. Am. **66**, 753–788, 1976

Shapira, A., Kulhánek, O., Wahlström, R.: Threshold magnitudes for weak regional seismic events. Seismological Institute, Uppsala, Rep. No. **2–78**, 24, 1978

Received December 19, 1978; Revised Version June 5, 1979; Accepted August 13, 1979

J. Geophys. 46, 135–149, 1979

Journal of Geophysics

Accuracy of Ray Theoretical Seismograms*

V. Červený

Institute of Geophysics, Charles University, Ke Karlovu 3, 121 16 Praha 2, ČSSR

Abstract. Theoretical seismograms computed by the ray method with certain modifications in the critical region are compared with theoretical seismograms computed by more exact methods (the reflectivity method). It is shown that the ray method with some simple modifications gives qualitatively satisfactory results for a broad class of models of media, even in certain singular regions (such as critical regions, the neighbourhood of caustics, the half-shadow, etc). Poor results are obtained for some non-ray waves. It would be possible, however, to use some modifications of the ray method (such as the ray method with a complex eikonal) even for certain non-ray waves and to include these waves into ray theoretical seismograms.

Key words: Ray theoretical seismograms – Accuracy of ray method – Modifications of ray method.

I. Introduction

At the present time, many different methods can be used to construct theoretical seismograms. Unfortunately, there is no method generally suitable for all situations. For various problems, various epicentral distances, and various models of media, certain methods are suitable and other cannot be used at all. Great differences exist also in the accuracy and in the computer time requirements.

The ray method may be applied to construct theoretical seismograms even for rather complicated laterally inhomogeneous media with curved interfaces, for which other methods can be hardly used.

The basic problem in the application of ray methods to the construction of ray theoretical seismograms lies in its insufficient accuracy in certain situations. There are three main reasons for this limited accuracy:

1. The ray method is not applicable in certain singular regions, such as

* Presented at the Workshop Meeting on Seismic Waves in Laterally Inhomogeneous Media, Liblice, ČSSR, February 27 – March 3, 1978

0340-062X/79/0046/0135/$03.00

the critical region, the neighbourhood of caustics, the transition zone between the illuminated region and the shadow zone, etc.

2. The standard ray method cannot be used to describe the properties of certain types of non-ray waves, such as various inhomogeneous waves, channel waves, tunnel waves, etc. (Of course, certain of these waves can be studied by the ray method with a complex eikonal.)

3. The number of waves which arrive at the receiver within a finite time window is often rather high. It is necessary to exclude a priori certain classes of waves (even regular zero-order ray waves) to make computations possible. Of course, this may lead to inaccuracies.

The object of this paper is to study the accuracy of ray theoretical seismograms. It would be difficult (practically impossible) to appreciate the accuracy of ray theoretical seismograms by the ray method itself. The simpler way is to compare the ray theoretical seismograms with the seismograms computed by another, more accurate method. We must, of course, perform computations only for simpler types of media, to which the more accurate methods are applicable. From this point of view, the most natural thing to do is to start with vertically inhomogeneous media, for which a number of various more exact methods can be now used to compute thoretical seismograms.

In this paper we shall compare the ray theoretical seismograms with theoretical seismograms computed by the reflectivity method, see Fuchs (1968), Fuchs and Müller (1971). We shall be mainly interested in the accuracy of ray theoretical seismograms in singular regions, see point (1) above. The accuracy of ray theoretical seismograms depends greatly on the epicentral distance. As we are interested here mainly in the interpretation of explosion seismology data, we shall study the range of epicentral distances from about 40 to 300 km for standard Earth's crust structures. The most important waves in this region are the refracted waves (also called diving waves) and the supercritical reflections from various first-order discontinuities. The frequently occurring singular regions at these epicentral distances are those listed sub (1) above. At shorter epicentral distances (~ 40 km) we have mainly subcritical reflections and the accuracy of ray theoretical seismograms is usually satisfactory even for more complicated structures. For large epicentral distances (~ 500 km), we expect lower accuracy in the ray description of waves propagating within the Earth's crust due to various reasons (such as the strong interference character of the wave field). The accuracy of ray theoretical seismograms in this region would require special investigation.

The problem of the accuracy of ray theoretical seismograms is far from being solved by the examples presented in this paper. The conclusions presented here are of course only of a very limited character. It would be necessary to perform a lot of other computations to be able to make some conclusions of more general character.

2. Construction of Ray Theoretical Seismograms

In the following we shall describe very shortly the main principles of the construction of ray theoretical seismograms. Details can be found in Červený et al. (1977), Hron and Kanasewich (1971), etc.

In the ray theory the wave field is decomposed into elementary waves, corresponding to individual rays. The seismograms of elementary waves, called elementary seismograms, are computed independently one after another. The resulting theoretical seismograms obtained as a superposition of elementary seismograms are called ray theoretical seismograms.

In the case of the medium being composed of homogeneous plane parallel layers, the elementary seismograms can be computed exactly (Cagniard-deHoop, Smirnov-Sobolev). It is, however, also possible to apply the standard ray theory, or some of its modifications.

For the computation of elementary seismograms, it is necessary to know the travel-time of the corresponding wave, its complex amplitude and the source-time function. The numerical methods of computation of these values in the case of vertically inhomogeneous media are well known. To compute elementary seismograms, it is also necessary to determine the Hilbert transform of the source-time function. For some important classes of functions, simple approximate formulae for the Hilbert transform were derived in Červený (1976). These functions can simulate with a good accuracy many real wavelets observed in explosion seismology studies. The approximate formulae expedite the computation of theoretical seismograms considerably.

To describe the type of the elementary wave, it is necessary to introduce certain numerical codes specifying the elementary waves. The main difficulty of the computation of ray theoretical seismograms does not lie in the computation of elementary seismograms, but in the algorithms for the generation of numerical codes and in the selection of waves. In case of a medium consisting of a finite number of plane-parallel homogeneous layers, the problem of the automatic generation of numerical codes becomes relatively simple, mainly when we do not consider converted phases. In this case, the elementary waves can be grouped into families of kinematically analogous waves. This grouping makes the computation substantially faster. Unfortunately, even the number of groups which arrive at the receiver within a finite time window is often rather high. This applies mainly to models with a large number of thin layers to large epicentral distances and to the long time window. In these cases, it is necessary to make certain a priori assumptions and to exclude from computations the waves that are expected to influence the theoretical seismograms only slightly.

To improve the accuracy of elementary ray theoretical seismograms, we can use certain modifications of the ray theory. In the case of a medium of parallel homogeneous layers, this applies mainly to the critical region modification. The critical region is of great importance in applications since the amplitude-distance curve of the corresponding reflected wave reaches its maximum there. It is not complicated to use critical region modifications when ray theoretical seismograms are computed. The computation is then substantially faster than the computation based on exact methods, and more exact than the computations based on the standard ray method. This modification was first used in the construction of ray theoretical seismograms in 1974, see Červený (1978). See also details in Červený et al. (1978).

It should also be noted that the ray theoretical seismograms can be easily supplemented by certain non-zero waves such as the head waves. In the ray theoretical seismograms presented in this paper the head waves are included automatically.

3. Discussion of the Accuracy of Ray Theoretical Seismograms

In this section we shall give some examples of theoretical seismograms computed by the ray method (or by some of its modifcations) and by the reflectivity method. To appreciate the accuracy of ray theoretical seismograms we shall choose certain simple vertically inhomogeneous structures. In the reflectivity method the smooth distribution of velocity and density with depth is simulated by a thin-layered medium with horizontal plane interfaces. Therefore, we shall consider only such types of media.

In all the computations (with the exception of the upper system in Fig. 11), the source-time function is given by the formula

$$f(t)=\exp\,(-4\pi^2 f_M^2 t^2/\gamma^2)\cos\,(2\pi f_M t+v), \tag{1}$$

with three free parameters: f_M, γ and v. The quantity f_M corresponds approximately to the prevailing frequency of the source-time function. In all the examples we consider an explosive point source of P-waves with the symmetrical characteristics situated near the Earth's surface. Only ideal registration is considered, and a possible distortion of theoretical seismograms by the recording equipment is not taken into account. The receiver is also situated near to the Earth's surface, and the vertical component of the displacement is presented. Only P-waves are taken into account(no S-waves and converted waves). For simplicity, only primary reflections (with corresponding head waves) are considered, no multiple reflections. Some scaling of amplitudes with respect to the epicentral distance is applied in most cases (proportional to the epicentral distance). The scaling is, of course, the same for the ray theoretical seismograms and for theoretical seismograms computed by the reflectivity method.

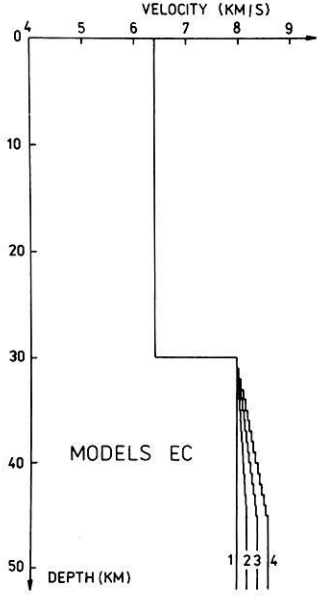


Fig. 1. P-wave velocity distribution for vertically inhomogeneous models EC1–EC4

3.1. Critical Region

First we shall consider simple models of a homogeneous Earth's crust with the Mohorovičić discontinuity situated at the depth of 30 km, see Fig. 1. In the uppermost mantle the velocity is either constant (see model EC1), or increases linearly with depth (models EC2, EC3, EC4). In this section we shall deal with model EC1 with the homogeneous upper mantle.

Figures 2 and 3 give theoretical seismograms computed by three different methods. The source time function in Figs. 2 and 3 is given by formula (1), with $\gamma=4$, $\nu=0$, it differs only in the predominant frequency, viz., $f_M=4$ Hz in Fig. 2 and $f_M=2$ Hz in Fig. 3. In both these cases, the uppermost figure (A) is computed by the reflectivity method, the middle figure (B) by the ray method with the critical region modification, and the lower figure (C) by the standard ray method.

In the standard ray method, the amplitude-distance curve of the reflected wave reaches its maximum right at the critical point. In our case, the critical

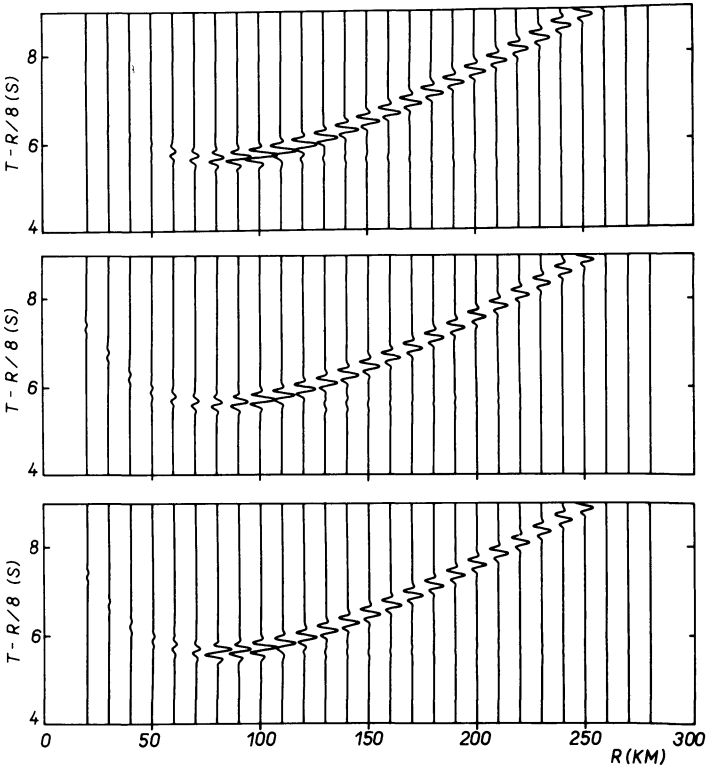


Fig. 2. A comparison of theoretical seismograms for the model EC1 shown in Fig. 1, computed by three different methods: *Upper* figure (A): the reflectivity method. *Middle* figure (B): the ray method, with a critical region modification. *Lower* figure (C): the standard ray method. The explosive point source and receiver are located close to the Earth's surface. The source-time function is given by (1), with $\gamma=4$, $\nu=0$ and with the predominant frequency $f_M=4$ Hz

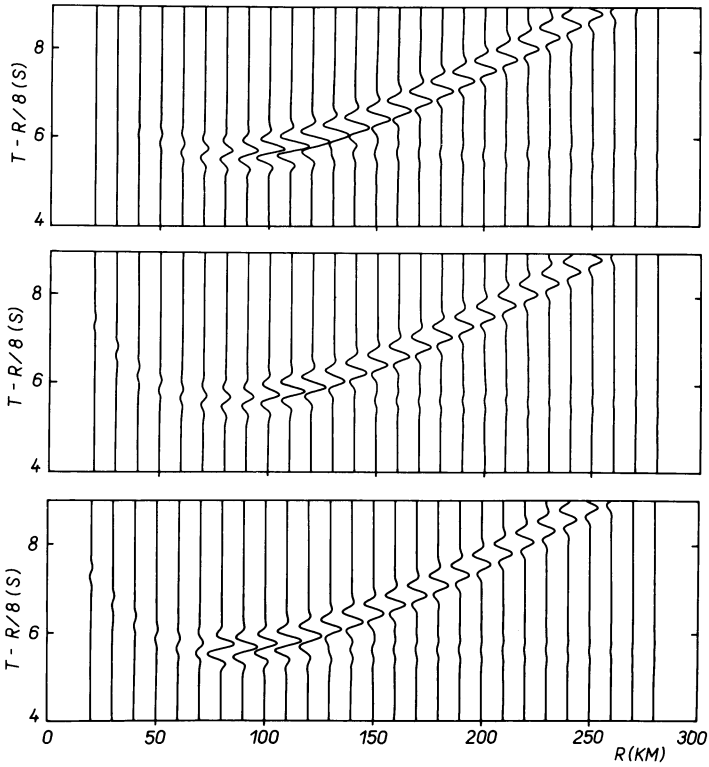


Fig. 3. The same as in Fig. 2, for the predominant frequency $f_M = 2$ Hz

point is situated at the epicentral distance of 80 km. The maximum amplitudes at the critical point can be clearly seen in both sets of theoretical seismograms computed by the standard ray method. In reality, the maximum of amplitude curves of reflected waves is shifted beyond the critical point, and the shift is frequency dependent. Both these effects can be clearly seen in theoretical seismograms computed by the reflectivity method and in theoretical seismograms computed by the ray method with the modification in the critical region (see B and C). In the case $f_M = 2$ Hz, the maximum amplitudes are shifted approximately to 110 km, and in the case of $f_M = 4$ Hz approximately to 100 km.

Although the results obtained by the two methods (see A and B) differ in details, the overall agreement of most important pecularities of wave fields is satisfactory. This agreement was obtained due to the applied modification of the ray method in the critical region.

(Let us mention one difference between the two systems of seismograms which is connected only with the numerical computation effects. Weak amplitudes of the wave reflected from the Moho at small epicentral distances obtained by the reflectivity method are caused by the velocity filtration applied in the reflectivity method.)

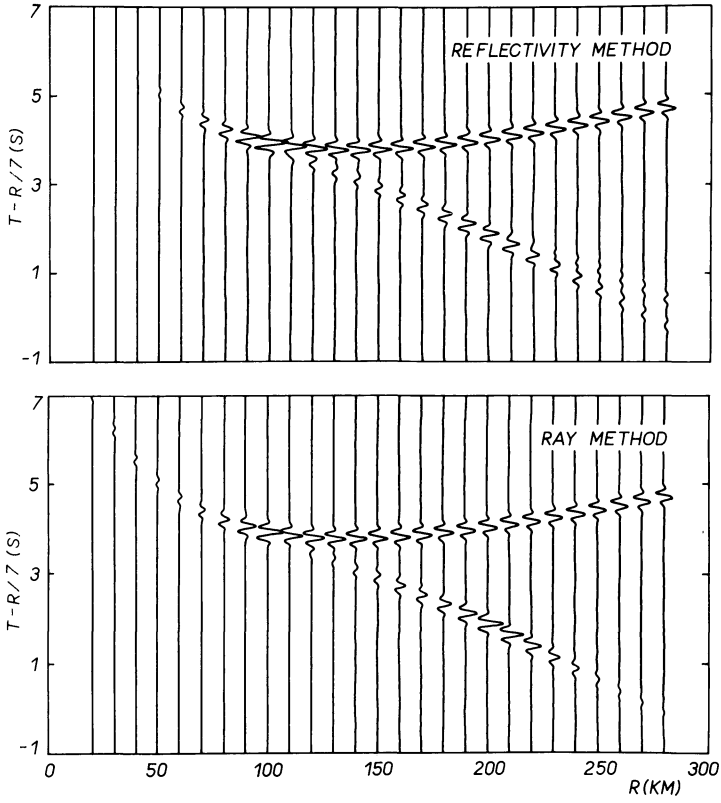


Fig. 4. A comparison of theoretical seismograms for the model EC4 shown in Fig. 1, computed by two different methods: *Upper* figure: reflectivity method. *Lower* figure: the ray method with a critical region modification for reflections from the Moho. The explosive source and the receiver are located close to the Earth's surface. The source-time function is given by (1) with $\gamma=4$ and $\nu=0$ and with the predominant frequency $f_M=4$ Hz

3.2. Interference Head Waves

Now we shall present theoretical seismograms computed for models EC2, EC3 and EC4, see Fig. 1. The continuous increase of velocities in the uppermost mantle is simulated by the sequence of thin homogeneous layers. The corresponding systems of theoretical seismograms are presented in Figs. 4–6 for the prevailing frequency $f_M=4$ Hz. The critical region modification is used for reflected waves from the Moho (in the computation of ray theoretical seismograms).

Two dominant waves are evident in these figures: the wave reflected from the Moho and the wave refracted in the uppermost mantle. This wave is also called the interference head wave, see Červený and Ravindra (1971). Let us note that the refracted wave is formed by a superposition of waves reflected from individual fictitious interfaces in the uppermost mantle.

The properties of the wave reflected from the Moho remain the same as in Figs. 2 and 3, see the previous ection. Here we shall be interested in the

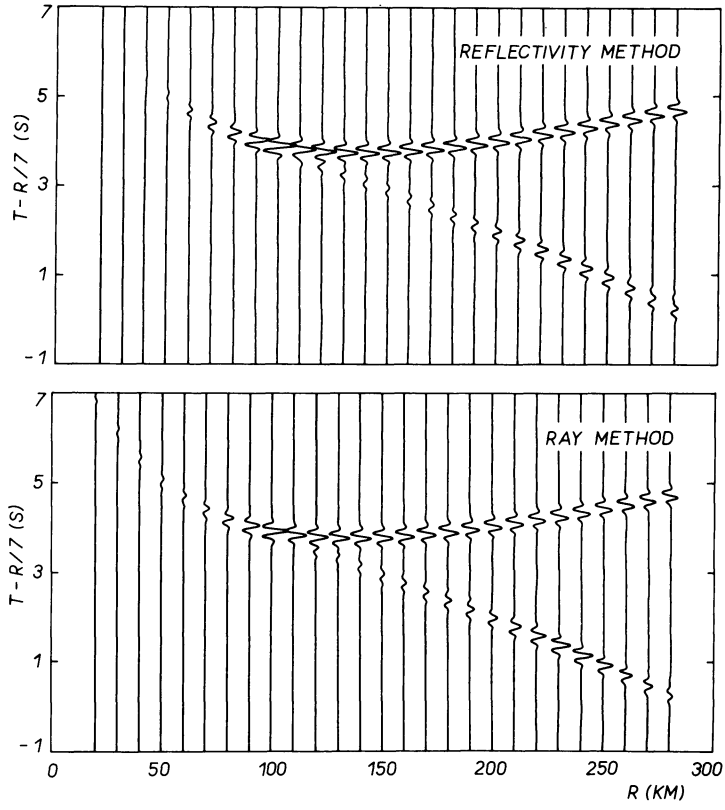


Fig. 5. The same as in Fig. 4, for the model EC3

properties of the refracted wave in the uppermost mantle. For this wave, the standard ray method is not applicable, see Červený and Ravindra (1971). Certain interesting pecularities of these waves, described earlier in the above shown reference, may be easily verified in Figs. 4–6: the amplitudes of refracted waves beyond the critical point first decrease with the increasing epicentral distance, then they increase, and at a certain epicentral distance the amplitudes reach their maximum values. The position of this maximum depends on the velocity gradient below the Moho. The larger the gradient, the smaller is the epicentral distance at which the maximum is situated.

Again, we can see that the theoretical seismograms computed by the two methods differ in details, but the most important pecularities (such as the position of maxima and minima of amplitude-distance curves) are satisfactory for many practical purposes. As a rule, the ray method gives larger amplitudes than the reflectivity method.

(Let us mention another difference which is connected with certain differences in computational procedures. The wave, which separates from the refracted wave at large epicentral distances in the reflectivity computations for the model EC4 corresponds to the multiple reflection from the bottom side of the Mohorov-

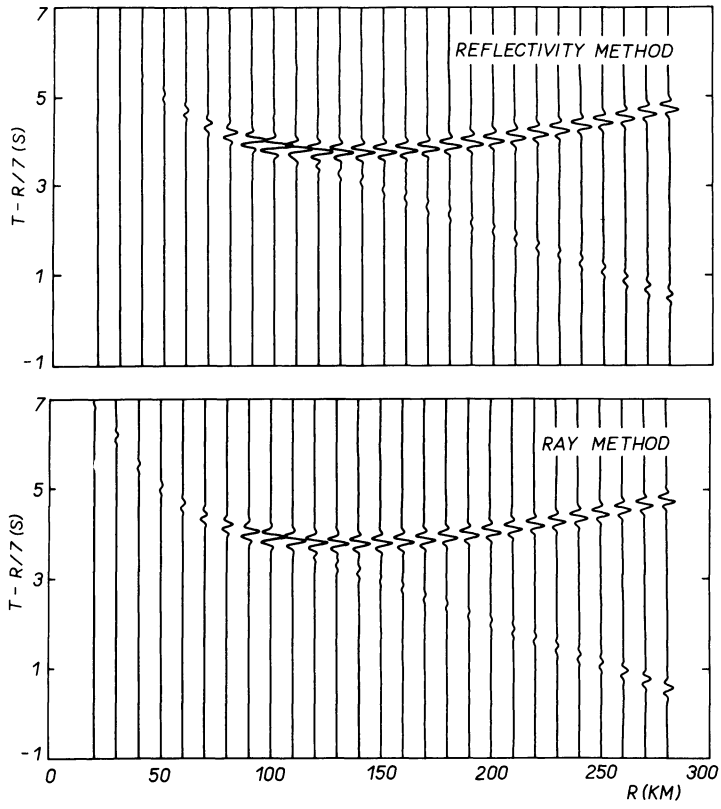


Fig. 6. The same as in Fig. 4, for the model EC2

ičić discontinuity. In ray theoretical seismograms, this wave is missing since only primary reflections were considered. It would be easy, though, to obtain this wave even in ray theoretical seismograms by choosing different input data controlling the generation algorithms. This possibility is included in the programs used for computations.)

3.3. Transition Region Between the Shadow and the Illuminated Zone

We shall now consider the models ECG shown in Fig. 7. The models are similar to those discussed above (see Fig. 1), but assume a positive velocity gradient within the Earth's crust. For the models ECG shown In Fig. 7, we obtain a strong refracted wave propagating within the Earth's crust, which did not exist in the case of models EC. The travel time curve of this refracted wave has a common tangent point with the travel-time curve of the wave reflected from the Moho. This tangent point is situated at about 195 km. At larger epicentral distances, a geometrical shadow zone for both waves is formed. We are not interested here in the deep shadow, as the amplitudes are very

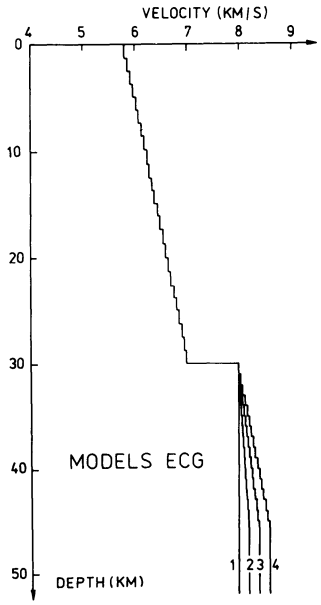


Fig. 7. *P*-wave velocity distribution for vertically inhomogeneous models ECG1–ECG4

small there, but in the wave field in the half-shadow (~ 195 km). It is well known that the standard ray theory fails there.

Now we shall present theoretical seismograms for the model ECG1, see Fig. 8 for $f_M = 4$ Hz and Fig. 9 for $f_M = 2$ Hz. Again, the upper figure is computed by the reflectivity method, the lower by the ray method, with the modification in the critical region for the wave reflected from the Moho.

We are interested mainly in the properties of the interference wave formed by the superposition of the refracted wave and the wave reflected from the Moho, in the vicinity of the gometrical boundary of the shadow zone (~ 195 km). The behaviour of this interference wave is very interesting. The amplitudes of the interference wave first strongly increase, form a maximum at about 170 km, and they continuously decrease. The level of amplitudes of the interference wave in the region of their maximum (~ 170 km) is really rather high, of the same order or even higher than the amplitudes in the critical region.

Similarly as in previous cases, the agreement between the theoretical seismograms computed by the two methods is satisfactory. The agreement is in a way surprizing, no modification of the ray method was used to compute the interference wave in the transition region. The agreement is mostly caused by the simulation of the smooth velocity-depth distribution by thin layers. The agreement could not be obtained for a smooth velocity-depth distribution.

Similar results were obtained for the models ECG2–ECG4. We shall not present them here for they would be, to some extent, only a combination of theoretical seismograms for the model ECG1 with those for models EC2, EC3, and EC4. As an example, we shall present here only the ray theoretical

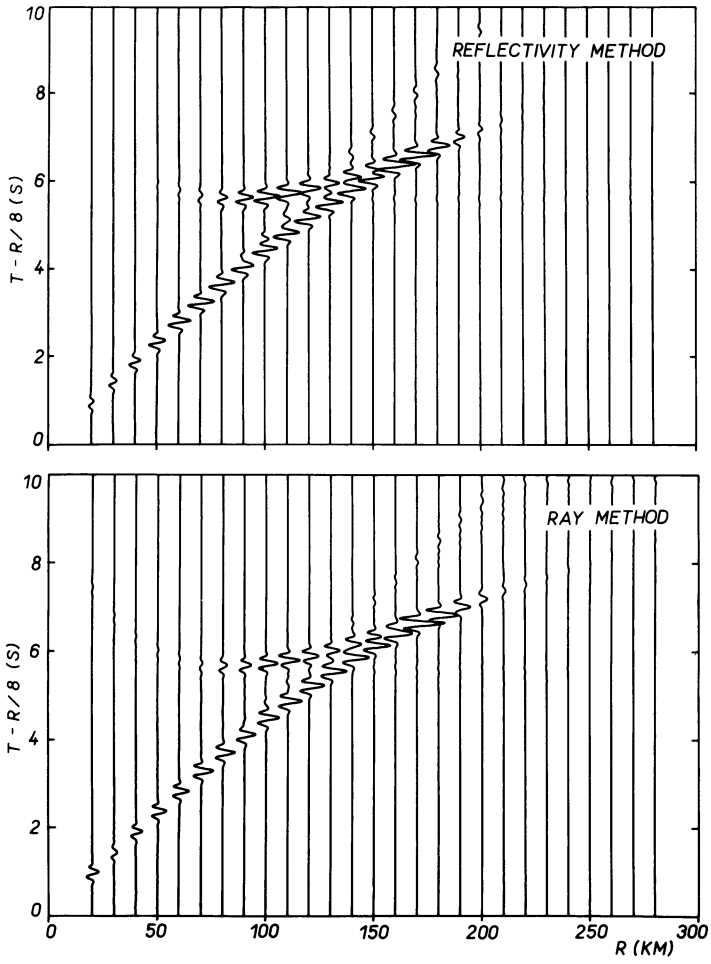


Fig. 8. A comparison of theoretical seismograms for the model ECG1 shown in Fig. 7, computed by two different methods: *Upper* figure: the reflectivity method. *Lower* figure: the ray method with a critical region modification for reflections from the Moho. The explosive point source and receiver are situated close to the Earth's surface. The source-time function is given by (1), with $\gamma=4$ and with $v=0$. The predominant frequency f_M equals 4 Hz

seismograms for the model ECG4 (with the critical region modification for reflections from the Moho), computed for the predominant frequencies $f_M=4$ Hz and $f_M=2$ Hz, see Fig. 10.

3.4. Other Singular Regions

A number of other ray theoretical seismograms were computed and compared with the reflectivity method. This applies to various layered structures composed

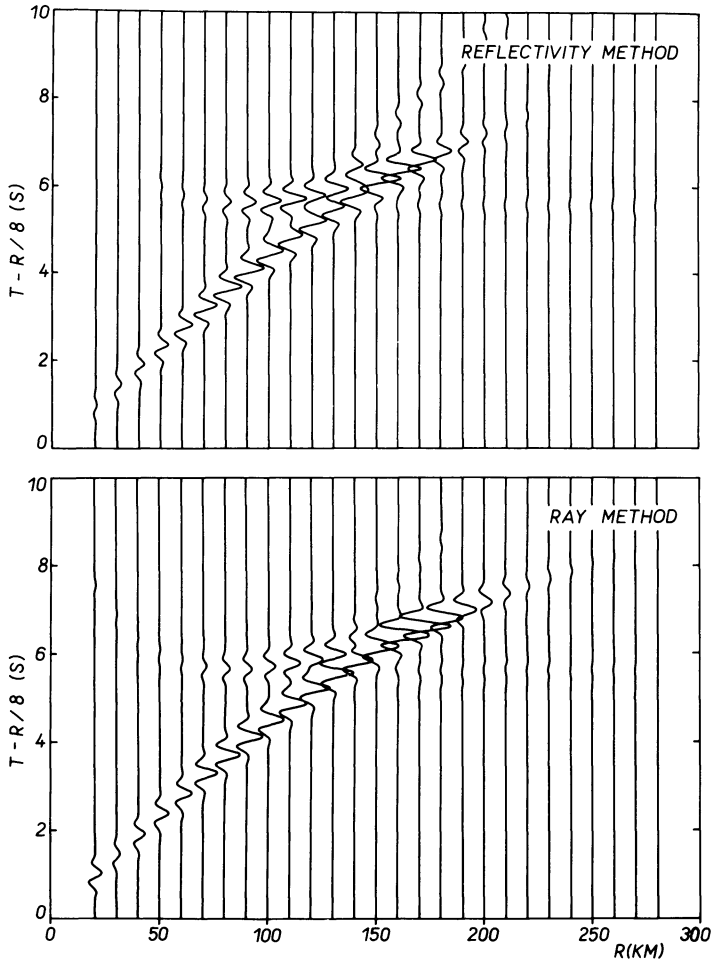


Fig. 9. The same as in Fig. 8, for the predominant frequency $f_M = 2$ Hz

of thick layers, to more realistic structures, etc. The agreement of the results was usually satisfactory. Great attention was also paid to the neighbourhood of a caustic. Some results of these computations can be found in Červený et al. (1977), Červený (1978), Červený and Pšenčík (1977). Similarly as in the case of the half-shadow the aggreement is satisfactory.

3.5. Non-Ray Waves

As shown above, the ray theory supplemented by some simple modifications may yield satisfactory results, even in singular regions.

As was mentioned in the introduction, the standard ray method does not include certain types of non-ray waves, such as various inhomogeneous waves,

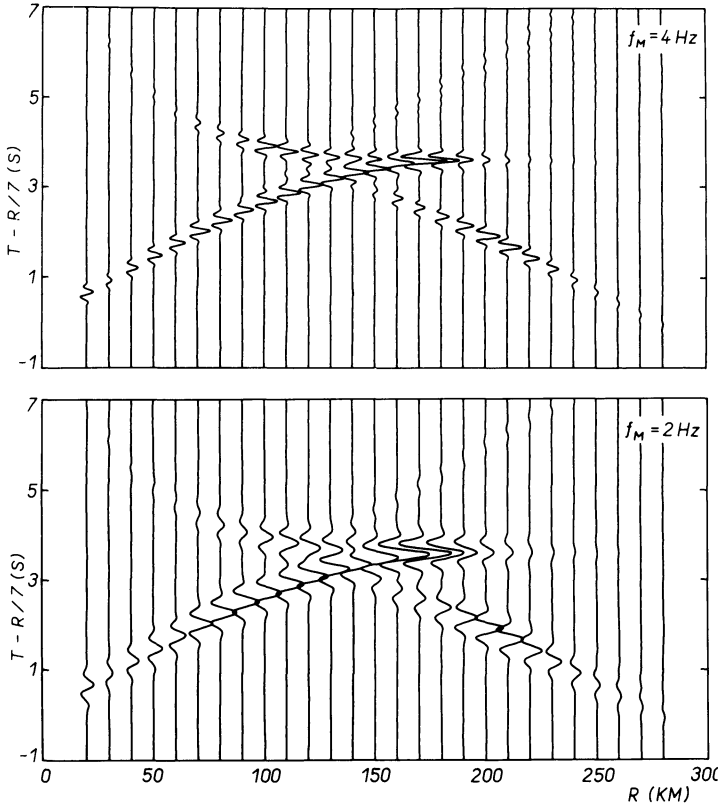


Fig. 10. Ray theoretical seismograms for the model ECG4. The critical region modification is applied to reflections from the Moho. The explosive point source and the receiver are situated close to the Earth's surface. The source-time function is given by (1), with $\gamma = 4$, $\nu = 0$. The predominant frequency $f_M = 4$ Hz in the upper figure, $f_M = 2$ Hz in the *lower* figure

channel waves, tunnel waves, diffracted waves. Some types of diffracted waves can be obtained by a proper simulation of the smooth velocity distribution by a thin-layered medium, see the waves penetrating to the shadow zone in Figs. 8 and 9. For other types of waves, such as the tunnel waves, the comparison of ray theoretical seismograms with the theoretical seismograms computed by the ray method has not given satisfactory results.

We shall present one example, see Fig. 11. Fuchs and Schulz (1976) studied the effect of the thin-high-velocity layer on the wave field in an attempt to explain the combination of low-frequency Moho reflections and high-frequency P_n arrivals which were observed in several cases. The theoretical seismograms for a structure given in the left upper corner of Fig. 11, computed by the reflectivity method, are shown in the upper part of the figure. They clearly demonstrate the tunnelling of waves through a thin high-velocity lamina. The first arrival is the reflection from the high-velocity layer, the second arrival comes from the Moho. As soon as the incidence of the wave from the source

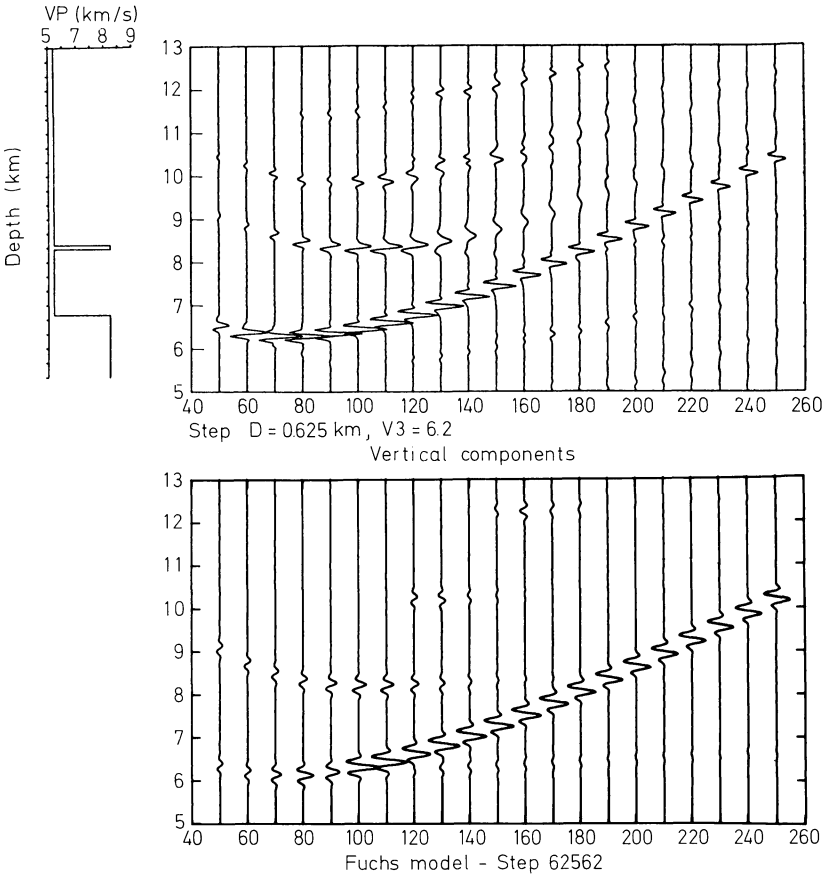


Fig. 11. A comparison of theoretical seismograms for the model of medium containing a thin high-velocity lamina, see upper left corner. *Upper* figure: theoretical seismograms computed by the reflectivity method. *Lower* figure: ray theoretical seismograms, with the critical region modification for reflections from lamina and from the Moho. The explosive point source and the receiver are located close to the Earth's surface, the source-time function in both sections are slightly different

on the lamina is supercritical, only the low frequencies propagate (tunnel) through the lamina, and the reflection from the next interface contains only low frequencies. It must be emphasized that the tunnel wave cannot be described by the standard ray theory. The ray theoretical seismograms computed for the same model are shown in the lower part of Fig. 11. The figure clearly demonstrates that the standard ray method does not properly describe the tunnel effects.

It would be possible, however, to use some modifications of the ray method (such as the ray method with a complex eikonal, see Červený et al., 1977) even for certain on-ray waves. These modifications have not yet been included in our programs for ray theoretical seismograms, but they have been checked in some test computations (e.g., for tunnel waves).

4. Conclusions

The comparisons of the ray theoretical seismograms with the reflectivity theoretical seismograms have shown that the ray method with some simple modifications gives qualitatively satisfactory results for a broad class of models of media, even in some singular regions. This applies, e.g. to the critical region, to the neighbourhood of a caustic, the half-shadow, etc. Satisfactory results have been also obtained for interference head waves, converted waves, etc. On the other hand, poor results have been obtained for some non-ray waves and inhomogeneous waves. We may expect that certain modications (the ray method with a complex eikonal, etc.) will be used in the near future to remove many of the remaining limitations and improve effectively the accuracy of ray theoretical seismograms.

These comparisons show that the ray theoretical seismograms can give valuable results even for laterally inhomogenous media with curved interfaces, especially when we use certain modifications to improve their accuracy in singular regions and to describe properly certain non-ray waves.

Acknowledgements. The author wishes to thank to Profesor K. Fuchs for the invitation to work for 3 months at the Institute of Geophysics, University of Karlsruhe, and to Deutscher Akademischer Ausstauschdienst for the financial support of this stay. All the computations presented in this paper were performed during this stay. The author is also greatly indebted to his colleagues from the Charles University (Prague) and from the University of Karlsruhe for helpful suggestions and valuable discussions.

References

Červený, V.: Approximate expressions for the Hilbert transform of a certain class of functions and their application to the theory of seismic waves. Stud. Geophys. Geod. **20**, 125–132, 1976

Červený, V.: Ray theoretical seismograms. In: Geophysikalische Interpretationsmethoden. Sammelwerk des Symposiums der Arbeitsgruppe 1.7 KAPG, Bratislava, Mai 1974, L. Kubáčková, ed.: pp. 49–64. Bratislava: Veda 1978

Červený, V., Fuchs, K., Müller, G., Zahradník, J.: Theoretical seismograms for inhomogeneous elastic media. Voprosy dinamicheskoy teorii rasprostraneniya seismicheskikch voln. Vol. 20, G.I. Petrashen, ed. Leningrad: Nauka 1978

Červený, V., Molotkov, I.A., Pšenčík, I.: Ray method in seismology. Praha: Univerzita Karlova 1977

Červený, V., Pšenčik. I.: Ray theoretical seismograms for laterally varying layered structures. In: Proceedings XV. General Assembly European Seismological Commission, Kraków 1976, Part I. Publ. Inst. Géophys. Pol. Acad. Sci A-4 (115), R. Teisseyre, ed.: pp. 173–185. Warszawa-Lódź: PWN 1977

Červený, V., Ravindra, R.: The theory of seismic head waves. Toronto and Buffalo: University Toronto Press 1971

Fuchs, K.: The reflection of spherical waves from transition zones with arbitrary depth-dependent elastic moduli and density. J. Phys. Earth **16** (Special Issue), 27–41, 1968

Fuchs, K., Müller, G.: Computation of synthetic seismograms with the reflectivity method and comparison with observations. Geophys. J.R. Astron. Soc. **23**, 417–433, 1971

Fuchs, K., Schulz, K.: Tunneling of low-frequency waves through the subcrustal lithosphere. J. Geophys. **42**, 175–190, 1976

Hron, F., Kanasewich, E.R.: Synthetic seimograms for deep seismic sounding studies using asymptotic ray theory. Bull. Seismol. Soc. Am. **61**, 1169–1200, 1971

Received August 1, 1978; Accepted July 2, 1979

J. Geophys. 46, 151–169, 1979

Journal of Geophysics

Auroral Particle Fluxes in the Ionosphere

K. Wilhelm

Max-Planck-Institut für Aeronomie, D-3411 Katlenburg-Lindau 3,
Federal Republic of Germany

Abstract. The observations discussed in this paper were conducted in the framework of the 'Polar High Atmosphere Sounding Rocket Project' with a view to determining the contributions of auroral particle fluxes to both the energy budget and the current system of the upper atmosphere. The experiment was flown on board four payloads that were launched into various phases of magnetospheric substorm events to peak altitudes of approximately 270 km. During undisturbed portions of the flight times the electron flux below 500 eV was nearly isotropic and could be described by a power law spectrum. At high energies the spectra were steeper and exhibited a loss signature in the atmospheric backscatter cone. In disturbed periods electron fluxes with peaked spectra in the keV energy range were often superimposed on these distributions. In addition, strongly field-aligned electron fluxes were frequently observed at low energies. It seems difficult to devise any other mechanism to explain some of the observed events than field-aligned electric field acceleration. Results will also be presented on the energy flux carried by electrons and protons and the relationship to optical auroral emissions.

Key words: Magnetospheric substorm – Auroral particle fluxes – Magnetically aligned electric fields.

1. Introduction

The low-energy auroral particle experiment (AL1) was flown on four complex payloads in the framework of the 'Polar High Atmosphere Sounding Rocket Project'. The main goal of the experiment was to determine the bulk properties of low-energy particle fluxes during various phases of magnetospheric substorm events in order to provide input data for interpreting the atmospheric response to magnetospheric disturbances. This response was observed on the ground and by other instruments of the payload compliment comprising plasma experiments, mass spectrometer, magnetometer, and optical instruments. A description

0340-062X/79/0046/0151/$03.80

of the scientific objectives, the payload and the geophysical launch conditions has been given by Theile (1978).

Another scientific objective of the low-energy particle experiment was to study the particle distribution function as a means to deduce particle acceleration and precipitation mechanisms. Low-energy electron fluxes can be considered as sensitive probes of magnetospheric electric fields. In view of the importance of these fields, the data presented here will mainly be considered in the context of possible acceleration and precipitation mechanisms involving electric fields. Other aspects such as energy input into the high atmosphere, electric current distributions and wave-particle interaction processes will have to be treated in more detail, in a later paper together with supporting information from other experiments flown on the payloads or operated elsewhere. All observations discussed here were made over or near auroral forms in various stages of activity. To further the understanding of this exciting phenomenon of the polar night was one of the motivations for performing the measurements. In earlier investigations it has been recognized that the observed spectra can be divided into primary and secondary electron fluxes. Secondary electrons normally have energies of less than 500 eV and tend to be completely isotropic with a power law dependence of the flux and spectral parameters between 1 and 2 (Cahill et al., 1974; Arnoldy et al., 1974; Peterson et al., 1977). This behaviour was theoretically explained by considering the interaction of primary electron beams with the atmosphere by Banks et al. (1974) and by taking into account the reflection of upwards going low-energy electrons at potential barriers that supposedly are located above auroral displays (Evans, 1974).

In the energy range of several keV electron fluxes with peaked energy spectra were often observed above bright auroral displays. It was also recognized that these fluxes were normally field-aligned (O'Brien and Reasoner, 1971; Whalen and McDiarmid, 1972; Arnoldy et al., 1974; Cahill et al., 1974; Boyd, 1975). Mende and Shelley (1976) demonstrated that the spectral peaks could not be correlated with observations on a geosynchronous satellite thus supporting a mechanism of low-altitude post-acceleration of auroral particles. Boyd (1975) arrived at a threshold intensity for peaked spectra of about 12 erg $cm^{-2} s^{-1}$ corresponding to an IBC II aurora. Low-energy electron fluxes have also been identified as being important carriers for magnetospheric electric currents (Arnoldy et al., 1974; Cahill et al., 1974).

2. Instrumentation

The low-energy auroral particle experiment was designed to achieve good energy and angular resolution for electrons and protons. These requirements resulted in a nested-electrostatic analyser system, named differential energy analyser (DEA), followed by continuous channel electron multipliers (CEM) for separate electron and proton detection. A schematic sketch of the analyser is given in Fig. 1. The outer and inner spheres were grounded and the potential of the middle sphere was varied between 0 and -4 kV. Plasma rejection was achieved by a double grid in front of the entrance aperture at ground and

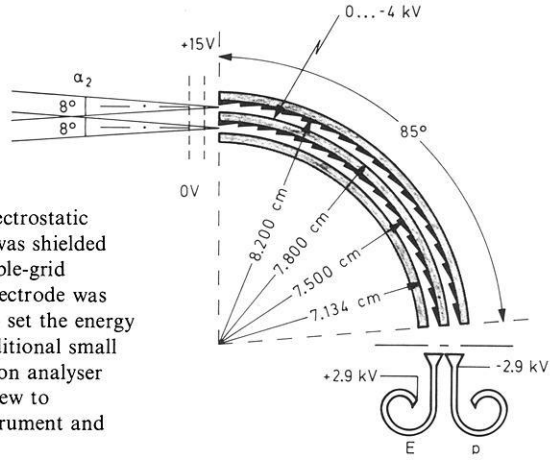


Fig. 1. Schematic sketch of the nested-electrostatic analyser (DEA). The entrance aperture was shielded from thermal plasma particles by a double-grid structure. The potential of the middle electrode was varied between 0 and -4 kV in order to set the energy acceptance band of the analyser. An additional small CEM (e) was mounted behind the electron analyser outside the plane of this figure with a view to increasing the dynamic range of the instrument and providing some redundancy

$+15$ V. As, in general, the rocket skin was at slightly negative potentials of a few Volts with respect to the plasma, both thermal electrons and protons were excluded from the inside of the analyser by this arrangement. The outer channel deflected protons and the inner one electrons into the detection devices. Before entering the CEM, the protons were post-accelerated by a voltage of -2.9 kV to enhance the instrument efficiency for low-energy protons. The concave surfaces were serrated in order to reduce scattering of particles inside the spherical analysers. In addition, the aluminium spheres were chromium-black plated, which, at the same time, guaranteed a good electrical conductivity of the surfaces of the spheres.

Two electron channels with different geometric factors and one proton channel were looking approximately 25 deg off the rocket axis. Another identical detector arrangement was mounted under an angle of 115 deg with respect to the payload axis. The sampling time was 15.8 ms followed by a dead time of 0.2 ms. The calibration procedure applied during the laboratory tests of the instruments has been described by Wilhelm (1979). It can be summarized in Fig. 2 by presenting an example of the relative detector efficiency ε_1 as a function of the particle energy, the azimuthal angle α_2 defined in Fig. 1 and the meridional angle α_1 normal to α_2. The two expressions 'meridional' and 'azimuthal' refer to the payload spin axis and imply a special mounting configuration of the instrument with respect to the payload. Given the nominal direction of the payload axis, i.e., parallel to the local vertical, and the magnetic field vector, it can be seen that the pitch angle resolution is predominantly determined by the acceptance range of the 'meridional' angle. For this reason the angle α_1 was mounted parallel to the meridional plane of the rocket. The evaluation of the calibration led to the instrument characteristics given in Table 1.

Taking into account both the CEM properties and the efficiency as a function of energy, one can perform an absolute conversion of count rates into particle fluxes.

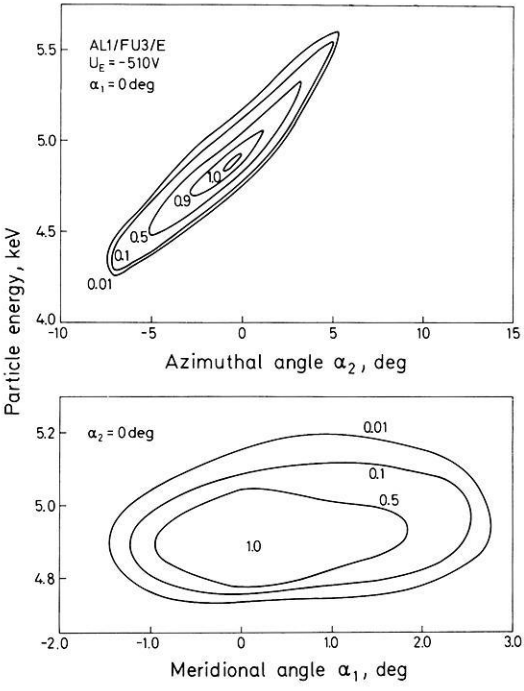


Fig. 2. Contours of the relative efficiency ε_1 of one of the large-geometric-factor electron channels. The efficiency is plotted as a function of the two angular directions and the particle energy. Note the different scales

Table 1. Characteristics of instrument

Energy range,	eV	15–35,000
Resolution $\Delta W/W$,	%	14
Geometric factor,	cm^2 sr	2.3×10^{-3}
Field of view,	deg^2	2.7×8.0
Time resolution,	ms	16
Payload spin period,	ms	330

3. Observations

Four payloads were launched from the Andoya Rocket Range, Andenes, Norway by Skylark 7 motors. A compilation of the launch times and the geophysical conditions encountered is given in Table 2. For details the reader is referred to the description of the project by Theile (1978).

Table 2. Launch times and geophysical conditions

Flight	Payload designation	Launch time (GMT)		Apogee (km)	Geophysical conditions
1	F2A	January 22, 1977	21:28:00	276.8	Pre-breakup phase
2	F1B	February 8, 1977	22:11:51	272.7	Active auroral display
3	F3C	February 20, 1977	21:14:00	256.0	Quiet diffuse aurora
4	F4D	March 16, 1977	22:04:40	259.3	Active auroral display

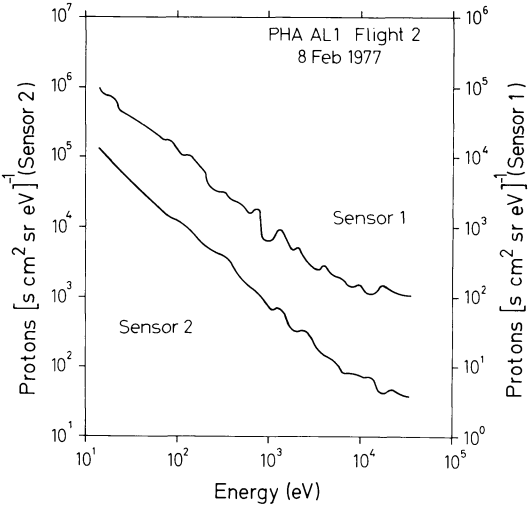


Fig. 3. Proton spectra observed during the second flight (F1B) above an altitude of 250 km. The graphs define an upper limit for the proton observations for fluxes below an energy of several keV. *Sensor 1* covered the upper halfsphere, whereas *sensor 2* mainly looked downwards. Notice that the scales for the two sensors are displayed by a factor of 10

With two exceptions the payloads performed nominally as far as the requirements of the low-energy particle experiment were concerned. A spin-modulated telemetry loss occurred during the later portions of the first flight (F2A) and payload 3 (F3C) was exposed to humidity during ground handling procedures resulting in an unacceptable dark count rate in one sensor system that, as a consequence, had to be disconnected before launch. All the other channels worked according to expectation. The pitch angle scans of the two sensor heads were rather restricted due to the stabilisation of the payloads under nominal conditions. In this sense, an accidentally induced high coning angle on the second flight (F1B) was profitable for this experiment.

Proton Flux Observations. The proton fluxes measured during all flights were very low and did not permit any significant time history study. The measurements can be summarized in two representative spectra that are given in Fig. 3. They have been produced by integrating all the proton measurements above 250 km during the second flight (F1B) separately for both directional channels. Taking the count statistics and the noise levels of the detectors into account, the spectra can only be regarded as significant above several keV proton energy. Below that energy, it can be concluded that the flux levels stayed below the threshold sensitivities of the detectors represented by the graphs.

Electron Distributions. Measurements of electron energy spectra will conveniently be displayed either as differential electron flux $F(\alpha, W)$ in units of electrons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}\,\mathrm{eV}^{-1}$ assuming azimuthal symmetry around the magnetic field direction or as velocity distribution function $f(\mathbf{v})$ in $\mathrm{s}^3/\mathrm{km}^6$ in order to study some of their characteristics. The symbols α, W, and $\mathbf{v}$ denote electron pitch angle, energy and velocity vector, respectively. The electron flux and the velocity distribution function are related by

$$f_e(\mathbf{v}) = m_e^2 F(\alpha, W)/2W \tag{1}$$

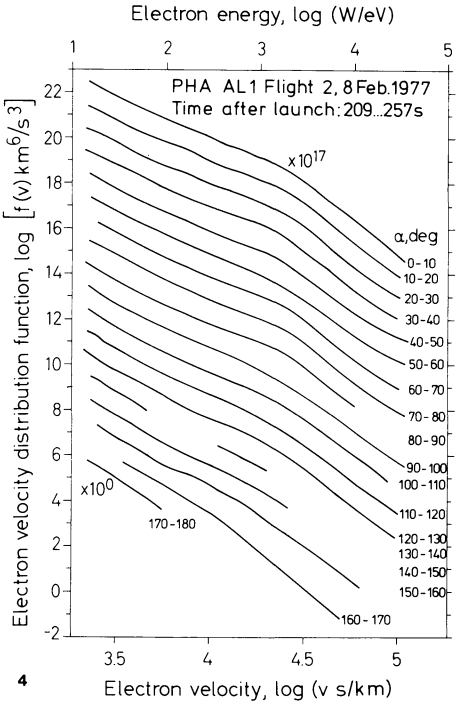


Fig. 4. Electron velocity distribution function plotted versus electron velocity (and energy) with pitch angle ranges as parameter. The format of the presentation allows 18 pitch angle ranges of 10 deg width each to be displayed by displacing the curves by a factor of 10 with respect to each other. The scale is given for the 170–180 deg curve. The observation were made in the time interval from 209 ... 257 s after launch of flight 2 (F1B). The instrument was looking upwards along the field line when 0 deg pitch angle particles were observed

with m_e the electron mass and $W = m_e v^2/2$ the non-relativistic electron energy. It will first be attempted to illustrate the normal electron velocity distribution function that was observed for most of the flight times. A typical observation is shown in Fig. 4. Characteristic for this distribution is a knee at approximately 3×10^4 km/s corresponding to $W \approx 2$ keV with a steep decrease to higher velocities. The differential energy spectrum of this distribution can be described by two power law functions as

$$F \propto W^{-\gamma} \tag{2}$$

with

$$\gamma = 1.0 \quad \text{for } W < 2\,\text{keV}$$

and

$$\gamma = 2.2 \quad \text{for } W > 2\,\text{keV}.$$

Except for a loss cone signature at high velocities the distribution is more or less isotropic in the range covered by the observations.

Near bright auroral arcs the distribution was similar but the knee shifted to higher velocities, the high-energy slope steepened and a plateau or a relative maximum developed below the knee velocity. Illustrations of this type of distribution are given in Figs. 5 and 6 in terms of the electron distribution

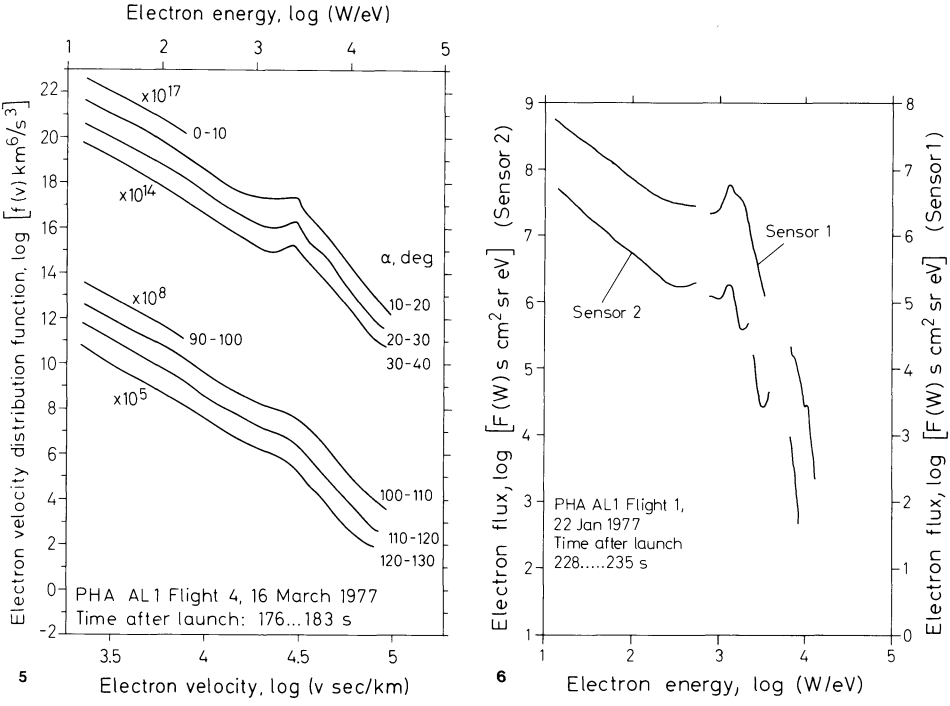
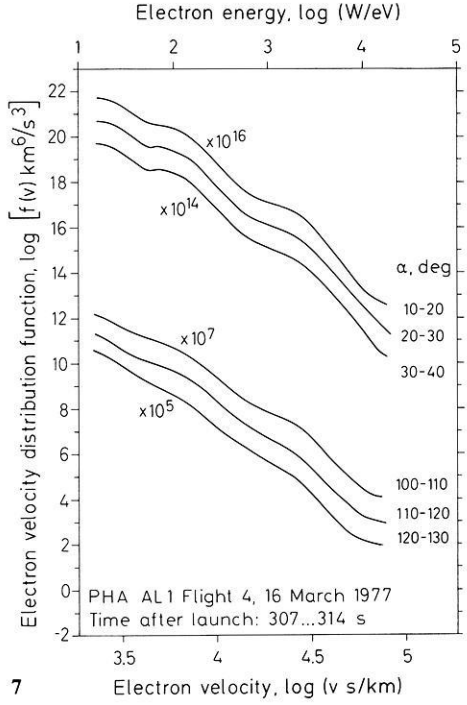


Fig. 5. Electron velocity distribution function between 176 ... 183 s after launch of F4D. For details of the format see Fig. 4. Due to the limited pitch angle coverage only 8 ranges are available for F4D

Fig. 6. Differential electron flux versus energy measured during the time interval from 228 ... 235 s on the first flight (F2A). The gaps result from data losses caused by the telemetry failure mentioned in Sect. 3. The *sensors 1 and 2* scanned pitch angle ranges of 0 to 40 and 90 to 130 deg, respectively

function and the differential energy spectrum. The fluxes exhibit field-aligned collimation near the peak intensities. Predominantly north of the auroral displays the distributions were more complex as can be seen from Fig. 7. Again the knee signature near 2 keV electron energy is present but, in addition, another change of slope of the velocity distribution function can be detected near 200 eV. Similar to the simple distribution in Fig. 4, it is typical for the double knee distribution that the electron fluxes are nearly isotropic, although a tendency of field-alignment can be noticed at low energies. Distributions such as those in Figs. 4 and 7 can be looked at as reference distributions. Superimposed on them, strongly field-aligned electron fluxes were observed on many occasions that exhibited a marked difference in their characteristics to the reference distributions. Figure 8 illustrates one of these events in terms of the differential energy flux in two different pitch angle ranges. The apparent spin modulation between 0.2 and 2 keV with peak fluxes near 0 deg pitch angle is the instrumental response to strongly field-aligned electron precipitation within this wide energy range. The degree of alignment can best be seen in flux versus pitch angle plots given in Fig. 9 for another event having similar features. For still



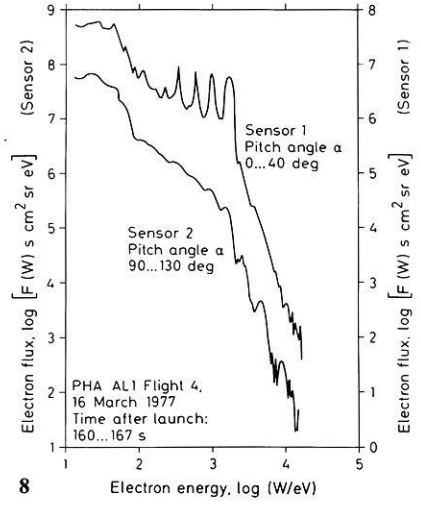


Fig. 7. Electron velocity distribution function observed from 307... 314 s after launch of F4D. For details see Fig. 4

Fig. 8. Electron differential energy spectra measured in two pitch angle ranges on F4D between 160 ... 167 s elapsed time

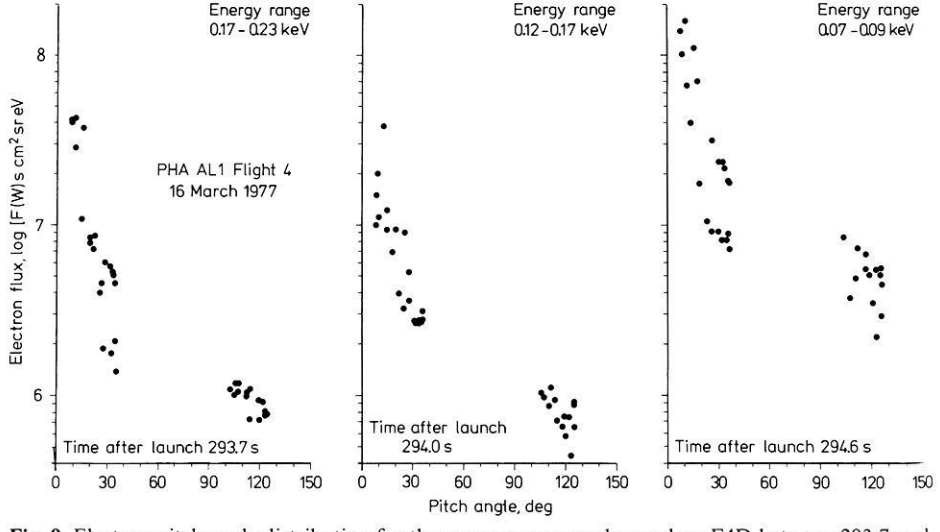


Fig. 9. Electron pitch angle distribution for three energy ranges observed on F4D between 293.7 and 294.6 s elapsed time

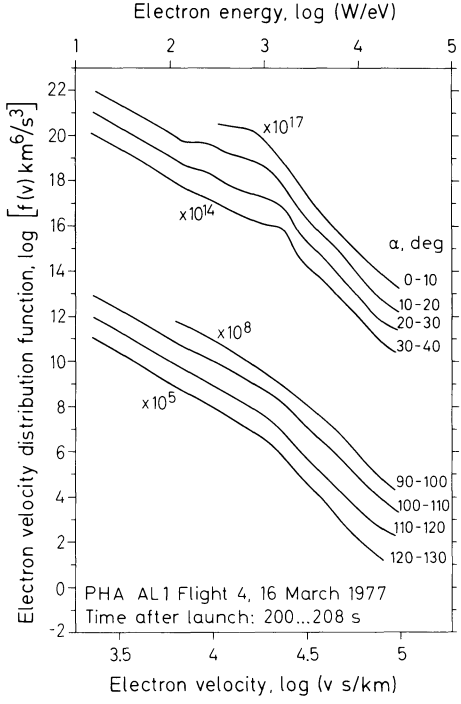


Fig. 10. Electron velocity distribution function in the time period from 200 ... 208 s after launch of F4D. For details see Fig. 4

another event the velocity distribution function has been plotted in Fig. 10 demonstrating that a pronounced density increase at small pitch angles is characteristic for this type of event. Moreover it should be noted that the increase is situated at velocities smaller than the knee velocity. Field alignment of the electron flux was present as low as $v = 8.4 \times 10^3$ km/s ($W = 200$ eV) in this example but could be observed down to $v = 3.2 \times 10^3$ km/s ($W = 30$ eV) on other occasions. Considering the low altitude at which the measurements were made and the mean free path length of the order of 10 km for the low-energy electrons involved, the existence of such strongly field-aligned fluxes is highly surprising and merits detailed investigations.

The knee distributions were very stable in time. The field-aligned events, on the contrary, exhibited a burst-like-structure with a duration of several seconds for individual events. With a view to understanding the causes of these events with strong field alignment of the electron flux, part of the data of Fig. 10 have been reorganized as contours of constant distribution functions in velocity space in Fig. 11. The solid lines represent those regions in velocity space that were directly observed by the instrument. The broken lines are interpolations believed to be possible without ambiguity. There are several interesting features that should be noted in Fig. 11:

– in the first quadrant the distribution function contours for energies greater than 2 keV are nearly circular;

– a loss cone effect is present for pitch angles greater than 105 deg in the second quadrant, corresponding to a local loss cone angle of 75 deg;

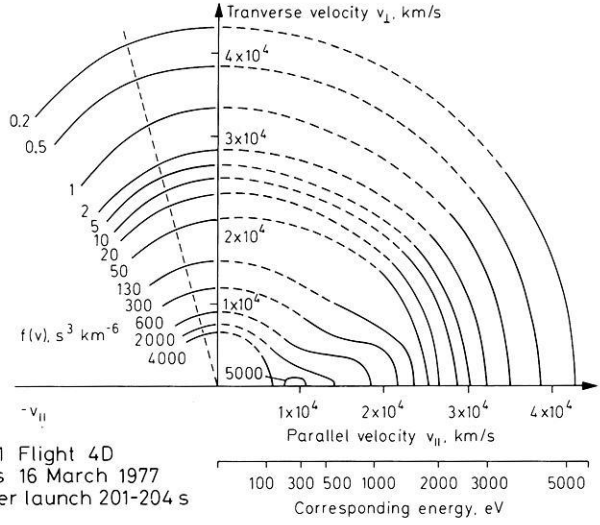


Fig. 11. Constant electron distribution function contours in velocity space. In addition to the velocity scales, the corresponding energy of electrons is given. *Left of the dashed line*, intensity decreases can be seen caused by loss cone effects. The distrition was measured on F4D between 201–204 s elapsed time

– the field-aligned event clearly stands out along the $+v_\parallel$ axis with a relative maximum at $v \approx 10^4$ km/s.

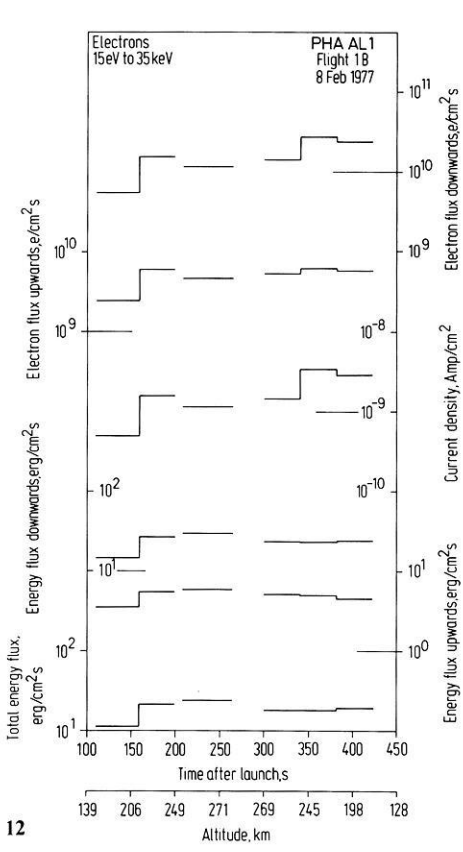
Electron Number and Energy Fluxes. In order to arrive at the bulk properties of the electron flux, the differential electron flux $F(\alpha, W)$ can be integrated to give the number and energy flux densities according to

$$n_{\text{down}} = 2\pi \int_0^{\pi/2} \int_{W_1}^{W_2} F(W, \alpha) \sin\alpha \cos\alpha \, dW d\alpha \tag{3}$$

$$Q_{\text{down}} = 2\pi \int_0^{\pi/2} \int_{W_1}^{W_2} WF(W, \alpha) \sin\alpha \cos\alpha \, dW d\alpha \tag{4}$$

Similarly, the upwards going fluxes can be determined. The electric current density carried by electrons in the energy range under consideration then follows from the difference of the two opposite number flux densities. As the instrumental conditions differed for all four flights, a separate discussion is required in each case.

Starting with the second flight (F1B), it should be recalled that the large coning angle of the payload was favourable as far as a complete pitch angle scan of this instrument was concerned. However, as another consequence, the time resolution was substantially deteriorated. Fortunately, time variations of the electron flux were not very pronounced during this flight, thus allowing a long integration time without serious implications. The results are presented in Fig. 12. Optical observations showed that the payload was engulfed in auroral



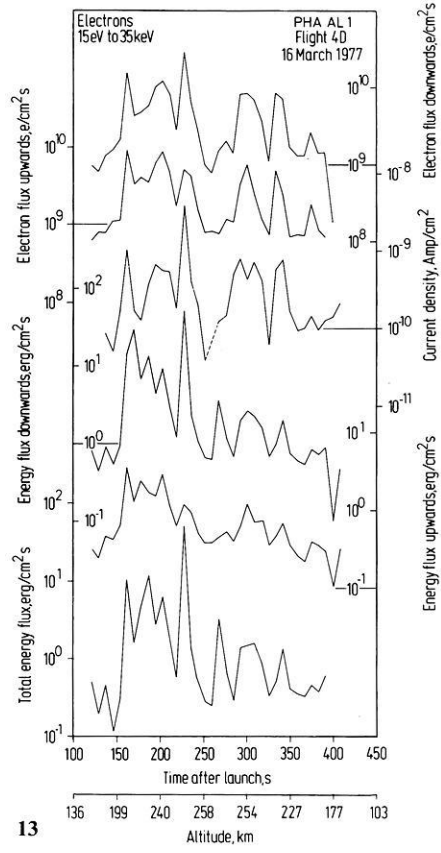


Fig. 12. Electron flux parameters for the second flight (F1B) on February 8, 1977. The upper three histograms show the electron flux densities and the lower ones indicate the corresponding energy fluxes. Time variations of the electron flux in certain intervals prevented long integration times and consequently no flux densities and energy fluxes could be computed

Fig. 13. Electron flux parameters for the fourth flight (F4D) on March 16, 1977

activity with an intensity of 3 to 4 kR at 557.7 nm for all of the flight time (Theile, private communication).

On the fourth flight (F4D), the pitch angle scan was restricted to intervals of 2×40 deg. Before the integration of Eqs. (3) and (4) could be performed, inter- and extrapolation methods had to be used following a procedure discussed by Wilhelm (1979). It was shown that the computations could be considered with confidence for spectra with reasonably smooth behaviour. In Fig. 13 the results of F4D have been compiled. The dramatic increase of all quantities between approximately 150 and 240 s elapsed time coincided with the encounter of the payload with the visible auroral are as can be seen from Fig. 14 where the intensities contours derived from 557.7 nm scanning photometer measurements have been plotted as a function of time and geographic location.

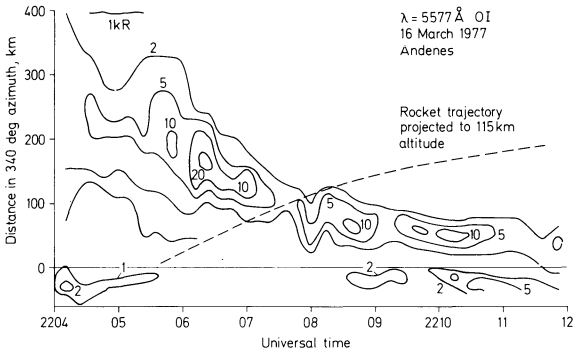


Fig. 14. Photometer observations of the 557.7 nm oxigen line during the flight time of F4D. The scan was performed in the 340 deg azimuth plane corresponding to the launch directions of the sounding rocket. The trajectory of the payload projected down to 115 km altitude has also been included

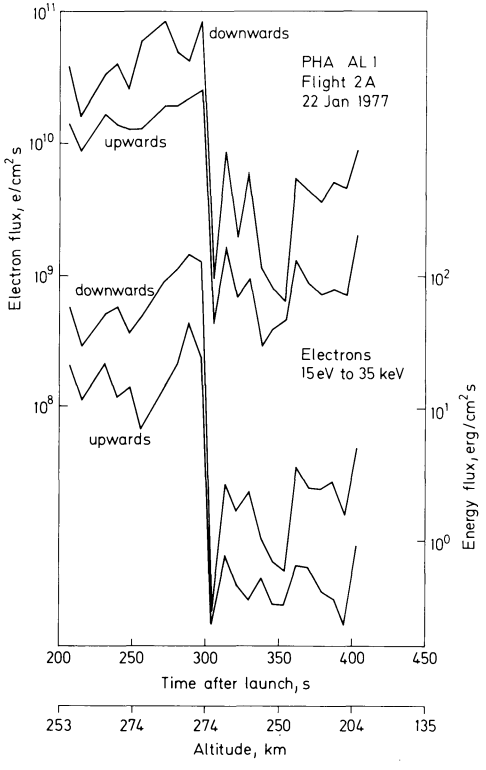


Fig. 15. Electron flux parameters for the first flight (F2A) on January 22, 1977

The technical limitations prohibited a numerical evaluation for both the third (F3C) and the first (F2A) flight. Graphical methods were therefore used that could only provide an estimate of the quantities under consideration. The results have been displayed in Fig. 15 for F2A. Of particular interest seems to be the decrease near 300 s after launch. Even if the uncertainties in absolute values

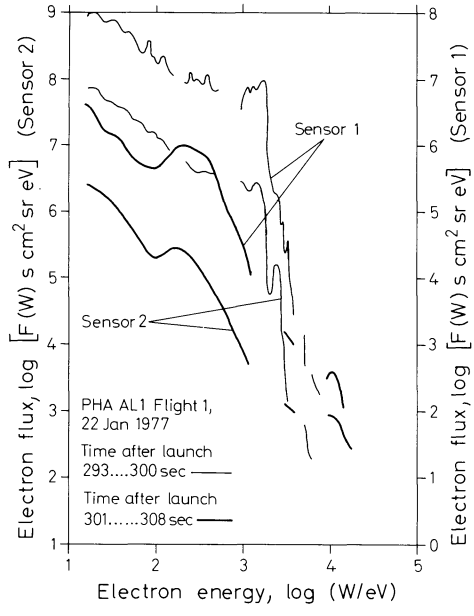


Fig. 16. Electron energy spectra before and after the flux intensity decrease at 300 s elapsed time on F2A

should be considerable, the relative decrease of more than two orders of magnitude in energy flux can be considered as genuine. The spectra before and after that discontinuity have been plotted in Fig. 16.

4. Discussion

There is increasing evidence supporting the hypothesis that magnetospheric charged particles are being accelerated or post-accelerated near the ionosphere before they precipitate into the atmosphere (Maehlum and Moestue, 1973; Evans, 1974; Mozer, 1976; Raitt and Sojka, 1977; Bryant et al., 1978; Mizera and Fennell, 1977). The signature of electron distributions should be a sensitive indicator of fieldaligned static electric field acceleration as was discussed by Whipple (1977) for natural and Wilhelm (1977) for artificially injected particles. Under the assumption of adiabatic motion, equivalent to the constancy of the magnetic moment

$$\mu = \{W_0 - q[\phi(s) - \phi(s_0)]\} B^{-1}(s) \sin^2 \alpha \qquad (5)$$
$$= m_e v_\perp^2(s)/2B(s) = \text{const}$$

with

W_0 initial kinetic energy
q charge of particle
$\phi(s)$ electric potential at position s
$B(s)$ magnetic field strength at s
$v_\perp(s)$ perpendicular velocity with respect to the magnetic field at s

the main conclusion was that the presence of an accelerating electric field parallel to the magnetic field would subdivide the particle population into several distinct regimes in velocity-space. The identification of these regimes and the corresponding boundaries in the observed particle distributions should thus be indicative of the operation of the assumed mechanisms. Chiu and Schulz (1978) have given a full description of the different phase space regimes and the corresponding demarcation lines. Using their analysis, Croley et al. (1978) have pointed out that the region in velocity space which defines the trajectories of charged particles that are turned around by an electric field above the point of observations s_1 is given by the interior of the ellipse

$$v_{\parallel}^2(s_1)+[1-B(s_0)\,B^{-1}(s_1)]\,v_{\perp}^2(s_1)=-2\,q\,\phi(s_1)/m_e \tag{6}$$

with $v_{\parallel}(s_1)$ parallel velocity with respect to the magnetic field at s_1 and $B(s_0)$ magnetic field strength where the potential function ϕ vanishes.

Similarly, the loss of particles in the atmosphere at the position s_2 can be projected upwards by

$$v_{\parallel}^2(s_1)+[1-B(s_2)\,B^{-1}(s_1)]\,v_{\perp}^2(s_1)=2q\,[\phi(s_2)-\phi(s_1)]/m_e \tag{7}$$

resulting in a region bounded by a hyperbola at the altitude of observation s_1. Furthermore, Croley and co-workers have presented an example of electron flux measurements in their Fig. 1b that showed, except for the loss cone effect, circular contour lines for $v>1.87\times10^4$ km/s and a sharp demarcation line at that velocity. They interpret these findings as evidence of a field-aligned electrostatic acceleration of magnetospheric electrons in a potential drop of 1 kV above the satellite. The circular domain boundary could be described by Eq. (6) under the assumption of $B_0 \ll B(s_1)$, i.e., a rather high upper limit of the electric field region.

The observations presented in Fig. 11 of this paper resemble those discussed above to a high degree for electron velocities greater than 2.5×10^4 km/s. Similarly, it can be seen from Figs. 5 and 6 that presentations of the same data in velocity space would result in circular demarcation lines at velocities near 3×10^4 km/s. As far as the high-velocity electrons are concerned, the measurements obtained during the time periods discussed are thus consistent with an electrostatic field acceleration. It can, therefore, be concluded that the distribution shown in Figs. 5 and 6 are indicative of a field-aligned acceleration process operating at great altitudes. All electrons with velocities greater than the knee velocity were of magnetospheric origin and fell through the total potential drop. Electrons with small velocities, on the other hand, belonged to an ionospheric population unable to surmount the potential barrier above the point of observation as suggested by Evans (1974).

The distribution shown in Fig. 7 exhibits a double knee structure and, transferred into velocity space contours, could be characterized by two circular boundaries. As an explanation it is suggested that the cause be sought in two spatially separated electric field regions along the field line of observation. High-velocity electrons were of magnetospheric origin and experienced both acceleration steps, low-velocity ionospheric electrons were reflected by the low-

altitude field, whereas electrons with intermediate velocities originated from the region between the potential gradients.

The ionospheric population consisting of secondary and degraded primary electrons (Banks et al., 1974) was in most cases observed to be rather uniform and, in particular, did not indicate any distinction between particles trapped between the electric field region and the magnetic mirror point and reflected ones. This is in contrast to the findings of Croley and co-workers and might be explained by the low height of the sounding rocket observing point compared to the satellite altitude of several thousand kilometres. The most important difference between the satellite observations as presented by Croley et al. (1978) and the sounding rocket measurements, however, manifests itself in the field-aligned electron flux at ionospheric altitudes. An understanding of this signature of the electron distributions can be achieved by generalising the analysis by Croley and co-workers and modifying the model calculations of Kaufmann et al. (1976) as follows:

For simplicity, a Maxwellian distribution with temperature T and density n_e will be assumed above the electric field region at a station s_0 along the magnetic field line

$$f_e = n(m_e/2\pi kT)^{3/2} \exp(-W_0/kT) \tag{8}$$

where $W_0 = m_e\,[v_\parallel^2(s_0) + v_\perp^2(s_0)]/2$ is the initial energy of a given particle. Constant f_e contours thus are circles in velocity space with radii

$$r_0 = (2W_0/m_e)^{1/2} \tag{9}$$

The question at hand is, how contours of constant f_e project downwards to station s_1 in a converging magnetic field configuration with an accelerating potential gradient. It can immediately be seen from

$$\begin{aligned} v_\parallel^2(s_1) + v_\perp^2(s_1) &= 2\{W_0 - q[\phi(s_1) - \phi(s_0)]\}/m_e \\ &= r_1 \end{aligned} \tag{10}$$

and Liouville's theorem that again circles with radii r_1 describe the resulting projection. It has, however, to be noted that the maximum of $v_\perp$ at s_0 equals r_0 of Eq. (9). As a consequence of the adiabatic assumption in Eq. (5), there exists, independent of the constraint resulting from Eq. (10), a maximum for $v_\perp(s_1)$, namely

$$\mathrm{Max}\,[v_\perp(s_1)] = r_0\,[B(s_1)\,B^{-1}(s_0)]^{1/2} \tag{11}$$

Combining Eqs. (10) and (11), an essential distinction can be defined by

$$\mathrm{Max}\,[v_\perp(s_1)] \gtrless r_1. \tag{12}$$

If, as case 1, both sides are equal, a one-to-one map of the object circle onto the image circle is involved. If, in case 2, the maximum of $v_\perp(s_1)$ is greater than r_1, portions of the object circle are projected on the full image circle. Particles populating the remaining parts of the object circle have no access to the altitude

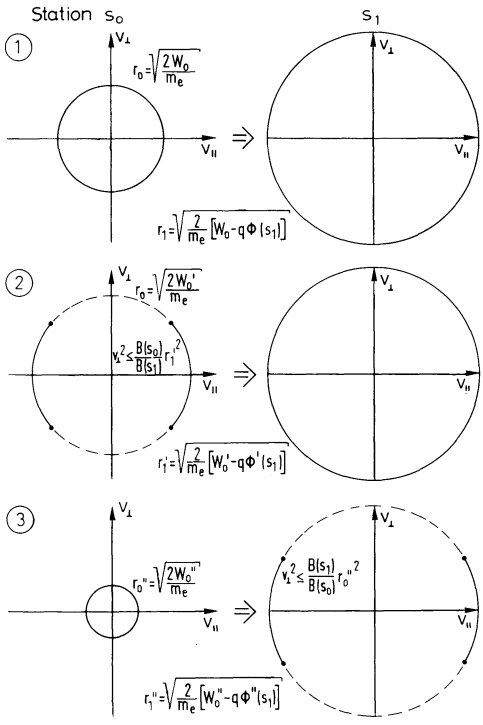


Fig. 17. Schematic illustration of projections in velocity space for the three different cases discussed in the text. In this figure, it was assumed that ϕ vanishes at s_0

of observation as their magnetic mirror points lie above that height. Should, finally, as case 3, the left-hand side be smaller than the radius r_1, the complete object circle would be mapped into portions of the image circle. The three cases are demonstrated in Fig. 17 assuming three different initial energies and specific acceleration potentials. Atmospheric loss processes have not been taken into account in this schematic diagram. Mapping of Eq. (8) under case 3 assumptions for the energy range $0 \leqq W \leqq W_0$ will thus result in an area in velocity space similar to a crescent moon that is bounded by the ellipse of Eq. (6) and the circle of Eq. (10) with $\phi(s_0)=0$.

By rewriting Eq. (12) one obtains

$$B(s_1)\,B^{-1}(s_0)-1 \gtrless -q\,[\phi(s_1)-\phi(s_0)]\,W_0^{-1} \tag{13}$$

The right-hand side is the ratio of the acceleration energy to the initial energy of an electron, whereas the left side can be regarded as a measure of the acceleration distance along the field line. Applying these projections to magnetospheric Maxwellian electron populations, it can be concluded that concentric circles would be observed as constant distribution graphs around the origin, if either the accelerating potential was small compared to the initial energy or the distance along the field line was very long. Conversely, crescentlike projections would be generated when the opposite statements should be true.

As was already discussed, most of the observed constant distribution diagrams contain concentric circular contours with distinct demarcation lines and loss cone effects at high velocities. In the light of the foregoing discussion, this signature suggests, consistent with the conclusion reached earlier, an electrostatic acceleration of hot magnetospheric electrons down to the observing point.

Field-aligned distributions such as that in Fig. 11 have quite frequently been observed superimposed on the more regular distributions. It should be noted that they are characterized by a broad maximum over a wide range in parallel velocity sharply confined to small perpendicular velocities. It is clear that this is the velocity space region that would be occupied by case 3 projections. The wide range in parallel velocity, however, precludes the assumption that all electrons have been accelerated by the total potential drop. Bryant et al. (1978) suggested a time-dependent acceleration or energy dissipation process in order to explain the high flux values at energies below the peak of the spectrum. The measurements presented here confirm earlier observations that these fluxes cannot readily be understood by the reflection model of Evans (1974). However, no indication of fast varying precipitation could be detected either given the time resolution of this experiment. In a simple model it might be assumed that thermal electrons inside the electric field region can under certain conditions be accelerated as well. In accordance with their point of origin these run-away electrons (Mozer, 1976) only gain the corresponding portion of energy giving rise to the extended range in energy observed for these field-aligned events. If, in addition, the temperature of the source plasma was sufficiently low or if the starting point of the acceleration occurred sufficiently near the observing site, the perpendicular velocities would be restricted as required for field-aligned precipitation events. As an abundant supply of cold electrons can only be expected to be available in or near the ionosphere, both conditions point to a local electric field acceleration.

The lack of significant low-energy proton fluxes observed on all four flights might be considered as another indication of electric field acceleration. Even taking into account the effects of charge exchange in the atmosphere that would have neutralized more than 90% of a 2 keV primary proton beam before reaching the detector system (Miller and Whalen, 1976), the measurements support the existence of a strongly selective acceleration process. The restricted pitch angle coverage on the three flights 2A, 3C, and 4D unfortunately prevented the decisive observation of upwards streaming protons as a consequence of a local parallel electric field.

In this context, it is also interesting to study the abrupt changes in the bulk properties of the electron flux. In particular, the sharp decrease at 300 s elapsed time of flight 2A in Fig. 15 suggests that the detector encountered a completely different electron population. On the basis of the optical observations it was concluded that the variation could be interpreted as a spatial structure that was transversed by the payload. Given this dramatic change, it is surprising to note that the electron energy spectra before and after the decrease shown in Fig. 16 can approximately be derived from the same distribution function by an accelerating potential of 0.25 and 1.5 kV, respectively, at least for energies above the peak fluxes.

Field-aligned electron fluxes in the ionosphere led Raitt and Sojka (1977) to the conclusion that local electrostatic field acceleration would be consistent with their observations. They considered atmospheric interactions of the electrons and argued that the degree of alignment observed could not be explained without local acceleration or post-acceleration. The observed and calculated pitch angle distribution did, however, not exhibit the steep increase near 0 deg shown in Fig. 9. Without detailed discussion of the prevailing geophysical conditions, it appears to be difficult to decide whether or not the effects of the same process have been detected in both investigations. Nevertheless, it can be concluded that an electrostatic electric field acceleration process in or near the ionosphere is able to explain a variety of observations and thus is a strong candidate for causing field-aligned auroral electron precipitation.

Acknowledgements. The contributions of C. Becker, R. Schmidt, and W. Engelhardt during the instrument development and operation phases are greatly acknowledged. I thank H.-J. Müller and H.J. Baur for their assistance in the data analysis. The experiment was financially supported by the German Bundesministerium für Forschung und Technologie through DFVLR-PT under grant No. RV 14 - B 112/73 - AI/AL 1.

References

Arnoldy, R.L., Lewis, P.B., Issaacson, P.O.: Field-aligned auroral electron fluxes. J. Geophys. Res. **79**, 4208–4221, 1974

Banks, P.M., Chappell, C.R., Nagy, A.F.: A new model for the interaction of auroral electrons with the atmosphere: Spectral degradation, backscatter, optical emission, and ionization. J. Geophys. Res. **79**, 1459–1470, 1974

Boyd, J.S.: Rocket-borne measurements of auroral electrons. Rev. Geophys. Space Phys. **13**, 735–740, 1975

Bryant, D.A., Hall, D.S., Lepine, D.R.: Electron acceleration in an array of auroral arcs. Planet. Space Sci. **26**, 81–92, 1978

Cahill, L.J., Potter, W.E., Kintner, P.M., Arnoldy, R.L., Choy, L.W.: Field-aligned currents and the auroral electrojet. J. Geophys. Res. **79**, 3147–3154, 1974

Chiu, Y.T., Schulz, M.: Self-consistent particle and parallel electrostatic field distributions in the magnetospheric-ionospheric auroral region. J. Geophys. Res. **83**, 629–642, 1978

Croley, D.R., Mizera, P.F., Fennell, J.F.: Signature of a parallel electric field in ion and electron distribution in velocity space. J. Geophys. Res. **83**, 2701–2705, 1978

Evans, D.S.: Precipitating electron fluxes formed by a magnetic field aligned potential difference. J. Geophys. Res. **79**, 2853–2858, 1974

Kaufmann, R.L., Walker, D.N., Arnoldy, R.L.: Acceleration of auroral electrons in parallel electric fields. J. Geophys. Res. **81**, 1673–1682, 1976

Maehlum, B.N., Moestue, H.: High temporal and spatial resolution observations of low energy electrons by a mother-daughter rocket in the vicinity of two quiescent auroral arcs. Planet. Space Sci. **21**, 1957–1967, 1973

Mende, S.B., Shelley, E.G.: Coordinated ATS 5 electron flux and simultaneous auroral observations. J. Geophys. Res. **81**, 97–110, 1976

Mizera, P.F., Fennell, J.F.: Signatures of electric fields from high and low altitude particles distribution. Geophys. Res. Lett. **4**, 311–314, 1977

Miller, J.R., Whalen, B.A.: Characteristics of auroral proton precipitation observed from sounding rockets. J. Geophys. Res. **81**, 147–154, 1976

Mozer, F.S.: Anomalous resistivity and parallel electric fields. In: Magnetospheric particles and fields, B.M. McCormac, pp. 125–136. ed.: Dordrecht-Holland: D. Reidel Publ. Comp. 1976

O'Brien, B.J., Reasoner, D.L.: Measurements of highly collimated short-duration bursts of auroral electrons and comparison with existing auroral models. J. Geophys. Res. **76**, 8258–8278, 1971

Peterson, W.K., Doering, J.P., Potemra, T.A., McEntire, R.W., Bostrom, C.O., Hoffman, R.A., Janetzke, R.W., Burch, J.L.: Observations of 10-eV to 25-keV electrons in steady diffuse aurora from Atmosphere Explorer C and D. J. Geophys. Res. **82**, 43–47, 1977

Raitt, W.J., Sojka, J.J.: Field-aligned suprathermal electron fluxes below 270 km in the auroral zone. Planet. Space Sci. **25**, 5–13, 1977

Theile, B.: Initial results of the sounding rocket campaign 'Polar High Atmosphere', European Space Agency SP-135, pp. 113–118, 1978

Whalen, B.A., McDiarmid, I.B.: Observations of magnetic-field-aligned auroral electron precipitation. J. Geophys. Res. **77**, 191–202, 1972

Whipple, E.C. Jr.: The signature of parallel electric fields in a collisionless plasma. J. Geophys. Res. **82**, 1525–1531, 1977

Wilhelm, K.: Remote sensing experiment for magnetospheric electric fields parallel to the magnetic field. J. Geophys. **43**, 731–750, 1977

Wilhelm, K.: Beobachtungen von suprathermischen Polarlichtteilchen, Bundesministerium für Forschung und Technologie, Forschungsbericht W 79-03, 1979

Received February 7, 1979; Revised Version June 11, 1979; Accepted June 30, 1979

J. Geophys. 46, 171–183, 1979

Journal of Geophysics

The Shape of the Cosmic Ray Modulation Region of the April 30, 1976, Event, as Deduced From HELIOS-1, HELIOS-2, IMP-8, and Neutron Monitor Data

A. Geranios

Max-Planck-Institut für Extraterrestrische Physik, D-8046 Garching bei München, Federal Republic of Germany

Abstract. During STIP (Study of Travelling Interplanetary Phenomena) II period, several disturbances of the cosmic ray flux have occured due to the enhanced activity of the sun, although this period (March-May 1976) was at the end of the past solar cycle. One of the most prominent 'events' was caused by the 2 B solar flare on 30 April (121st day of the year), at 20:47 UT, with coordinates S8 W46 in Mc-Math region 14179. This event has been 'seen' as a tangential discontinuity by both HELIOS spacecraft and as a shock wave by IMP-8. Neutron monitors recorded an increase of the counting rate on the ground level, and two days later, the onset of a Forbush decrease.

In this paper, we estimate the extension of the modulation region of cosmic rays in the ecliptic plane by correlating the observations which are related to the above mentioned solar flare.

We conclude that in the ecliptic plane the shock characteristics of this event are found only in the neighbourhood of the heliographic longitude of the causitive solar flare, i.e., of an extension 45° and 120°. The tangential discontinuity observed by HELIOS-2 and then by HELIOS-1 was of a corotative nature. The feature of the tangential discontinuity and the Forbush decrease suggest a corotation of the above modulation region.

Key words: Cosmic rays – Solar wind protons.

A Brief Description of the Event

On April 30, 1976, at 20:48 UT, day of the year 121 (shortly DOY 121), a 2B solar flare (S.FL.) occurred at S8 W46 in Mc-Math region 14179 (Dodson and Hedeman, 1977). Two days later, HELIOS-1 and -2 were radially lined up about 120° west of the flare site at radial distances of 0.67 (H-1) and 0.43 (H-2) astronomical units, respectively from the sun (Fig. 1). About 20 minutes after the occurence of the flare, a Ground Level Event (GLE) was detected

0340-062X/79/0046/0171/$02.60

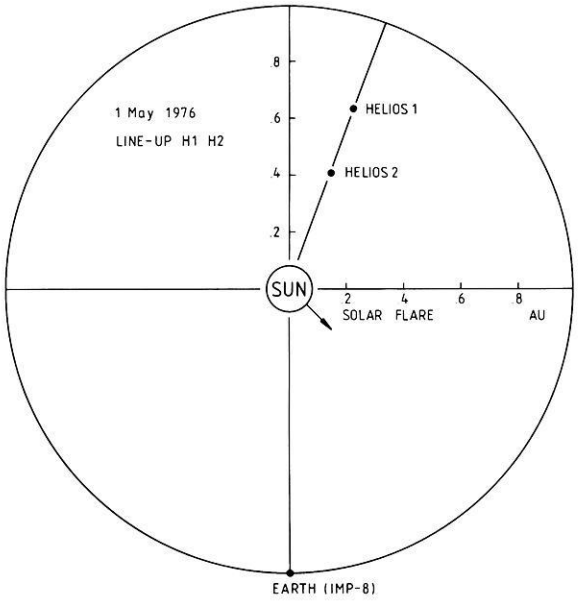


Fig. 1. The longitudinal distribution of *HELIOS-1*, *HELIOS-2*, IMP-8 and the solar flare

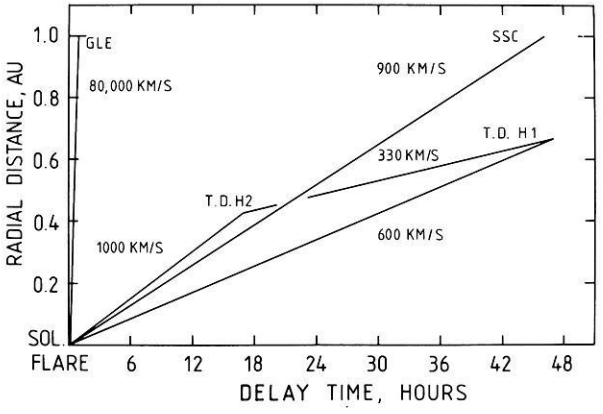


Fig. 2. The detection delay-time for the ground level event, the tangential discontinuity and the shock, or the storm sudden commencement, by neutron monitor, HELIOS-1, HELIOS-2, IMP-8 spacecraft

at the earth by neutron monitors (NM) with a cut-off rigidity below 3 GV (Duggal and Pomerantz, 1977), (Fig. 3). Later, during the same day, the electron, proton, and α-particle channels of IMP-8 detected abrupt increases (Armstrong et al., 1977). HELIOS-2 (DOY 121) and -1 (DOY 123) detected electron- and proton- but not α-particle intensity enhancements (Kunow et al., 1977). A tangential discontinuity (TD) passed HELIOS-2 and -1, 17 and 46 h respectively after the onset of the solar flare (Schwenn et al., 1977). A Storm Sudden Commencement (SSC, DOY 123, 18:29 UT), marked the passage of a shock front at the earth. This shock was also detected by IMP-8, and the commencement of a typical Forbush decrease (FD) was observed by ground NM's, (Fig. 3). A series of other correlated observations (For example type II and IV radio bursts, X-ray enhancements) are described elsewhere (Shea, 1977).

The Ground Level Event

GLE's are generally rare, especially during a minimum of solar activity. Such an event occured at the end of the past solar cycle on April 30, 1976, DOY 121, at about 21:00 UT, only about 20 min after the onset of the above mentioned flare. The enhancement of the energetic proton fluxes ejected from the solar flare was detected by low cut-off rigidity NM stations, for example Deep River (Fig. 3), Calgary and Sulphur Mountain showed relative increases of 3%, 5%, and 7% respectively (SGD, 1976). The maximum energy of the accelerated protons in this flare was 2 or even 3 GeV (Duggal and Pomerantz, 1977).

The longitudinal site of the solar flare (46°) was favorable for such a short delay time, which suggests that the particle flux was directed along the interplanetary magnetic field (IMF) lines and very little scattering occured.

According to Chirkov et al., 1977, the integral energy spectrum of the cosmic rays of this event discussed here, was:

$$A(E > 1.3 \text{ GeV}) \sim E \exp(-2.6)$$

for energies $E > 1.3$ GeV.

A similar event which orginated also from a S.FL. in the western hemisphere occured during a complex event in August 1972, where the 3B S.FL. (August 7, 37° W) caused the Ground level Event with a time-delay of 30 min and a cosmic ray increase of 8% (Mihalov et al., 1973; Agrawal et al., 1974; Pomerantz and Duggal, 1974; Tanskanen et al., 1973).

Observations of Particles in the Energy Range Between 2 and 37, and Greater Than 51 MeV/Nucleon

All three spacecraft measured an enhancement of energetic particles. HELIOS-2 and HELIOS-1 (shortly, H2 and H1) observed proton flux increases on April 30 (DOY 121) and on May 2 (DOY 123) respectively in the energy channels:

4–13 MeV
13–27 MeV
27–37 MeV of H1, and

4–13 MeV
27–37 MeV
>51 MeV of H2 (Kunow et al., 1977).

At the end of the first of May (DOY 122), IMP-8 detected abrupt increases in the proton energy channel 2–4.6 MeV, in addition, it recorded flux enhancements of α-particles of energy 1.8–4.2 MeV/nucl. later during the same day.

A main difference between the HELIOS and the IMP-8 observations was the lack at both HELIOS of any α-particle flux enhancement. These flux increases normally are regarded as a signature of a flare shock wave. Typical differences between the observations of H1 and H2 are:

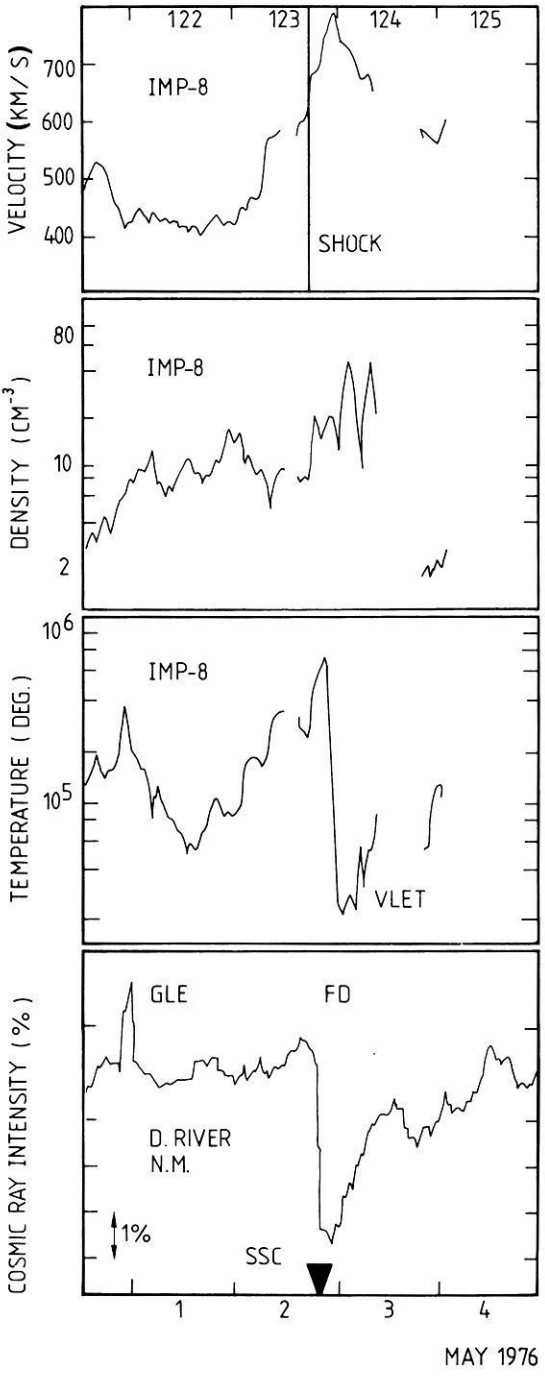


Fig. 3. Solar wind velocity, density and temperature measured by *IMP-8;* cosmic ray intensity ($E > 1$ GeV) measured by the *D. River* neutron monitor and *SSC* around the time of shock on May 2, 1976

1. The proton flux increase in the energy channel $E > 51$ MeV was clearly detected only by H2 (Kunow et al., 1977), and

2. the rise time of the proton flux increase observed by H2 compared with that of H1 was shorter by about a factor of 3, (Fig. 4a and b).

The difference between the intensity profiles during the increase phase detected by H2 and H1 may be due to the different distances of H2 and H1 from the sun.

A step-like proton intensity enhancement can be recognized as a superposition on the intensity profile during the increase phase at both H1 and H2. This step-like increase coincides with that of the solar wind velocity observed also by H1 and H2 (Fig. 4a and b).

The Tangential Discontinuity

After the occurence of the 2B S.FL. a tangential discontinuity (TD) was observed by H2 and H1 on May 1 (DOY 122) at 14:00 UT and on May 2 (DOY 123) at 19:30 UT, respectively. Apparently, the solar wind parameters are not shock characteristics, as will be explained further (Fig. 4a and b). The shape of the discontinuity is observed to be quite similar on both HELIOS spacecraft, which were about 0.25 AU distant from each other on a radial line (Schwenn et al., 1977).

According to Burlaga's classification (1968), this TD is of type T3.

For a check, the total pressures (P) just before and after the TD are calculated using preliminary solar wind data (Schwenn, 1977a; Neubauer, 1977).

$$P_1 = N_1\, k(T_1(e) + T_1(p)] + \langle B_1^2 \rangle / 8\pi$$

$$P_2 = N_2\, k[T_2(e) + T_2(p)] + \langle B_2^2 \rangle / 8\pi$$

N, k, $T(e)$, $T(p)$, $\langle B \rangle$ are the number density, the Boltzmann constant, the electron temperature, the proton temperature and the magnitude of the IMF respectively.

Parameters before the TD are marked with index 1, and after the TD with index 2. At the moment, no electron temperature data from H1 and H2 are available, therefore, we assume the following relation between the electron and proton temperatures:

$$T(e) = 1.2\, T(p) \qquad \text{(Burlaga, 1969a)}$$

and we find

for H1 $P_1 = 0.92\, P_2$ and
for H2 $P_1 = 1.3\, P_2$.

These pressures may be considered as equal within the errors of measurements, if we take under consideration that for usual shocks the total pressure in the shocked plasma is an order of magnitude higher than in the unshocked plasma.

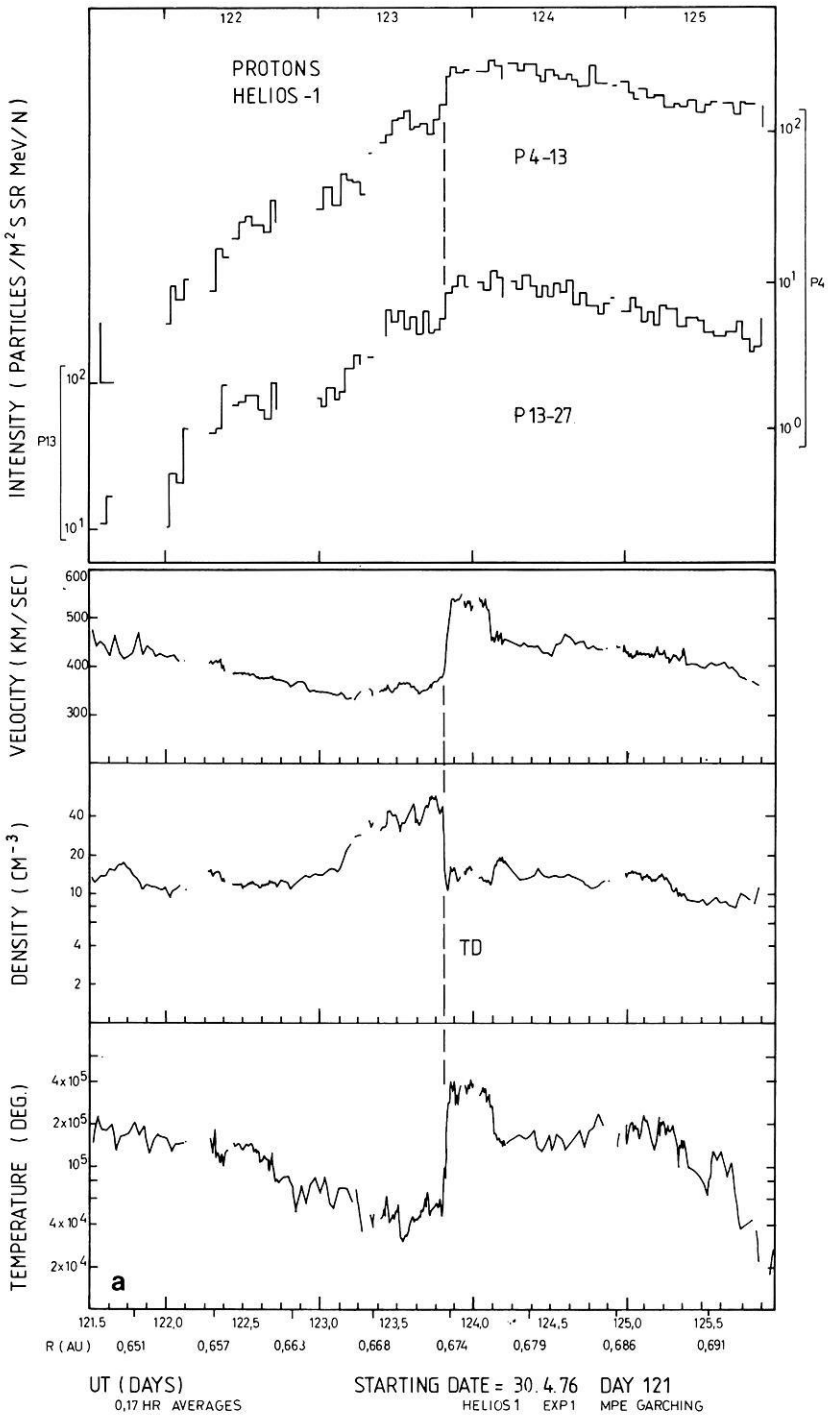
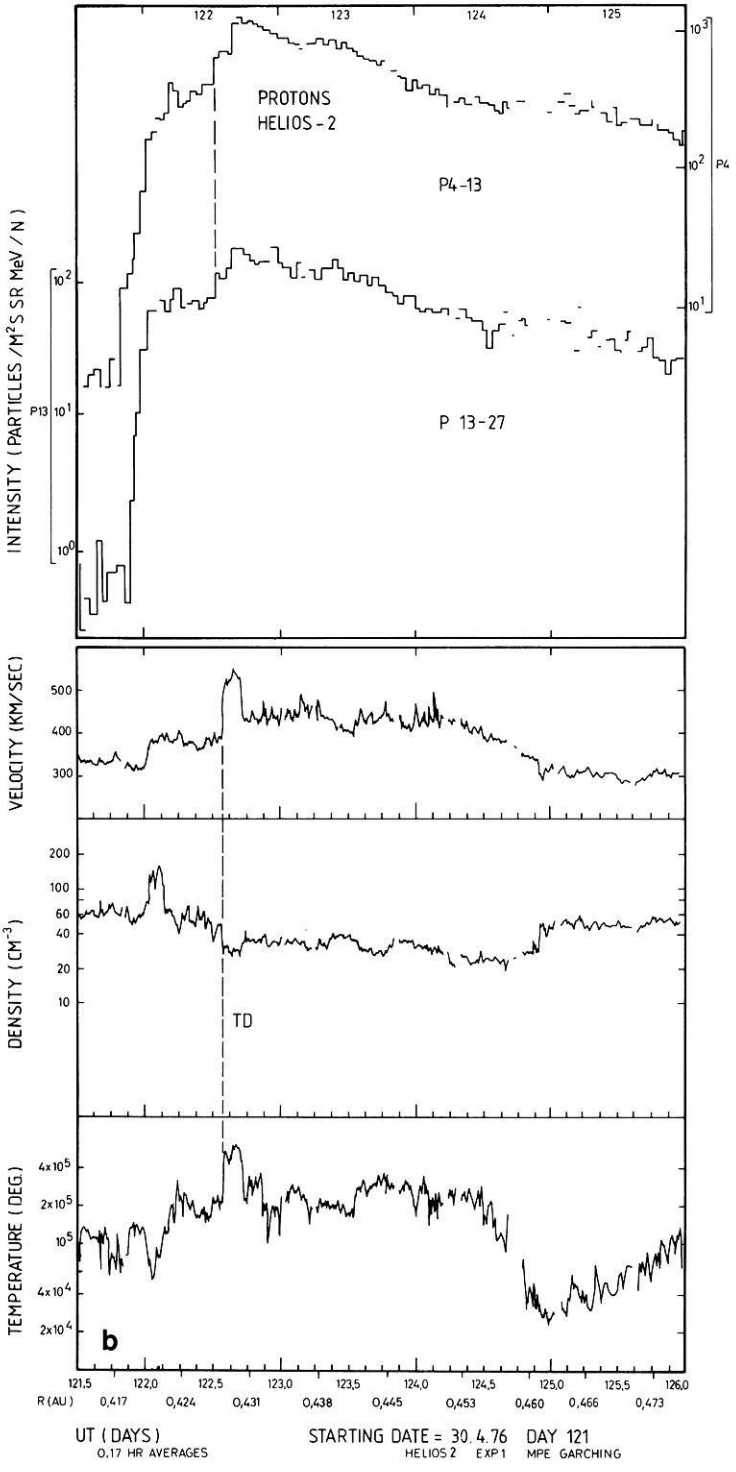


Fig. 4. **a** Solar wind proton parameters measured by HELIOS-1. **b** Solar wind proton parameters measured by HELIOS-2


PROTONS
HELIOS - 2
P4 -13
P 13 - 27
INTENSITY (PARTICLES /M²S SR MeV / N)
VELOCITY (KM/SEC)
DENSITY (CM⁻³)
TEMPERATURE (DEG.)
TD
b
R (AU)
UT (DAYS)
0.17 HR AVERAGES
STARTING DATE = 30.4.76 DAY 121
HELIOS 2 EXP 1 MPE GARCHING

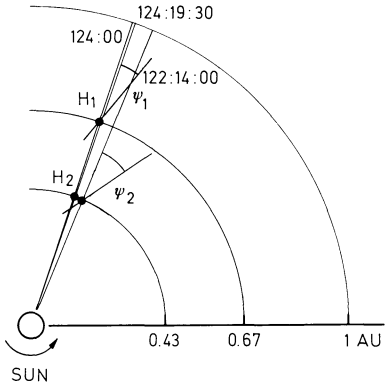


Fig. 5. The calculated angles ψ_1 and ψ_2 of the orientation of the tangential discontinuity detected by *HELIOS-1 HELIOS-2*

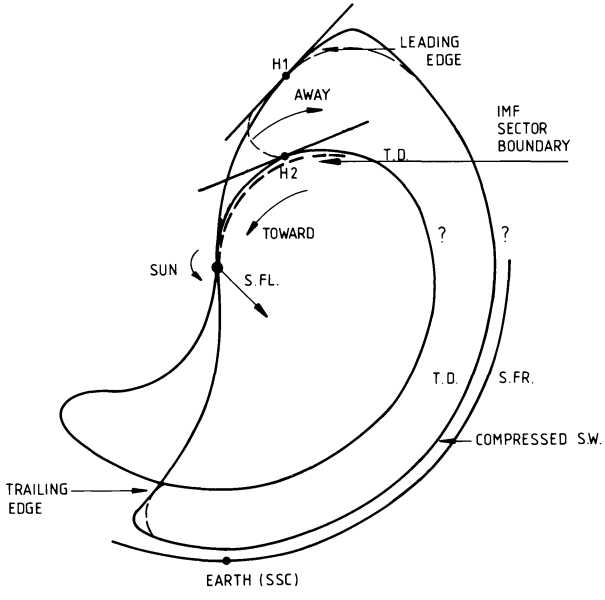


Fig. 6. A possible representation of the corotating structure, the region indicated by questionmarks is not covered by spacecraft that time. The two circumference lines represent the possible spatial structures on day 122 and day 123. The region between the shock front and the tangential discontinuity is the compressed solar wind. The *thin dashed line* indicates the possible large-scale closed magnetic structure

In order to calculate the orientation of the TD surface, we apply the method of Siscoe et al. (1969). Using the solar wind velocity data and assuming no mass flux across the discontinuity, we calculate the angle of the intersection between the surface of the TD with the ecliptic and the sun-HELIOS line (Fig. 5) for both H1 and H2 with the equation:

$$\tan \psi = \Delta V_T / \Delta V_R$$

ΔV_T, ΔV_R, are the variations of the transvers and radial solar wind velocities across the TD respectively. For H1 we obtain:

$$\psi_1 = (20 \pm 7)^\circ$$

and for H2

$$\psi_2 = (34 \pm 10)^\circ$$

According to Siscoe et al., 1968; Burlaga and Ness, 1969a; Burlaga, 1969b; the line formed by the intersection of a TD and the ecliptic plane lies along the spiral field direction.

Before the TD was detected by H2 the IMF changed direction from TOWARD to AWAY from the sun, (Fig. 6). This new sector lasted about 5 days. The time-delay of the sector boundary between the passages at H2 and H1 was about one and half days (Neubauer, 1977), and during this time the earth was in an IMF sector directed TOWARD the sun (SGD, 1976).

The Interplanetary Shock

While no shock indication was found by H1 and H2, IMP-8 observed on May 2 the passage of a shock front simultanously with an SSC, identified by ground based magnetometers, (Fig. 3), (Bame, 1977; SGD, 1976).

Assuming a radial direction of the shock normal, we calculate the shock velocity from the equation of mass conservation:

$$N_1(V_1-U)=N_2(V_2-U)$$

N, V, being the number density, the solar wind velocity, and U the shock front velocity. Indices 1 and 2 indicate the parameters immediately before and after the shock respectively. With:

$$N_1=8\ \mathrm{cm}^{-3}$$

$$N_2=20\ \mathrm{cm}^{-3}$$

$$V_1=600\ \mathrm{km\ s}^{-1}$$

$$V_2=680\ \mathrm{km\ s}^{-1}$$

we obtain for U:

$$U=730\ \mathrm{km\ s}^{-1}.$$

This figure compares well to the mean shock velocity $U_m=900\ \mathrm{km\ s}^{-1}$ (Fig. 2), as calculated from the time difference between the onset of the flare and shock arrival at the earth, since the shock must be expected to have slowed-down at 1 AU, compared to its propagation velocity closer to the sun. (Dryer, 1974).

The Forbush Decrease

Simultaneously, (within measurement accuracy) with the passage of the shock front the onset of a FD occured at 18:30 UT on May 2, about two days after the Ground Level Event.

The solar flare related 10 cm radio wave emission, which generally accompagnies a FD producing solar flares, was of type II and IV (Dodson and Hedeman, 1977).

Discussion

Trying to understand the described event, its following characteristics are to be taken into account:

1. the absence of any shock at HELIOS-1 and -2,
2. the relatively long time (29.5 h) spent by the tangential discontinuity to propagate from *H*2 to *H*1,
3. the occurence of a shock at the earth about the same time which the tangential discontinuity reached H1,
4. the abrupt decrease and the short recovery time of the high-energy cosmic ray particle fluxes, and
5. the orientation of the tangential discontinuity surface.

The abrupt increase of the solar wind velocity and the proton temperature but the decrease of the proton number density detected at H2 and H1 (123:14:00, 123:19:30, respectively, Fig. 4a and b) as well as the lack of any α-particle flux enhancement (due to the piston gas emitted by the solar flare) suggest that at least in the line-up region, shock- or flare material are not found. On the other hand, the abrupt increase of the solar wind velocity, the proton number density, the temperature and the α-particle flux (Armstrong et al., 1977), detected at IMP-8 (123:18:29), as well as the storm sudden commencement detected at ground level (SGD, 1976) support the existence of a shock front and a piston gas at the earth.

According to the spatial separation among IMP-8, HELIOS-1 and HELIOS-2, and the solar flare, it is suggested that the longitudinal extension of this perturbation region was at least 160° wide. Usually, such extension is about 130° wide, (Iucci et al., 1977).

If the shock with such a large width does not extend to the longitudes at HELIOS-1 and HELIOS-2, it should probably be observed close to the heliographic longitude of the flare, rather than far away. In the described event, the position of the earth relative to the flare was much nearer than of H1 and H2, and therefore, the shock has been detected at the earth but not at H1 and H2 region.

Assuming the time at the onset of the flare as zero, we give in Fig. 2 the mean propagation velocity of the tangential discontinuity between H1 and H2.

Surprisingly, the calculated angles ψ_1 and ψ_2 are in disagreement with the rule, according to which the Parker angle increases with increasing radial distance. This fact leads to the conclusion that between HELIOS-1 and HELIOS-2 the interplanetary magnetic field has an unusual shape. Ness (1966) reported a directional discontinuity observed also by two spacecraft, Pioneer-6 and IMP-3, and showed that the observed transit time from Pioneer-6 to IMP-3 was about the time required for an Archimedean spiral line corresponding to a solar wind velocity of 410 km s^{-1} to corotate from Pioneer-6 to IMP-3.

According to the feature of flare-produced interplanetary shock wave, proposed by Hundhausen (1972, p. 192), the region of the compressed ambient solar wind behind the shock is 0.1 to 0.2 AU thick. The following tangential discontinuity that is expected to separate the compressed ambient plasma from the flare ejecta is sometimes followed by a thin shell (0.01 to 0.1 AU thick)

of helium-rich material. The same feature we find also here (Fig. 6), where after the shock a helium-rich material is detected by IMP-8 (Armstrong et al., 1977). For the expected tangential discontinuity, which separates the compressed ambient plasma from the helium-rich shell, it is yet not clear if it was detected at IMP-8.

In order to estimate the thickness (D) of the compressed plasma, we consider the onset of the shock (123:17:05), the peak of the α-particle flux observed at IMP-8 (123:24), and the mean solar wind velocity, (700 km/s). Under the assumption that the peak, superimposed on the main increase, is really due to the helium-rich material (Armstrong et al., 1977) we obtain:

$$D \cong 0.1 \text{ AU}.$$

The abrupt decrease of the high-energy cosmic ray particles of about 5% only, and especially the short recovery time (Fig. 3), suggest that the earth was engulfed by the so called 'trailing' edge of a corotative-like modulation region (Iucci et al., 1977). A longer recovery period should be expected if the modulation region would expand only radially. The other edge, called 'leading' edge, considered as the boundary of the modulation, is considered also as corotative, bacause of the relatively long time spent by its passage from HELIOS-2 to HELIOS-1. After the shock, a very low proton temperature region (Gosling et al., 1973; Montgomery et al., 1974; Geranios, 1978) appeared, lasting for about 5 h, and coinciding with the observed Forbush Decrease (Fig. 3). Whether or not, this low temperature region suggests a closed magnetic loop, still remains an open question. For this answer, simultaneous electron temperature and interplanetary magnetic field data are necessary.

During the described event, there were unfortunately no other spacecraft between the lined-up HELIOS-1 and HELIOS-2, and the earth's heliographic longitude, which could complete the observations between 0° and 160°, (Venera-9 and -10 were fairly close to HELIOS-1 and HELIOS-2, Pioneers were in the eastern part of the ecliptic (Hilberg et al., 1977).

Conclusion

All detected enhanced interplanetary parameters obtained from data measured at HELIOS-1, HELIOS-2, IMP-8, and ground level, had a common origin: the 2B solar flare on April 30, 1976. Correlating the data measured at a separation in heliographic longitude at 160° we arrive to the following conclusions:

1. For unusually large modulation regions, the shock characteristics are not extended in the whole region, but only in heliographic longitudes close to the causitive flare. In this described event the shock characteristics were to be seen in a longitude more than 45° and less than 120° relative to the flare site.

2. The tangential discontinuity observed a HELIOS-1 and HELIOS-2, and the Forbush Decrease observed at the earth were of a corotative type.

3. During the corotation, the Archimedean spiral changed its angle in an unexpected way on the way between HELIOS-2 and HELIOS-1. This fact could

signify the detection of the upper part of a 'Wurst-like' magnetically closed region, (dashed line, Fig. 6).

4. If the modulation region is controlled by the tangential discontinuity, 'leading edge' and the Forbush Decrease, 'trailing edge' then, the whole region can be expected to corotate with the sun, (Fig. 6).

Acknowledgements. The author would like first to acknowledge Professor K. Pinkau and Dr. H. Rosenbauer for his stay at Max-Planck-Institut für Extraterrestrische Physik, and Drs. H. Rosenbauer and R. Schwenn for the support with all necessary plasma data from both HELIOS-1 and HELIOS-2 spacecraft.

The author expresses his gratitude to Professor F. Neubauer for the preliminary magnetic field data from HELIOS-1 and -2, to Dr. S. Bame for the plasma data from IMP-8 satellite, and to Professor G. Wibberentz for the permission to reproduce the upper part of Fig. 4a and b. For final remarks and corrections I am gratefull to Dr. H. Rosenbauer.

References

Agrawal, S., Ananth, A., Bemalkhedkar, M., Kargathra, L., Rao, U.: High-energy cosmic-ray intensity increase of nonsolar origin and the unusual forbush decrease of August 1972. J. Geophys. Res. **79**, 2269–2280, 1974

Armstrong, T., Decker, R., Krimigis, S., Kohl, J.: Solar and interplanetary particles observed in the interval March 20–May 5 with IMP-8. Collected data reports for STIP intervall II, March 20–May 5, 1976, UAG-61, WDC-A, 1977

Bame, S.: Private Communication, Los Alamos Scientific Laboratories, Los Alamos, New Mexico 87544, 1977

Burlaga, L.: Micro-scale structures in the interplanetary medium. Sol. Phys. **4**, 67–92, 1968

Burlaga, L., Ness, N.: Tangential discontinuities in the solar wind. Sol. Phys. **9**, 467–477, 1969a

Burlaga, L.: Directional discontinuities in the interplanetary magnetic field. Sol. Phys. **7**, 54–71, 1969b

Chirkov, N., Novikov, A., Filippow, A., Kuzmin, A., Sergeyer, A., Okara, A., Borisov, V., Sokolova, T.: Solar cosmic ray increase on April 30, 1976. Collected data reports for STIP interval II, March 20–May 5, 1976, UAG-61, WDC-A, 1977

Dodson, H., Hedeman, E.: Overall evaluation of major centers of activity on the solar disc, March 20–May 5, 1976. Collected data reports for STIP intervall II, March 20–May 5, 1976, UAG-61, WDC-A, 1977

Dryer, M.: Interplanetary shock waves generated by solar flares. Space Sci. Rev. **15**, 403–468, 1974

Duggal, S., Pomerantz, M.: Relativistic solar cosmic rays on April 30, 1976. Collected data reports for STIP interval II, March 20–May 5, 1976. UAG-61, WDC-A, 1977

Geranios, A.: A search of the origin of VLETs. Planet. Space Sci. **26**, 571–579, 1978

Gosling, J., Pizzo, V., Bame, S.: Anomalously low proton temperatures in the solar wind following interplanetary shock waves, evidence for magnetic bottles? J. Geophys. Res. **78**, 2001–2009, 1973

Hilberg, R., Teaque, M., Vette, J.: Positions of satellites March 20–May 5, 1976. Collected data reports for STIP interval II, March 20–May 5, 1976, UAG-61, WDC-A, 1977

Hundhausen, A.: Coronal expansion and solar wind. Berlin, Heidelberg, New York: Springer 1972

Iucci, N., Parisi, M., Storini, M., Villoresi, G.: Type IV solar radio burst and forbush decreases. Proc. Int. Conf. Cosmic Rays, Plovdiv **3**, 329, 1977

Kunow, H., Müller-Mellin, R., Iwers, B., Witte., M., Hempe, H., Wibberenz, G., Green, G., Fuckner, J.: MeV protons, alpha particles, and electrons as observed aboard HELIOS 1 and 2 during STIP Interval II. Collected Data Reports for STIP Interval II, UAG-61, WDC-A, 1977

Mihalov, J., Golburn, D., Collard, H., Smith, B., Sonett, C., Wolfe, I.: Pioneer solar plasma and magnetic field measurements in interplanetary space during August 2–17, 1972. Correlated

Interplanetary and Magnetosphere Observation. Saulgau, W. Germany: Proc. ESLAB Symp. 1973

Montgomery, M., Asbridge, J., Bame, S., Feldman, W.: Solar wind electrons temperature depression following some interplanetary shock waves: Evidence for magnetic merging? J. Geophys. Res. **79**, 3103–3110, 1974

Ness, N.: Simultaneous measurements of the interplanetary magnetic field. J. Geophys. Res. **71**, 3319–3324, 1966

Neubauer, F.: Personal Communication. Braunschweig: Technische Universität 1977

Pomerantz, M., Duggal, S.: Interplanetary acceleration of solar cosmic rays to relativistic energy. J. Geophys. Res. **79**, 913–919, 1974

Schwenn, R., Rosenbauer, H., Mühlhäuser, K.: The solar wind during STIP II interval: Stream structures, boundaries, shocks, and other features as observed by the plasma instruments on HELIOS-1 and HELIOS-2. Tel Aviv: STIP II. 1977

Schwenn, R.: Private Communication. Garching F.R.G.: Max-Planck-Institut für Extraterrestrische Physik 1977a

Shea, M.: Overview of solar-terrestrial physics phenomena for the retrospective world interval of March 20–May 5, 1976. Collected data reports for STIP Interval II, March 20–May 5, 1976, UAG-61, WDC-A, 1977

SGD.: Solar geophysical data, 380, 381. Part I: April, May, June, July 1976. Boulder, Colorado, USA: U.S. Department of Comerce 1976 p. 80302

Siscoe, G., Davis, L., Coleman, P., Smith, Jones, D.: Power spectra and discontinuities in the interplanetary magnetic field: Mariner 4. J. Geophys. Res. **73**, 61–82, 1968

Siscoe, G., Turner, J., Lazarus, A.: Simultaneous plasma and magnetic-field measurements of tangential discontinuity in the solar wind. Sol. Phys. **6**, 456–464, 1969

Tanskanen, P., Kananen, H., Blomster, K.: Observations relevant to the August 1972 Storm. Collected data reports on August 1972. Solar-terrestrial events, UAG-28, WDC-A, 1973

Received June 1, 1979; Accepted July 4, 1979

J. Geophys. 46, 185–199, 1979

Journal of Geophysics

Magnetic Rotational Hysteresis Loss in Titanomagnetites and Titanomaghemites – Application to Non-Destructive Mineral Identification in Basalts

A.J. Manson, J.B. O'Donovan, and W. O'Reilly

Department of Geophysics and Planetary Physics, School of Physics, The University, Newcastle upon Tyne NE1 7RU, Great Britain

Abstract. Magnetic rotational hysteresis loss provides a non-destructive method for the identification of magnetic minerals in rocks provided a catalogue of loss curves is available for well characterized minerals or synthetic analogues. In the present study such loss curves have been determined for synthetic titanomagnetites and titanomaghemites. The effect of the presence of aluminium and magnesium such as is found in naturally occurring minerals has also been determined. One feature of the loss curve, the field at which peak loss occurs (H_p), appears to be independent of the domain state of the material and depends only on chemical composition. The value of H_p may therefore be used to identify minerals in the same way as a value of Curie temperature (T_c), although neither H_p nor T_c alone will give an unambiguous answer. The loss curves of a number of basalts have also been determined and an attempt is made to determine the oxidation state of the titanomagnetites using the data from the synthetic materials.

Key words: Rock magnetism – Magnetic hysteresis – Titanomagnetite – Titanomaghemite.

1. Introduction

Studies of the magnetism of rocks are primarily aimed at providing the physical basis for the assumptions on which the palaeomagnetic method is founded. One basic problem is the identification of the magnetic minerals in rock samples, such minerals often being present in low concentration ($\sim 1\%$) and of submicroscopic grain size ($\lesssim 1$ μm). The identification and characterization of the magnetic fraction of a rock sample is essential to the interpretation of the magnetic properties of rocks, in assessing the usefulness of rocks as carriers of the palaeomagnetic record and can contribute to our knowledge of the petrological history of rocks.

0340-062X/79/0046/0185/$03.00

The magnetic properties of the minerals themselves may provide characteristic 'signatures' from which their presence may be inferred. This requires a catalogue of signatures obtained using well-characterized mineral samples or synthetic analogues. A widely-used measurement in this context is the determination of the temperature dependence of induced or remanent magnetization. The former yields the Curie temperature intrinsic to the magnetic phase. The latter may be used to identify magnetite and haematite by their 'remanence transitions' below room temperature. This method has the advantage of avoiding possible chemical change due to heating at elevated temperatures arising through the oxidizing or reducing conditions existing in the heated rock or inherent instability in the magnetic minerals (providing irreversible changes in the thermomagnetic curve occurring on heating are well understood, such changes can themselves be used to identify minerals). However, the remanence transition occurs only in minerals with a limited range of composition and the transitions themselves are strong functions of extrinsic parameters such as grain size and mechanical strain.

Even under ideal conditions, the results obtained using any single technique do not unambiguously identify the magnetic mineral. A measured Curie temperature, for example, simply locates the mineral on the appropriate contour of the titanomagnetite/titanomaghemite quadrilateral (Readman and O'Reilly, 1972) or on a contour in a titanomagnetite/titanomaghemite subsystem (e.g., O'Donovan and O'Reilly, 1977), or again may indicate a haemo-ilmenite or even a pyrrhotite or goethite if in the appropriate range. A combination of techniques is needed to reduce the field of composition, within which the mineral may lie, to a usefully narrow range.

Another magnetic property which may provide a signature is magnetic anisotropy and anisotropy-related characteristics such as coercive force, remanence coercive force, hysteresis loss, susceptibility and the shapes of magnetization and demagnetization curves. Of special interest among these is the determination of the field dependence of rotational hysteresis loss (Day et al., 1970). This loss is due to the irreversible magnetization changes which occur while rotating a specimen in a fixed magnetic field. There are a number of advantageous features of this method for identifying minerals. First the induced magnetism of the non-ferromagnetic minerals plays no role and cannot mask the characteristic signature of the ferromagnetic phases. Second, the effect depends on the anisotropy rather than the strength of magnetization of the material. Thus although haematite (α Fe_2O_3) is weakly magnetic ($\sim 1/200$ the saturation magnetization of magnetite, Fe_3O_4) the rotational hysteresis loss may be quite large especially if in the finely divided pigment form (of about the same order as that of coarse grain magnetite). Similarly whereas fine grain (pigment) and coarse grain (specularite) haematite have the same Curie temperature, their rotational hysteresis characteristics are different and the two forms of haematite may in principle be distinguished (Cowan and O'Reilly, 1972). Finally the technique is chemically non-destructive. The disadvantage of the technique is that the determination of a rotational hysteresis curve may take many hours and a highly automated instrument is desirable.

Rotational hysteresis loss is measured using a torque magnetometer. A single

crystal specimen suspended in a saturating magnetic field experiences a torque unless the field is parallel to an easy axis of the system, the torque being uniquely determined by the angle between the field and a reference direction. In lower fields the magnetization will lag or lead the field and the resultant torque is no longer unique but, for a certain orientation, will depend on the sense of rotation by which that orientation has been reached. Irreversible changes occur in the orientation of the magnetization during rotation of the field. The curves of the angular dependence of torque for rotation of the field in the clockwise and anticlockwise sense are displaced from one another. The area between these curves has the dimensions of energy and for a rotation $0°\rightarrow 360°\rightarrow 0°$ the area equals twice the hysteresis loss per cycle. A rock specimen contains a dispersion of magnetic grains. In the absence of any 'fabric' the specimen has no bulk easy axis. However even with a random array of crystallite easy axes, the bulk magnetization will lag a non-saturating rotating field, a torque will result and the area between the displaced torque curves gives the loss per cycle of the assemblage.

Theoretical loss curves have been derived for the case of monodomain particles with uniaxial anisotropy, the unique axis being the easy direction (Stoner and Wohlfarth, 1948; Jacobs and Luborsky, 1957). In the model curves, the loss is zero for fields below a critical field ($=K_u/M_s$ for coherent rotation of magnetization where K_u is the anisotropy constant and M_s the spontaneous magnetization). At the critical field the loss rises abruptly to a peak value ($=2.6\ K_u$ for an aligned assemblage) and falls steadily to zero at twice the critical field. Owens (1977) has considered the case of uniaxial symmetry in which the unique axis is the hard direction and the unique plane contains a single easy direction. In this case, the field at which the loss will vanish depends on the orientation of the unique plane to the plane of measurement, and can extend to very high fields. We have considered the case of monodomain grains with cubic anisotropy ($K_1 > 0$) and coherent rotation in the (100) plane. Figure 1 shows the resulting variation of rotational hysteresis loss with applied field. The general shape of the curve is similar to the uniaxial case but the critical field for the onset of loss is $^1/_2(K_1/M_s)$ and the peak loss is 0.908 K_1 for an aligned assemblage (0.606 K_1 for a random array). The loss vanishes at $2K_1/M_s$. Real materials are invariably more complicated than those described in simple idealized models. Nevertheless experimental loss curves (W_R–H curves) are in a general way similar to the model curves. In any event, the important point in the present context, that of mineral identification, is that an observed W_R–H curve can be recognized as belonging to a certain mineral or group of minerals. W_R–H curves are therefore needed for synthetic analogues or well-characterized mineral specimens of the titanomagnetites and their derivatives, produced by oxidation, 'substitution' and post deuteric alteration. These will include magnetite, maghemite, haematite, the titanomaghemites, and haemoilmenites. Other remanence-carrying minerals are the pyrrhotites, goethite and native iron.

The purpose of the present study is to provide some of the required specimen W_R–H curves. The systems studied are fine grain (monodomain) and coarse grain (multidomain) titanomagnetite and the oxidation products of such mate-

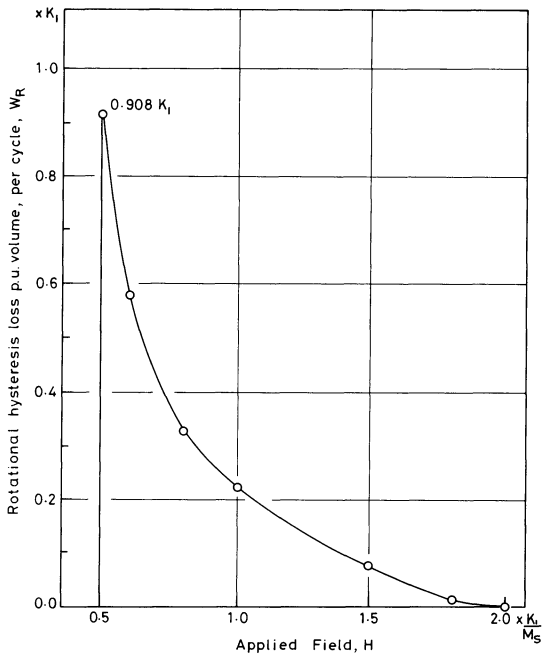


Fig. 1. Model rotational hysteresis loss (W_R–H) curve for aligned monodomain grains with cubic anisotropy ($K_1 > O$) and coherent rotation of magnetization in the (100) plane. W_R is expressed in units of K_1 and H in units of the ratio K_1/M_s. The curve for a random array of easy axes is similar but with a peak loss of 0.606 K_1

rials. The effect of 'substitution' of iron by aluminium and magnesium is also investigated together with the consequent changes in W_R–H curves due to maghemitization in the latter case.

2. Experimental

The rotational hysteresis loss curves were determined using the automatic torque magnetometer of Fletcher et al. (1969) operating at room temperature and in field range 0–20 k Oe. Torque curves for angle 0→360→0° were plotted on an X–Y recorder at each selected field, the areas enclosed by the curves measured, converted to energy per unit mass, and plotted against field to give the W_R–H curve.

The titanomagnetite specimens were prepared by sintering techniques, the unsubstituted titanomagnetites fired at 1,100° C under self buffering conditions and the substituted materials twice-fired using a technique of partial self buffering at 1,350° C. The homogeneity of the specimens was confirmed using X-ray powder pictures and Curie temperature determination. The synthesized material was hand-ground and sieved fractions in the required particle size range prepared.

Fine grain specimens of unsubstituted titanomagnetite were obtained by grinding in a wet slurry in a porcelain ball mill for four days. The finely divided substituted titanomagnetite was obtained by grinding, in high purity methanol, overnight in a tungsten carbide ball mill. Particle size determinations were attempted using X-ray line broadening and the electron microscope. Be-

cause of strain-broadening the former method provides an underestimate of the particle size. The latter method yields an overestimate because of flocculation. The combined measurements gave a particle size of the order of 0.1 μm.

Oxidation was carried out by heating in air in a thermogravimetric balance, the degree of oxidation being determined from a combination of weight increase data and wet-chemical analysis. Powder pictures of the oxidation products were obtained using Guinier de Wolff and Debye-Scherrer cameras.

3. The Rotational Hysteresis Characteristics

The rotational hysteresis curves of all samples were for the main part of very similar form. A representative suite of curves is shown in Fig. 2. All curves show a sharp rise in W_R to a peak value, followed by a gradual fall, as the applied field increases. The shape of the curves is thus similar to the model curves of Fig. 1 differing in that the measured W_R is not zero at low fields nor does it fall to zero when the applied field is (say) four times the field at which peak loss occurs. Given that the observed measured curves have a similar form to those in Fig. 2 it is not necessary to reproduce the curves themselves but simply to characterize each curve by the value of field, H_p, at which the peak occurs and the height of the peak itself, W_{Rp}.

3.1 Titanomagnetites

The variation of H_p and W_{Rp} for the titanomagnetites ($Fe_{3-x}Ti_xO_4$, $0<x<1$) as a function of composition x is shown in Fig. 3. The data refer to material

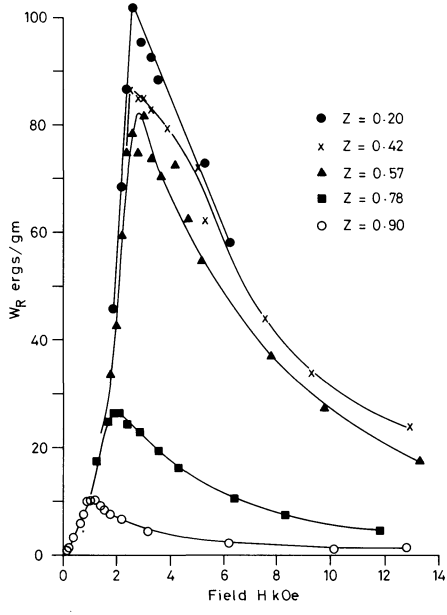


Fig. 2. Representative set of W_R–H curves for a suite of titanomaghemites (based on $Fe_{2.25}Mg_{0.15}Ti_{0.6}O_4$) showing the evolution of the loss curve as the degree of maghemitization (z) increases. The field (H_p) at which peak loss occurs falls, as does the peak height (W_{RP}), as z increases

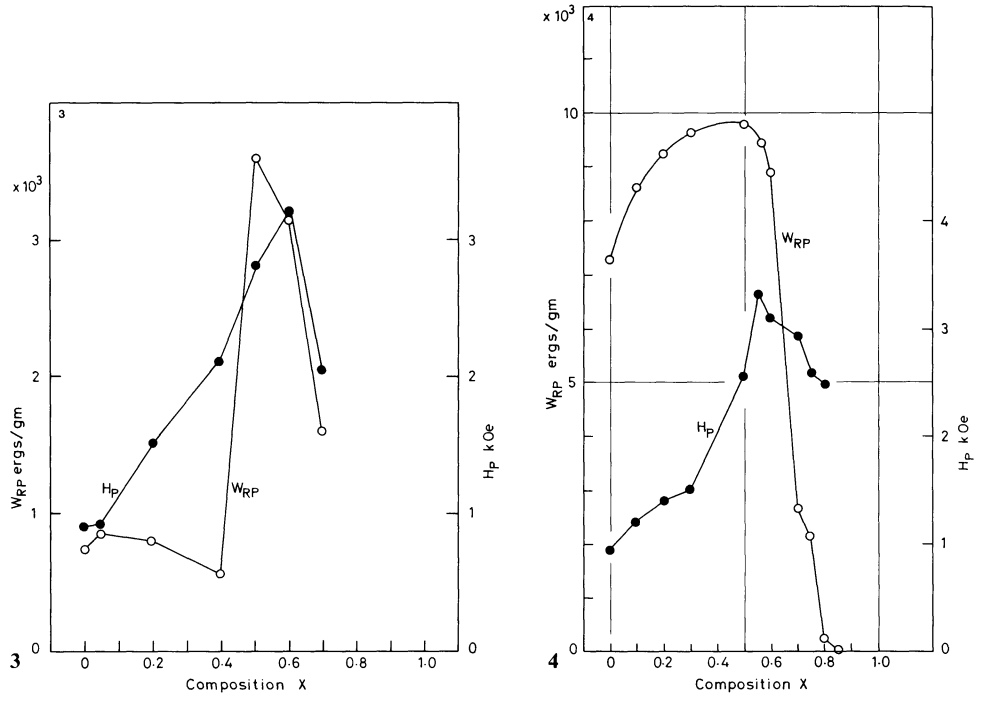


Fig. 3. Variation of H_p and W_{RP} with *composition* x in the titanomagnetite series $Fe_{3-x}Ti_xO_4$. The data refer to near-stoichiometric samples in the multidomain state, sieved to obtain a grain size range 37–44 μm

Fig. 4. Variation of H_p and W_{RP} with *composition* x for ball milled samples of titanomagnetite in the monodomain state. A low degree of non-stoichiometry is introduced by the ball milling. The H_p variation is the same as that in the multidomain samples (Fig. 3) whereas W_{RP} is an order of magnitude larger

which has been sieved to lie in the grain size range 37–44 μm and is expected to comprise mainly multidomain grains. However the possibility of smaller particles, in agglomerates or attached to larger particles cannot be discounted. H_p rises steadily from about 1 *k* Oe for magnetite ($x=0$) to 3 *k* Oe for $x=0.6$. The fall in H_p from $x=0.6$ to $x=0.7$ is associated with the fall in Curie temperature to near room temperature for $x=0.7$ and below room temperature for $x \gtrsim 0.8$. The behaviour of W_{Rp} is less systematic, falling from about 700 to 500 ergs/gm as x increases to 0.4 to a peak value of about 3,500 ergs/gm at $x=0.5$ before falling to 1,500 ergs/gm at $x=0.7$.

Figure 4 gives the corresponding data for fine grain ball milled specimens of grain size $\lesssim 0.1$ μm and expected to be in the monodomain state. The values and variation of H_p with composition x are similar to the coarse grain material. However the W_{Rp} values now show a more continuous and systematic variation with x and are about an order of magnitude greater than in the multidomain material, rising from about 7,000 ergs/gm ($x=0$) to a peak at about 10,000 ergs/gm ($x=0.5$). The ball milled samples are more non-stoichiometric (cation-defi-

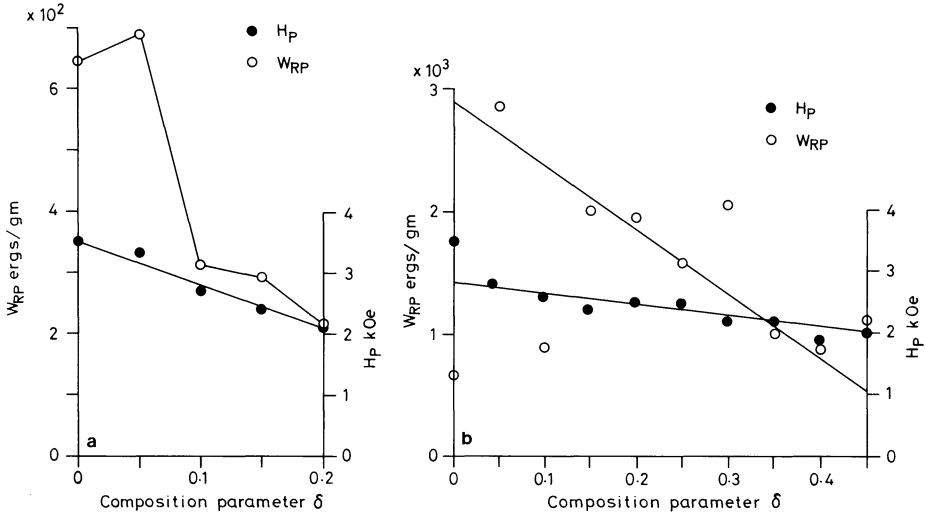


Fig. 5. **a** Variation of H_p and W_{RP} with composition in the titanomagnetite $Fe_{2.4-\delta}Al_\delta Ti_{0.6}O_4$ for multidomain specimens (37–44 μm). W_{RP} varies erratically with composition whereas the variation of H_p is systematic. **b** Variation of H_p and W_{RP} with *composition* δ in the titanomagnetite $Fe_{2.4-\delta}Mg_\delta Ti_{0.6}O_4$ for multidomain specimens (37–44 μm). W_{RP} varies erratically with composition whereas the variation of H_p is systematic

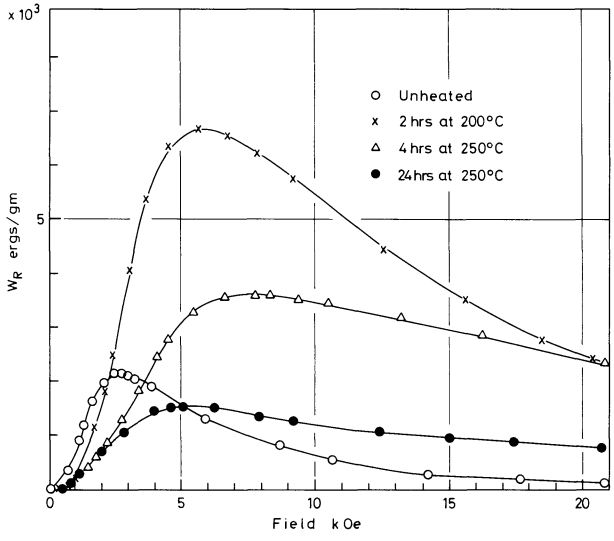


Fig. 6. Evolution of the W_R–H loss curve as a function of heat treatment in air (and increasing degree of maghemitization) of fine grain $Fe_{2.25}Ti_{0.75}O_4$ ($x=0.75$). The starting material has a Curie temperature T_c near to room temperature and, as maghemitization proceeds, the Curie temperature rises. In effect each curve, obtained at room temperature, corresponds to a different 'magnetic isotherm', T/T_c. This behaviour is to be contrasted with that of Fig. 2

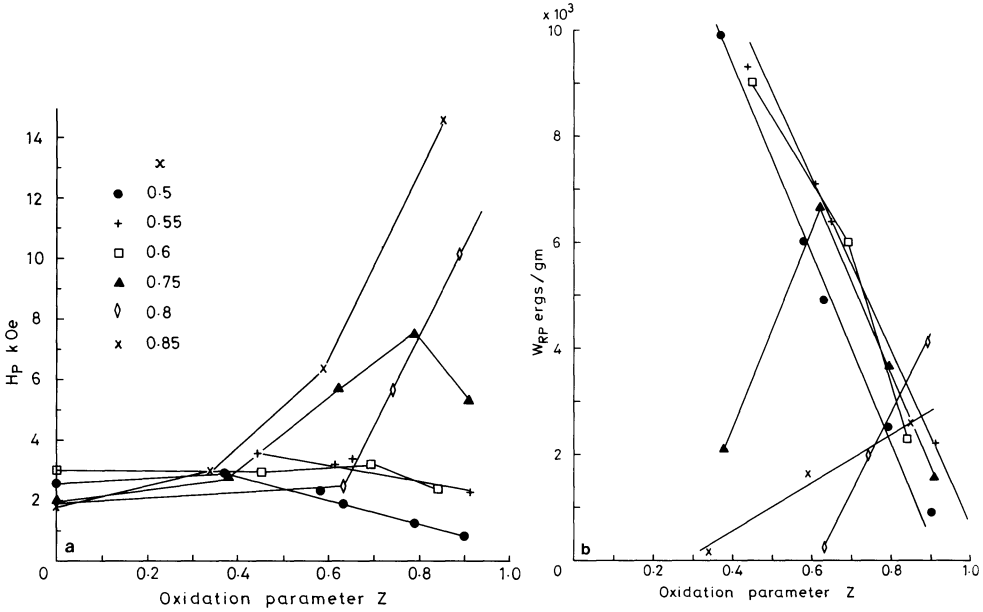


Fig. 7. **a** Variation of H_p with degree of oxidation, z, for maghemitized titanomagnetites of composition $x=0.5$, 0.55, 0.6, 0.75, 0.8, and 0.85. **b** Variation of W_{RP} with degree of oxidation, z, for maghemitized titanomagnetites of composition $x=0.5$, 0.55, 0.6, 0.75, 0.8, and 0.85 (see legend of **a**)

cient) than the large grain samples and therefore, having higher Curie temperatures, H_p and W_{Rp} persist at room temperature to a higher x value (~ 0.8).

The effect of the presence of cations other than Fe and Ti, such as are found in naturally occurring titanomagnetite has also been studied. The composition $Fe_{2.4}Ti_{0.6}O_4 (x=0.6)$ was chosen as being typical for many basalts and the Fe progressively substituted by up to 0.2 Al ions and 0.45 Mg ions per formula unit. The variation of H_p and W_{Rp} as a function of increasing concentration (δ) of substituting ion is shown in Fig. 5 for coarse grain material (37–44 μm). Al, replacing Fe^{3+}, is more effective in reducing H_p than Mg (replacing Fe^{2+}). Again the W_{Rp} values seem not to depend in a systematic way on composition for these multidomain specimens.

3.2 Titanomaghemites

Because of their relevance to naturally occurring titanomagnetites/titanomaghemites we have concentrated on compositions near to or based on $x=0.6$. The evolution (with respect to degree of maghemitization) of the W_R–H loss curve for a suite of titanomaghemites in which the Curie temperature is well above room temperature is shown in Fig. 2. The behaviour is more complicated when the starting composition is near (or below) room temperature as shown in Fig. 6. This may be described by saying that, at room temperature, the curves

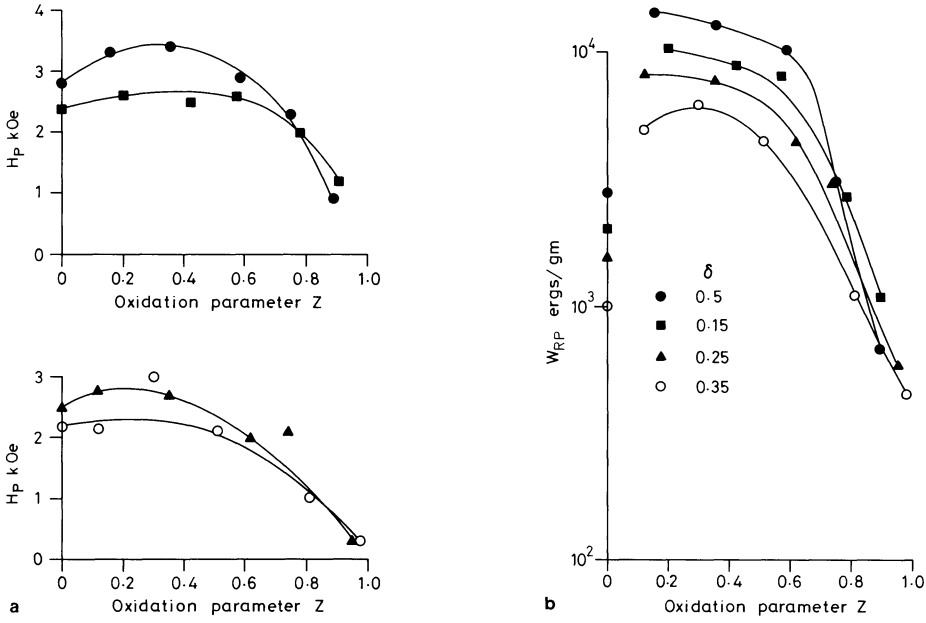


Fig. 8. **a** Variation of H_p with degree of oxidation, z, for maghemitized titanomagnetite based on $x=0.6$ and containing 0.5, 0.15, 0.25, and 0.35 ($=\delta$) Mg ions per formula unit (for identification of compositions see legend of **b**). **b** Variation of W_{RP} with degree of oxidation, z, for maghemitized titanomagnetite based on $x=0.6$ and containing 0.5, 0.15, 0.25, and 0.35 ($=\delta$) Mg ions per formula unit. The data points for $z=0$ refer to unground material in the multidomain strate. All other data points refer to fine grain monodomain material

do not correspond to the same 'magnetic isotherm', T/T_c. Figure 7 shows the variation of H_p and W_{Rp} with degree of maghemitization, z ($0 \leqq z \leqq 1$, z is the fraction of the Fe^{2+} concentration corresponding to stoichiometry, for a given x, oxidized to Fe^{3+}) for $x=0.5$, 0.55, 0.6, 0.7, 0.75, 0.8, and 0.85. All the titanomaghemite samples are in the fine grain monodomain state.

The combined effect of magnesium substitution and maghemitization has also been studied, again for ball milled monodomain specimens. The variation of H_p and W_{Rp} with z for a subsystem of $x=0.6$ containing up to 0.35 Mg ions per formula unit is shown in Fig. 8.

In all cases, except those complicated by the proximity of Curie temperatures to the measuring temperature (i.e., high x values), maghemitization results in a 'softening' of the material with reduced H_p values and W_R levels.

4. Discussion

4.1 Mineral Identification

The observed values of W_{Rp} indicate a dependence on some variable parameters such as microstructure rather than chemical composition, especially for multido-

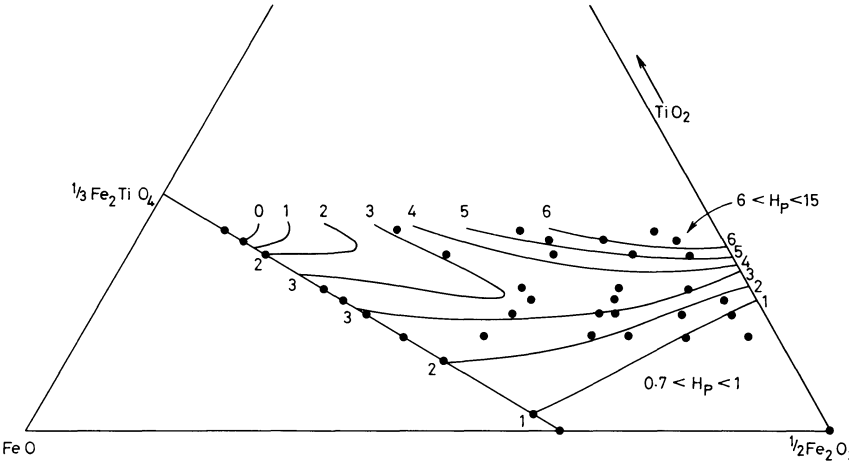


Fig. 9. Contours of constant value of H_p (in k Oe) for titanomagnetites/titanomaghemites in the $FeO-Fe_2O_3-TiO_2$ ternary diagram, constructed from the data of Fig. 4 and 7a. Solid circles indicate the compositions of the measured samples

main material. In any event W_{Rp} is an extensive property and would only be useful in mineral identification in rocks if the concentration of magnetic mineral were known. On the other hand, H_p is intensive and has a systematic variation with composition. It appears to be independent of grain size and other extrinsic influences and may be a useful signatory property.

As is traditional, the composition dependence of the characteristic property has been cast in the form of contours on compositional ternary diagrams. Figure 9 shows the H_p contours for the $Fe_3O_4-Fe_2TiO_4-Fe_2TiO_5-Fe_2O_3$ quadrilateral of the $FeO-TiO_2-Fe_2O_3$ ternary diagram. The contours are based on the data of Figs. 4 and 7 and, being derived from a finite number of data points (34), are of necessity a subjective interpretation.

The H_p data of Figs. 5 and 8 have similarly been recast into contours on the relevant quadrilateral (representing Mg substituted and maghemitized $Fe_{2.4}Ti_{0.6}O_4$) in the $FeO-TiO_2-Fe_2O_3-MgO$ tetrahedron (Fig. 10).

The use of ternary diagrams such as those of Figs. 9 and 10 to display the chemical composition of naturally occurring magnetic minerals is not strictly possible unless the concentration of cations other than Fe and Ti is very small, in which case the standard ternary diagram of Fig. 9 may be used. Alternatively if the concentrations or the cations other than Fe and Ti are small except for one species, e.g., Mg, then a specially constructed ternary diagram like the one of which Fig. 10 is a part might be employed. A suite of compositions can be displayed on such diagrams subject to certain constraints on their compositions, e.g., in Fig. 10 all compositions have a constant Ti: (Fe, Mg) atomic ratio. Visual display of compositions is therefore possible only under special circumstances which are unlikely to apply to naturally occurring minerals. It is only possible to describe a composition in terms of its coordinates in multi-dimensional composition space. The coordinates are the molecular fractions of the oxide of each species, preferably referred to one cation. The composition

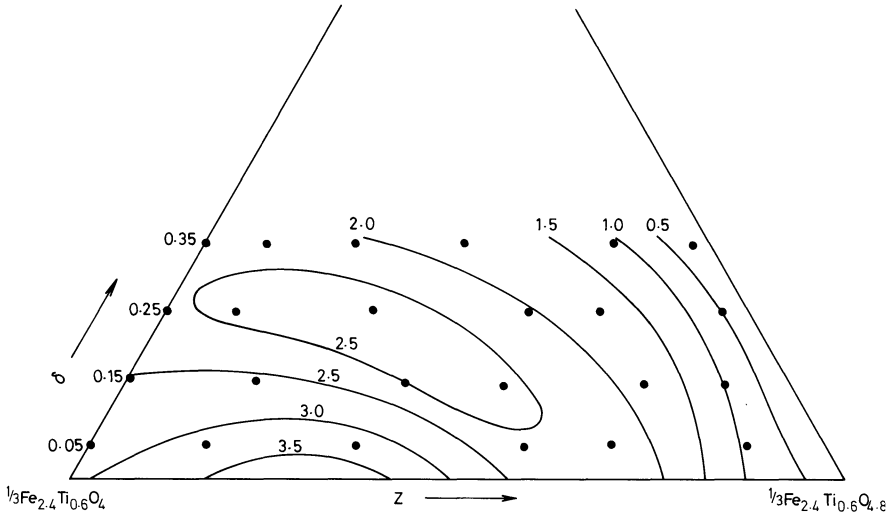


Fig. 10. Contours of constant value of H_p (in k Oe) for a magnesium substituted titanomaghemite subsystem based on $Fe_{2.4}Ti_{0.6}O_4$ and containing δ Mg ions per formula unit. The *solid circles* refer to the compositions of the measured samples (see also Figs. 5 and 8)

parameters x and z may still be retained but must be clearly defined (see O'Reilly, 1976). The determination of x and z for titanomagnetite in basalts has been attempted by a number of methods. Chemical analysis for each cation species may be carried out if a refined extract of the magnetic mineral can be made from the rock. Chemical analysis of large enough grains can be carried out in situ using the electron probe microanalyser (EPMA). However this technique cannot determine Fe^{2+} and Fe^{3+} separately, nor oxygen accurately enough to provide a value of z of useful precision. EPMA is therefore combined with the analysis of Fe^{2+} and Fe^{3+} of extracts using conventional chemical analysis, or the Mössbauer Effect, or combined with the determination of some composition dependent characteristic such as Curie temperature. A combination of EPMA and H_p determination could therefore be used.

To this purpose a number of basalt samples were selected which exhibited a single low Curie temperature and were therefore believed to contain only homogeneous unaltered titanomagnetite as the magnetic phase. The six basalts, the titanomagnetite in which had been subjected to EPMA analysis (Creer and Ibbetson, 1970), are listed in Table 1. An attempt had been made to determine the degree of maghemitization of these basalts by comparing their measured Curie temperatures to those expected for stoichiometric titanomagnetites of the indicated composition (Richards et al., 1973). The effect of high temperature oxidation ($>400°$ C) on the W_R–H curves of these basalts had also been studied (Manson and O'Reilly, 1976).

Column 1 in Table 1 gives the sample number and place of origin. The next four columns give the results of the EPMA analysis. The cation concentrations are expressed as the number of cations per four oxygen ions. Small concen-

Table 1. Rotational hysteresis data H_p(obs) (k Oe) for a number of basalts which exhibit a single, low, Curie temperature and are believed to contain homogeneous titanomagnetite. The microprobe chemical analyses of titanomagnetite grains are taken from Creer and Ibbetson (1970) and the oxidation parameter $z(T_c)$, derived from a consideration of the chemical composition and observed Curie temperature, taken from Richards et al. (1973). $H_p(x, \delta)$ are the values of field at peak rotational hysteresis loss expected from the chemical composition assuming a stoichiometric substituted titanomagnetite. The oxidation parameter based on the rotational hysteresis data, $z(H_p)$ is given in the final column (see text)

Sample	Microprobe Analysis							
	Fe	Mg	Al	Ti	$z(T_c)$	H_p(obs)	$H_p(x, \delta)$	$z(H_p)$
RK (Germany)	2.02	0.18	0.29	0.46	0.8	2.4	0.55	?
376 (Argentina)	1.98	0.24	0.15	0.60	0.75	1.6	1.6	low
293 (Argentina)	2.01	0.21	0.14	0.62	0.7	1.9	1.7	low
H21A (Aden)	2.05	0.13	0.15	0.64	0.6	3.5	1.2	~1 (?)
G143 (Turkey)	2.15	0.05	0.05	0.72	0.9	2.6	1.75	~0.4
D127 (Turkey)	1.99	0.10	0.04	0.85	0.75	2.1	0	0.2

trations of Mn and Cr (up to 0.03 ions per formula unit) were also detected but not listed here. One point of interest is that although the Fe concentration is more or less constant the other species vary widely and in an apparently systematic way. As the Ti concentration increases, the concentrations of Mg and Al decrease. Although the degree of oxidation could in principle be determined from the cation sum, the uncertainty in such a determination would be too large to be useful, and in the present analysis the original data have been normalized to give stoichiometric concentrations (Creer and Ibbetson, 1970).

The x value is simply the number of tetravalent cations per formula unit in a stoichiometric spinel and is therefore found in column 5 of Table 1. The z value of column 6, $z(T_c)$, was derived (Richards et al., 1973) using the Curie point data of synthetic materials. The Curie point corresponding to the stoichiometric composition of the indicated x value was reduced according to the concentrations of Al and Mg and the Curie temperature so derived compared to that observed. The observed Curie temperature was the higher in all cases and, knowing that maghemitization would increase the Curie temperature, this mechanism may be invoked to explain the difference. The degree of maghemitization is then estimated from the measured Curie temperatures of maghemitized synthetics. The same process has been applied in the present study using the H_p values. Column 7 of Table 1 gives the measured H_p values, H_p(obs), for the six basalts. The data for substituted synthetic titanomagnetites obtained in the present study is then used to obtain the H_p expected for a stoichiometric substituted titanomagnetite, $H_p(x, \delta)$. The difference between H_p(obs) and $H_p(x, \delta)$ may then be explicable in terms of maghemitization. Good agreement between H_p(obs) and $H_p(x, \delta)$ is found for two of the samples, 376 and 293 which may indicate near stoichiometric titanomagnetite. The value of oxidation parameter deduced from the rotational hysteresis measurements, $z(H_p)$, in column 9 is therefore considered to be low. For the other samples H_p(obs) is significantly

higher than $H_p(x, \delta)$. Inspection of the data of Figs. 7 to 10 shows that maghemitization does increase H_p values for high x (>0.6). Making the approximation that a given degree of maghemitization will produce the same elevation in H_p in both substituted and unsubstituted compositions a value of $z(H_p)$ may be derived. Thus samples G143 and D127 may be assigned $z(H_p)$ values of 0.4 and 0.2 respectively. H_p(obs) of H21A is not so easily explained as the H_p seems to be constant, independent of z for $x=0.6$ (Fig. 7). However the contours for high z (Fig. 9) are closely spaced for $x \gtrsim 0.6$ and it may be that for $z \sim 1$, H_p will increase for $x=0.6$. $z(H_p)$ for H21A may therefore be tentatively taken as being ~ 1. H_p(obs) for the remaining sample, RK, seems inexplicable. The large concentration of Al, together with Mg, depresses $H_p(x, \delta)$ to 0.55 k Oe whereas H_p(obs) is 2.4 k Oe. The effect of maghemitization on low x values (see $x=0.5$, Fig. 7) is to further depress the value of H_p. It might be concluded that the titanomagnetite in RK is near-stoichiometric but has an anomalously low H_p. RK is a much-studied basalt (see Petersen, 1962; Soffel, 1969) which is considered to contain unoxidized titanomagnetite and to be typical in many ways of continental basalts. However the EPMA analysis indicates a very large concentration of cations other than Fe and Ti. In fact with a total of 0.47 ions of Mg and Al per formula unit it might be arguable whether it qualifies as a 'titanomagnetite'. This high concentration of diamagnetic ions also presents problems in interpreting the observed Curie temperature. The 'compositional' Curie temperature for RK should be $\sim 90°$ C (Richards et al., 1973) whereas the observed value is 250° C. The high degree of maghemitization needed to bring the 'compositional' and observed Curie temperatures into line is not indicated by, for example, thermomagnetic curves. The role of the 'minor cation species' in naturally occurring titanomagnetites remains unclear.

4.2 Magnetic Anisotropy

It has been noted that H_p varies in a systematic way with composition, regardless of the domain state of the specimens. It is seen, for example, that H_p of a titanomagnetite of a given composition is about the same for coarse and fine grain specimens. However, the W_{Rp} values for the fine particles are about an order of magnitude bigger than for coarse grains of the same composition (Sect. 3.1). It appears that whereas H_p is a function of the intrinsic properties of a material W_{Rp} depends on extrinsic properties or microstructure. Similarly the coercive force of fine-particle, ball-milled titanomagnetites is (at about 2 k Oe for $x=0.6$) an order of magnitude bigger than that of coarse grain material.

According to the monodomain theory (Sect. 1) the peak rotational hysteresis loss, W_{Rp}, is proportional to the anisotropy energy per unit volume, as is the product of H_p and M_s, the spontaneous magnetization. The ratio $W_{Rp}:(H_pM_s)$ should then be constant for a given symmetry and type of anisotropy regardless of chemical composition. The magnesium substituted fine grain titanomaghemite specimens (Sect. 3.2) are in the monodomain state and range in W_{Rp} from about 500–15,000 ergs/gm, a factor of about 30 and in H_p by a factor of about 10 (0.3–3 k Oe). The ratios $W_{Rp}:(H_pM_s)$ for this system of

18 compositions lie between 0.1 and 0.2, relatively constant compared to the range in W_{Rp}. As W_{Rp} seems to depend on microstructure or extrinsic properties in the same way as coercive force, H_c, it seems likely that the ratio $W_{Rp}:(H_cM_s)$ may have a value nearer to a constant. This is indeed found to be the case for the suite of magnesium substituted titanomaghemites. Taking the H_c data for this system (O'Donovan and O'Reilly, 1977), this ratio has a value of 0.30 varying by about $\pm 10\%$ except two specimens out of the 18 which have values of about 0.20. The ratio does therefore appear to be a constant for the system. The actual values however are rather lower than predicted by the monodomain, coherent rotation, models which give 1.21 (cubic anisotropy) and 1.8 (uniaxial) for a random array of easy axes.

No model expressions for W_{Rp} and H_p are available for multidomain materials, however it is of interest to form the $W_{Rp}:(H_{p,c}M_s)$ ratios for the suite of 15 multidomain specimens of Al and Mg substituted $Fe_{2.4}Ti_{0.6}O_4$ (Sect. 3.1). The values of the two ratios are quite different. The $W_{Rp}:(H_pM_s)$ values for the Al substituted system lie between 0.005 and 0.007 and between 0.025 and 0.035 (with one value at 0.013 and one at 0.043) for the Mg substituted set. Using the H_c instead of the H_p values gives ratios over a closer range for both sets (0.5–1.7) having the same order of magnitude as for the monodomain system, although the degree of constancy is much poorer.

Magnetic 'hardness' whether expressed in terms of H_c, H_p, or W_{Rp} seems in the titanomagnetite system to be associated with the presence of Fe^{2+}. Thus as x increases, magnetic hardness increases. Similarly as z increases, magnetic hardness decreases. This effect is obscured at high x values where the Curie temperatures are low and measurements at room temperature appear to indicate a contrary variation of hardness with x and z. However this variation is due to variation in the magnetic isotherm (T/T_c) at room temperature. Of especial interest are the high values of H_p for the high x, high z, compositions (Fig. 7). It appears that the small residual concentrations of Fe^{2+} in $x=0.8$, $z=0.9$, and $x=0.85$, $z=0.85$ in the presence of the high Ti and vacancy concentrations make a large contribution to the anisotropy possibly via magnetostriction. It could be that a higher degree of oxidation would see the anisotropy fall for these samples as appears to be happening for $x=0.75$. Compositions in this area of high Ti concentration of the titanomaghemite quadrilateral of the ternary diagram seem to be the magnetically hardest part of the system. In nature, although the remanence carried by such material may be extremely stable, the near-stoichiometric first-formed titanomagnetite would not acquire a thermoremanent magnetization. The remanence would be acquired during subsequent maghemitization. However maghemitization is generally considered to occur fairly rapidly in nature ($\sim 10^6$ years) and although this would preclude using reversals of the geomagnetic field as a petrological tool (i.e., to infer rates of maghemitization from the polarity of the titanomaghemite), the chemical remanence, so acquired, would have palaeomagnetic usefulness.

5. Conclusions

The identification of the magnetic minerals in the rock samples or artefacts used in any investigation of the palaeomagnetic field is of fundamental impor-

tance. No single technique is available by which this may be done. Magnetic rotational hysteresis loss measurements can contribute to the identification of minerals, in situ, without danger of bringing about chemical change and may also give some indication of the domain state of the mineral if its concentration can be estimated. The rotational hysteresis loss curves obtained for synthetic analogues in the present study would form the basis for such investigations of rock samples.

Acknowledgments. This work forms part of NERC sponsored research programmes 'Rock and mineral magnetism' and 'Thermoremanence in titanomagnetites'. Two of the authors (A.J.M. and J.B.O'D.) have been in receipt of NERC studentships. Some of the samples were provided by Dr. P.W. Readman and we also thank Dr. Z. Hauptman for invaluable assistance in the preparation and analysis of materials.

References

Cowan, B.K., O'Reilly, W.: The effect of heat treatment on magnetic minerals in red sandstones, studied using the technique of rotational hysteresis. Geophys. J. R. Astron. Soc. **29**, 263–274, 1972

Creer, K.M., Ibbetson, J.D.: Electron microprobe analyses and magnetic properties of non-stoichiometric titanomagnetites in basaltic rocks. Geophys. J. Roy. Astron. Soc. **21**, 485–511, 1970

Day, R., O'Reilly, W., Banerjee, S.K.: Rotational hysteresis study of oxidized basalts. J. Geophys. Res. **75**, 375–386, 1970

Fletcher, E.J., De Sa, A., O'Reilly, W., Banerjee, S.K.: A digital vacuum torque magnetometer for the temperature range 300°–1,000°K. J. Phys. Earth **2**, 311–314, 1969

Jacobs, I.S., Luborsky, F.E.: Magnetic anisotropy and rotational hysteresis in elongated fine particle magnets. J. Appl. Phys. **28**, 467–473, 1957

Manson, A.J., O'Reilly, W.: Submicroscopic texture in titanomagnetite grains in basalts studied using the torque magnetometer and electron microscope. Phys. Earth Planet. Inter. **11**, 173–183, 1976

O'Donovan, J.B., O'Reilly, W.: The preparation characterization and magnetic properties of synthetic analogues of some carriers of the palaeomagnetic record. J. Geomagn. Geoelectr. **29**, 331–344, 1977

O'Reilly, W.: Magnetic minerals in the crust of the Earth. Rep. Prog. Phys. **39**, 857–908, 1976

Owens, W.H.: Rotational hysteresis in haematite. Geophys. J. Roy. Astron. Soc. **49**, 282 (Abstract), 1977

Petersen, N.: Untersuchungen magnetischer Eigenschaften von Titanomagnetiten im Basalt des Rauhen Kulm (Oberpfalz) in Verbindung mit elektronen-mikroskopischer Beobachtung. Z. Geophys. **28**, 79–84, 1962

Readman, P.W., O'Reilly, W.: Magnetic properties of oxidized (cation-deficient) titanomagnetites (Fe, Ti, □)$_3$O$_4$, J. Geomagn. Geoelectr. **24**, 69–90, 1972

Richards, J.C.W., O'Donovan, J.B., Hauptman, Z., O'Reilly, W., Creer, K.M.: A magnetic study of titanomagnetite substituted by magnesium and aluminium. Phys. Earth Planet. Inter. **7**, 437–440, 1973

Soffel, H.C.: The origin of thermoremanent magnetization of two basalts containing homogeneous single phase titanomagnetite. Earth Planet. Sci. Lett. **7**, 201–201, 1969

Stoner, E.C., Wohlfarth, E.P.: A mechanism of magnetic hysteresis in heterogeneous alloys. Philos. Trans. R. Soc. Ser. A: **240**, 559–640, 1948

Received May 24, 1979; Revised Version August 6, 1979; Accepted August 11, 1979

J. Geophys. 46, 201–216, 1979

Journal of Geophysics

The Geomagnetic Field and Its Secular Variation in Finland and Nearby Countries

H. Nevanlinna*

Department of Geophysics, University of Helsinki, Vironkatu 7B, SF-00170 Helsinki 17, Finland

Abstract. All magnetic observations made in Finland before 1910 and certain field measurements and observatory results up to 1970 were used to calculate polynomials representing the geomagnetic field in Finland before 1970. Fourth degree polynomials in time and first degree in latitude and longitude were found to fit the observations with r.m.s. errors of 235nT in H, 0.9° in D and 0.3° in I.

Equivalent dipoles for the magnetic field and its secular variation since 1840 were calculated. By comparing these dipoles with global data, it was found that the nondipole field has been increasing in Finland and nearby countries, and that the secular variation is closely connected with changes in the large nondipole anomaly in Central Asia and with the global 60-year oscillation of the Earth's magnetic field.

Key words: Geomagnetic field – Secular variation – Polynomials – Finland – Equivalent dipoles.

Introduction

Between 1650 and 1910 some 350 measurements of the declination (D), inclination (I) and horizontal intensity (H) of the geomagnetic field had been made in Finland (Nevanlinna and Sucksdorff, 1976). These measurements had only been made sporadically and there had been no long-term recordings to determine the secular variation. Thanks to the growing number of systematic geomagnetic recordings and large-scale surveys, the geomagnetic field and its secular variation in Finland has been well known since 1910.

The main purpose of this paper is to visualize the large-scale geomagnetic field and its secular variation, especially before 1910, by means of graphs and polynomials. Sources of the geomagnetic field in Finland and nearby countries have been analyzed using simple dipole models.

* Department of Geophysics, University of Helsinki, and Finnish Meteorological Institute, Helsinki

0340-062X/79/0046/0201/$03.20

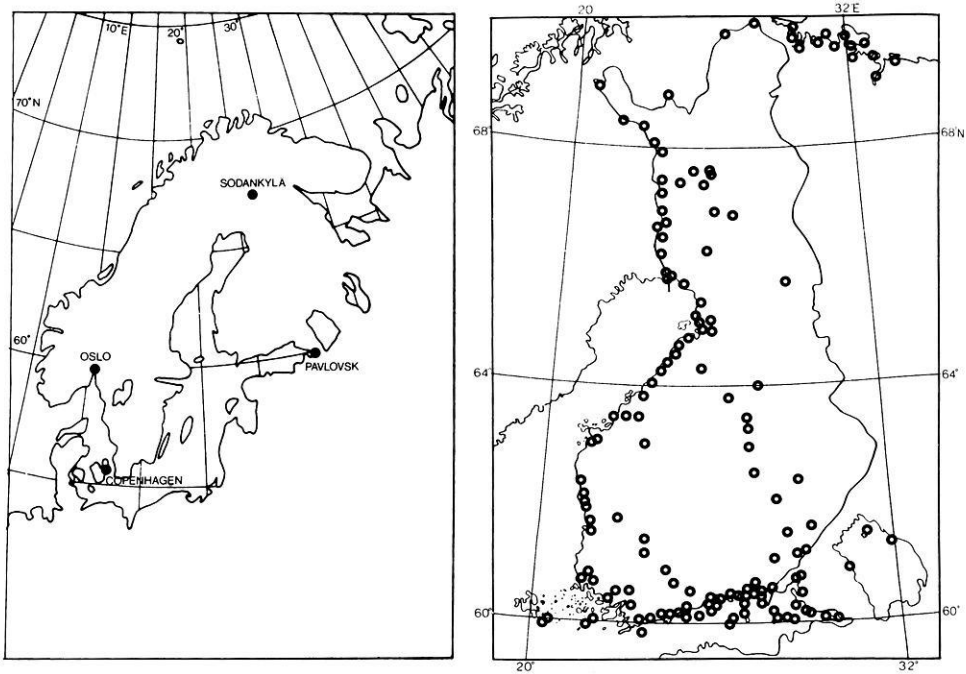


Fig. 1. *Left:* The geomagnetic observatories shown in Figs. 2–4. *Right:* Sites of geomagnetic measurements made before 1910

1. The Geomagnetic Field Expressed by Polynomials

Polynomials are widely used to depict evenly distributed and smoothly varying geomagnetic fields in limited areas like Finland. They enable any component C of the geomagnetic field to be expressed as a function of latitude ϕ, longitude λ and time t, as follows:

$$C(\phi,\lambda,t)=m_0(t)+m_1(t)(\phi-\phi_c)+m_2(t)(\lambda-\lambda_c)+m_3(t)(\phi-\phi_c)^2+\ldots \tag{1a}$$

where the coefficients $m_i(t)$ are also polynomials:

$$m_i(t)=n_{i0}+n_{i1}(t-t_c)+n_{i2}(t-t_c)^2+\ldots \tag{1b}$$

where the coefficients n_{ij} are constants and ϕ_c, λ_c, and t_c are fixed points generally near the middle point of the observations.

The only geomagnetic observations available for Finland before 1910 are 350 observations of H, D, and I, made sporadically at different sites and epochs. Figure 1 shows the sites of all these measurements; a full list of them has been published by Nevanlinna and Sucksdorff (1976). The polynomial selected to represent the H, D, and I components of these measurements, based on Eqs. (1a)

Table 1. The coefficients n_{ij} and their standard deviations σ_{ij} of the H, D, and I polynomials in Eq. (2). D and I are given in arcdegrees and H in nanoteslas (nT)

ij	H		D		I	
	n_{ij}	σ_{ij}	n_{ij}	σ_{ij}	n_{ij}	σ_{ij}
00	$0.14522 \cdot 10^5$	$0.15 \cdot 10^3$	-7.59	0.10	$0.7288 \cdot 10^2$	0.15
01	$0.16996 \cdot 10^2$	5.10	$0.967 \cdot 10^{-1}$	$0.34 \cdot 10^{-2}$	$-0.2327 \cdot 10^{-1}$	$0.71 \cdot 10^{-2}$
02	0.30336	0.16	$0.729 \cdot 10^{-3}$	$0.34 \cdot 10^{-4}$	$0.5619 \cdot 10^{-4}$	$0.18 \cdot 10^{-3}$
03	$-0.13086 \cdot 10^{-1}$	$0.29 \cdot 10^{-2}$	$-0.443 \cdot 10^{-5}$	$0.50 \cdot 10^{-6}$	$0.8771 \cdot 10^{-5}$	$0.24 \cdot 10^{-5}$
04	$0.75483 \cdot 10^{-4}$	$0.15 \cdot 10^{-4}$	$-0.249 \cdot 10^{-7}$	$0.26 \cdot 10^{-8}$	$-0.5685 \cdot 10^{-7}$	$0.12 \cdot 10^{-7}$
10	$-0.33973 \cdot 10^3$	$0.30 \cdot 10^2$	$0.386 \cdot 10^{-1}$	$0.23 \cdot 10^{-1}$	0.5272	$0.39 \cdot 10^{-1}$
11	-1.3979	1.11	$-0.114 \cdot 10^{-2}$	$0.37 \cdot 10^{-3}$	$0.1452 \cdot 10^{-2}$	$0.14 \cdot 10^{-2}$
12	$0.79346 \cdot 10^{-2}$	$0.92 \cdot 10^{-3}$	$0.282 \cdot 10^{-4}$	$0.64 \cdot 10^{-5}$	$-0.1290 \cdot 10^{-4}$	$0.12 \cdot 10^{-4}$
20	$0.49864 \cdot 10^2$	$0.35 \cdot 10^2$	0.643	$0.21 \cdot 10^{-1}$	$-0.3854 \cdot 10^{-2}$	$0.44 \cdot 10^{-1}$
21	$0.55469 \cdot 10^{-1}$	$0.14 \cdot 10^{-1}$	$-0.938 \cdot 10^{-3}$	$0.66 \cdot 10^{-3}$	$-0.1659 \cdot 10^{-3}$	$0.15 \cdot 10^{-2}$
22	$-0.55672 \cdot 10^{-2}$	$0.11 \cdot 10^{-1}$	$0.169 \cdot 10^{-4}$	$0.97 \cdot 10^{-5}$	$0.5494 \cdot 10^{-5}$	$0.55 \cdot 10^{-4}$

Table 2. R.m.s. differencies between polynomial and observed values of H, D, I, and Z. (Numbers of observations in brackets)

Year	-1800	$1800-25$	$1825-50$	$1850-75$	$1875-00$	$1900-25$	$1925-$	Total
H			245 (62)	230 (33)	415 (75)	165 (44)	135 (196)	$235nT$ (410)
D	2.7 (18)	1.2 (20)	1.2 (33)	0.5 (43)	0.5 (31)	0.4 (69)	0.4 (46)	0.9° (260)
I			0.3 (41)	0.4 (5)	0.3 (38)	0.3 (95)	0.3 (119)	0.3° (298)
Z			730	940	790	705	690	$770nT$

and (1b) is

$$C(\phi, \lambda, t) = n_{00} + n_{01}t + n_{02}t^2 + n_{03}t^3 + n_{04}t^4 + (\phi - \phi_c)(n_{10} + n_{11}t + n_{12}t^2) + (\lambda - \lambda_c)(n_{20} + n_{21}t + n_{22}t^2) \quad (2)$$

where t is the year of observations minus 1850, $\phi_c = 63° N$ and $\lambda_c = 26° E$. The coefficients n_{ij} were calculated using the method of least squares. The values used in the fits were those shown in Fig. 1 but also later values from secular points and from geomagnetic observatories were used in order to get the polynomials to fit to the present field as well as possible. Table 1 gives the coefficients n_{ij} and their standard deviations σ_{ij}. The numbers of observations used are shown in Table 2 (in brackets). The main figures in Table 2 are the r.m.s. differences between the observed and polynomial values grouped into 25-year period. The r.m.s. errors of Z were calculated from the r.m.s. errors of H and I.

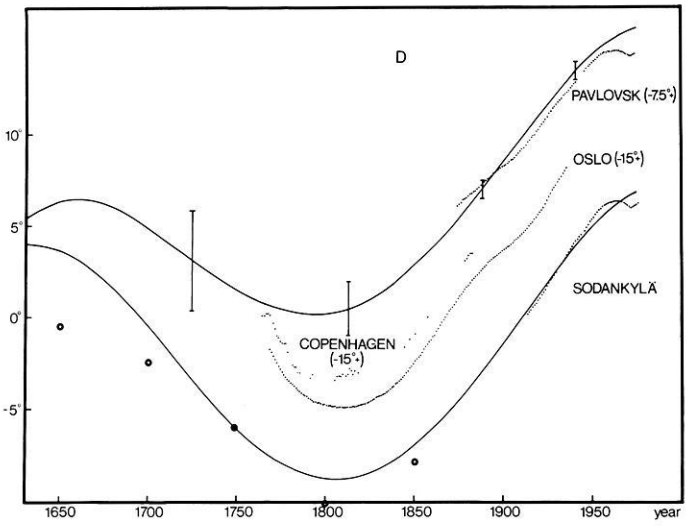


Fig. 2. Declination polynomials *(solid lines)* calculated for the magnetic observatories at Sodankylä and Pavlovsk (after 1947 Leningrad). *Dotted lines:* yearly means of observed values. *Circles: D* values taken from Barraclough's (1974) charts. *Vertical lines:* average r.m.s. error

Figures 2–4 show both polynomial and observed values of the H, D, and Z components for Sodankylä (67°22.2′N, 26°37,8′E) and Pavlovsk (59°41.2′N, 30°29.3′E) (after 1947 Leningrad) in the Soviet Union near the south-east border of Finland. To compare the polynomial values with observed ones not used in the least squares fits, yearly mean values from the geomagnetic observatory in Oslo (59°54.7′N, 10°43.4′E) compiled by Wasserfall (1950), D values from Copenhagen (55.69°N, 12.58°E) compiled by Abrahamsen (1973), and chart values of D from Barraclough (1974) have been included in Figs. 2–4.

As can be seen from Figs. 2–4, the polynomials give roughly the same shape as revealed by the long series of observatory values. Short-term variations can be seen in H and Z between 1890 and 1920 and in D after 1960. These variations are not visible in the polynomials because fourth degree time polynomials are too rigid to follow such rapid changes. The polynomials represent only the long-term (100 years or more) secular variation.

Systematic differencies between polynomial and observed values in Sodankylä are due partly to the sparse distribution of magnetic measurements in North Finland, but mainly to the fact that the Sodankylä observatory lies in a local Z anomaly covering at least 100×100 km. In 1975 the difference in Z between the chart value (Sucksdorff et al., 1975) and the annual mean value at Sodankylä was $600\,nT$ which is 10 times higher than the r.m.s. error for the whole Finland.

The D polynomials are more accurate than other components. They also extend about 150 years farther into the past than H and I. Even the extrapolated D for Sodankylä seems to fit the D values taken from spherical harmonic analysis for 1650–1850 (Barraclough, 1974).

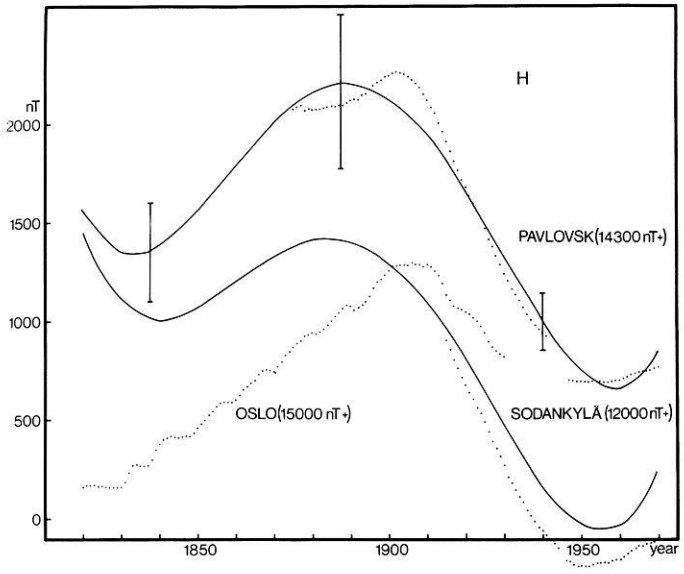


Fig. 3. Polynomials of the horizontal field *(solid lines)* calculated for the magnetic observatories at Sodankylä and Pavlovsk (after 1947 Leningrad). *Dotted lines:* yearly means of observed values. *Vertical lines:* average r.m.s. error

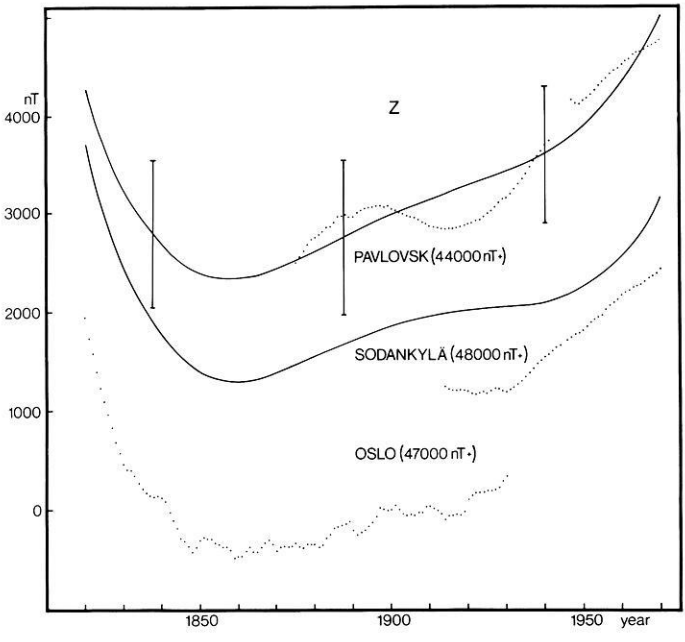


Fig. 4. Vertical field *(solid lines)* calculated from H and I polynomials for the magnetic observatories at Sodankylä and Pavlovsk (after 1947 Leningrad). *Dotted lines:* yearly means of observed values. *Vertical lines:* average r.m.s. error

Figure 2 also indicates that D changed cyclically with a period of about 400 years and an amplitude of 10°. It peaked around 1650 and dropped to a minimum around 1800, in agreement with other European records (Yukutake, 1962; 1967). The rapid oscillations since 1960 may forecast a new peak in the near future.

An interesting feature of the secular variation of Z (Fig. 4) is the rapid drop (c. 150 nT/year) at the beginning of the 19th century. From about 1850 to 1920 Z only increased slowly (c. 10 nT/year), but the secular change of Z has been faster in recent decades: 30–50 nT/year.

The rate of westward drift of the geomagnetic field in Finland has been calculated from the polynomial values of D because its isolines flow roughly north to south, thus causing a maximum gradient in the east-west direction. The average drift for 1750 to 1950 was found to be 0.03 ± 0.01°/year averaged over Finland. This is rather a low value compared to the global average of 0.18°/year for the drift of the whole nondipole field found by Bullard et al. (1950), but roughly the same as that calculated by Leaton (1962) for the Finnish latitudes at the epoch 1955.0. On the other hand, Yukutake (1967) obtained an average drift of 0.36°/year when he studied the drift of D using observatory values all around the world in the last 300 years.

2. Sources of the Magnetic Field and Its Secular Variation in Finland and Nearby Countries

2.1. Equivalent Dipole

The global distribution of the geomagnetic field is usually analyzed by means of spherical harmonic expansions (SHA). Second and subsequent terms of these expansions give the contribution of the nondipole or anomaly field, which averages 20% of the whole field. The nondipole field is concentrated mainly in large centres of continental size, such as the one in Central Asia.

It is well known that changes of intensity and the westward drift of the nondipole field are the main sources of global secular variation (e.g., Yukutake and Tachinaka, 1968). The centres of the secular variation will be located near the centres of the anomaly field if the anomalies are standing and changing their intensity. If the anomalies are drifting, the secular variation centres will be located near the border of the anomaly centres. Nagata and Rikitake (1957) found that in most cases the focus of a secular variation cell lies west of the corresponding anomaly-field focus.

The surface anomaly field can be approximated by radial dipoles within the core (Alldredge and Hurwitz, 1964). The secular variation can be depicted by intensity changes and by the drift of these dipoles (Alldredge and Stearns, 1969), or directly by radial dipoles (Lowes and Runcorn, 1951).

The parameters, location (r_0, ϕ_0, λ_0) and strength (M), of these dipoles are usually determined from observed field values by the method of least squares. The dipole field equations are linearized and the parameters are then determined iteratively. To obtain converging solutions for the parameters, the

minimizing procedure must be started close enough to the minimum. This means that we need to know where the foci of the Z isolines are located and take these as initial values in the least-square fits.

Here, the features of the anomaly field will be interpreted based on one geocentric dipole. From the observed values of Z, H, and D at (ϕ, λ) it is possible to calculate the direction (ϕ_0, λ_0) and strength (M) of this dipole, here termed the equivalent dipole, which gives exactly the observed field at the observation points (ϕ, λ).

The equation of an equivalent dipole governing the total field vector, $\mathbf{B}=\mathbf{H}+\mathbf{Z}$, on the surface of the Earth at the point $\mathbf{R}=(R, \phi, \lambda)$, is

$$\mathbf{B}(\mathbf{R})=3(\mathbf{M}\cdot\mathbf{R})\mathbf{R}R^{-2}-\mathbf{M} \tag{3a}$$

where $\mathbf{M}$ is the strength of the dipole in Teslas (T) defined by

$$\mathbf{M}=(\mu_0/4\pi)\,\mathbf{M}_0 R^{-3} \tag{3b}$$

where $\mu_0/4\pi=10^{-7}\,Tm/A$ and M_0 is the dipole moment in units of $A\,m^2$.

The direction (ϕ_0, λ_0) of the equivalent geocentric dipole defining the 'north pole' or virtual geomagnetic pole (VGP), can be calculated from Eq. (3a) as follows:

$$\begin{aligned}
&\sin\phi_0=\sin\phi\cos\rho+\cos\phi\sin\rho\cos D\\
&\lambda_0=\lambda-A+180^\circ \quad \text{if } \cos\rho\geqq\sin\phi\sin\phi_0\\
&\lambda_0=\lambda+A \quad \text{if } \cos\rho<\sin\phi\sin\phi_0
\end{aligned} \tag{4a}$$

where

$$\begin{aligned}
&\sin A=\sin\rho\sin D/\cos\phi_0\\
&\tan\rho=2H/Z.
\end{aligned} \tag{4b}$$

See also McElhinny (1973). The strength $\mathbf{M}$ obtained from Eq. (3a), $\mathbf{H}$ and $\mathbf{Z}$, is

$$\mathbf{M}=\mathbf{H}-\mathbf{Z}/2 \tag{4c}$$

$$M=(H^2+Z^2/4)^{1/2} \tag{4d}$$

M defined by Eq. (4d) is also termed 'the local magnetic constant' after Bauer (1914), and it has been called 'the length of the dipole vector' by As (1967).

The global distribution of M and its secular variation has been analyzed by Gaibar-Puertas (1953) and As (1967). Ispir et al. (1976) have recently studied the secular variation of M over Turkey in 1965–1970.

If the observed field at every point (ϕ, λ) is a pure dipole field then M, ϕ_0 and λ_0 are constants:

$$\begin{aligned}
&M=(g_{10}^2+g_{11}^2+h_{11}^2)^{1/2}\\
&\tan\phi_0=g_{10}/(g_{11}^2+h_{11}^2)^{1/2}\\
&\tan\lambda_0=h_{11}/g_{11}
\end{aligned} \tag{5}$$

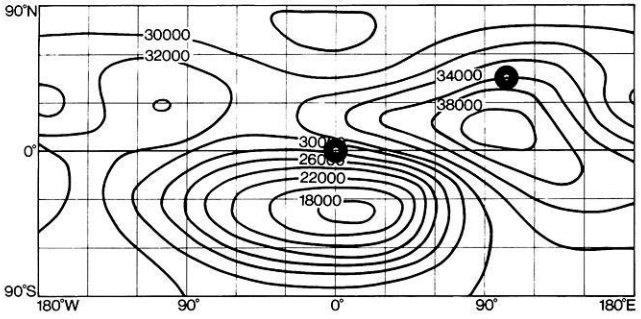


Fig. 5. Isolines (in nT) of the strength M of the equivalent dipoles calculated from IGRF-1965. The dots show Z foci of the Asian and African nondipole anomalies

where g and h are the gaussian coefficients of the spherical harmonic expansion of the magnetic field. Thus any deviation from these values is caused by the anomaly field. If, for example, the M values in a region are greater than the global average, this means there is a positive anomaly component in the observed magnetic field.

Although the isolines of M are related to those of the nondipole field, the foci of the M isolines do not coincide with those of the nondipole field. This can be seen in Fig. 5, which shows the global distribution of M calculated from the IGRF of 1965 (Zmuda, 1971). The M field consists of two large cells of isolines, one in Asia and the other covering Africa and the South Atlantic. The foci of these cells lie about 30° south of the corresponding nondipole centers in Central Asia and Africa.

Figures 6 and 7 show M and the location (ϕ_0, λ_0) of VGP's calculated from observed and polynomial values for Sodankylä, Pavlovsk and Oslo. Figure 6 gives locations of the geomagnetic pole calculated from the gaussian coefficients [Eq. (5)] obtained from SHA for 1840–1950 (Lucke, 1959). As can be seen, the curves for the equivalent dipole differ considerably from the values of the SHA field, which indicates the existence of a large nondipole field in Finland and nearby countries. The VGP's calculated from the Oslo values are nearest to the SHA geomagnetic pole, which means that the nondipole field in Scandinavia weakens from east to west. It can also be seen that the deviation from the SHA field has been growing steadily since the start of the 19th century. It can therefore be concluded that the nondipole field in Finland and nearby countries has been growing ever since. This is also true on a global scale as demonstrated by McDonald and Gunst (1967).

As can be seen in Fig. 6, the VGP's have been drifting westwards at an almost constant rate of 0.8°/year. This drift, however, is not necessarily connected with the general westward drift of the nondipole field, which was found to be 0.03°/year in Finland, as stated in Chap. 1. The rapid drift of the VGP's can also be explained by systematic intensity changes of the dipole and nondipole fields. By differentiating Eq. (4a), it can be shown that $d\lambda_0$ depends mostly on dD. In other words, the rapid westward drift of the VGP's is due to the linear growth of D (or Y) shown in Fig. 2.

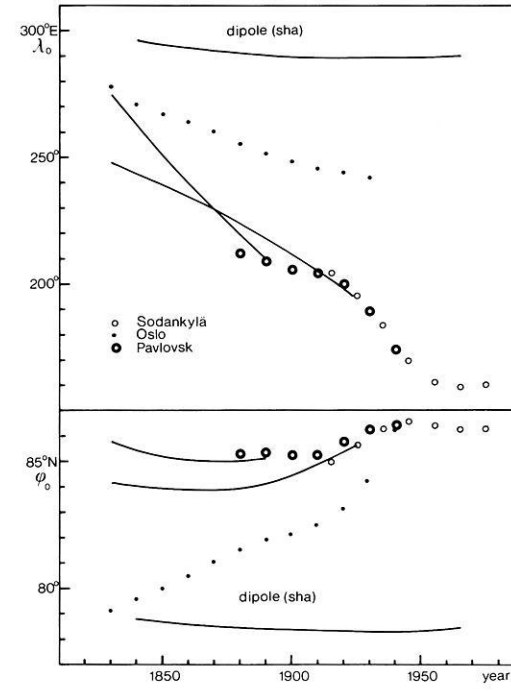


Fig. 6. Locations of the VGP's calculated from polynomials *(solid lines)*. The dots show 10-year means of VGP's calculated from observatory data. Also shown are the locations of the *sha* geomagnetic pole

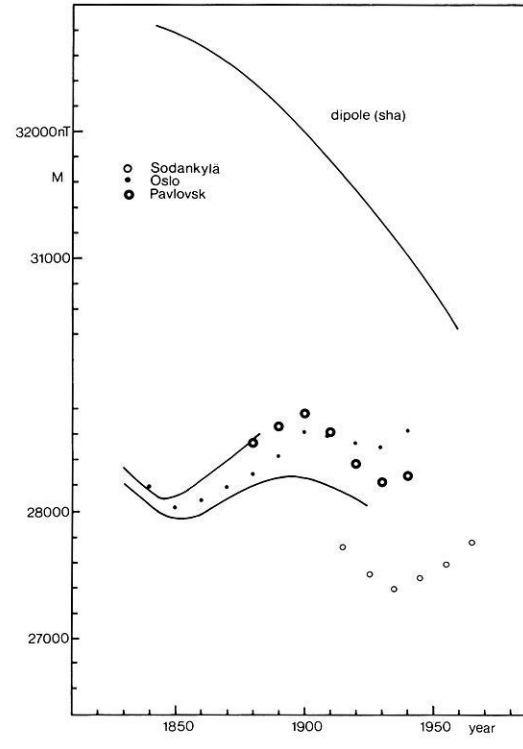


Fig. 7. Strength (M) of equivalent dipoles calculated from polynomials *(solid lines)*. The dots show 10-year means of M calculated from observatory data. Also shown is the strength of the *sha* dipole

Only 10% of the observed increase of D can be explained by changes of the SHA dipole and observed westward drift. Nearly all of the rapid increase of D has therefore been caused by the intensification of the Y component of the nondipole field.

As can be seen in Fig. 7, the curve calculated from the polynomials for the coordinates of Sodankylä differs about $500\,nT$ from the observations. This gap is due to the Z-anomaly covering the Sodankylä observatory.

2.2. *Eccentric Equivalent Dipole of the Secular Variation*

When radial dipoles are used to depict the secular variation, the dipoles must be located in the core near the core-mantle boundary, as shown by Lowes and Runcorn (1951) among others. For a more physical representation of the secular variation vector therefore, an eccentric equivalent dipole must be used instead of a geocentric one. Here we shall present a method for calculating an eccentric equivalent dipole of the secular variation vector starting from the geocentric equivalent dipole.

Using the secular variation vector, $\dot{\mathbf{B}}=\dot{\mathbf{X}}+\dot{\mathbf{Y}}+\dot{\mathbf{Z}}$, ($\dot{\mathbf{X}}=\Delta\mathbf{X}/\Delta t$, etc.), at (ϕ,λ), the magnitude $\dot{M}$ and the direction (ϕ_0,λ_0) of an equivalent geocentric dipole can be calculated from Eq. (4) if H, D, and Z needed in Eq. (4) are replaced by h, d, and z respectively defined by

$$h=(\dot{X}^2+\dot{Y}^2)^{1/2}$$
$$d=\tan^{-1}(\dot{Y}/\dot{X})$$
$$z=\dot{Z}. \tag{6}$$

The 'north pole' (ϕ_0,λ_0) of this dipole can be interpreted to be the virtual pole (focus) of the positive $\dot{Z}$ field.

Since the dipoles used here are radial, all the equivalent dipoles of a secular variation vector are located in the same diametral plane of the Earth. The intersection of this plane with the Earth's surface is a great circle, i.e., an S line, passing through the observations point and the poles of the equivalent dipoles. The orientation of this circle at a point (ϕ', λ') is defined by the declination D of the magnetic field of any equivalent dipole:

$$\cot D=\frac{\sin\phi'\cos\phi_0\cos(\lambda'-\lambda_0)-\cos\phi'\sin\phi_0}{\cos\phi_0\sin(\lambda'-\lambda_0)} \tag{7}$$

where (ϕ_0,λ_0) is the 'north pole' of the geocentric equivalent dipole of $\dot{\mathbf{B}}$. The total length (α) between the observation point and the pole (ϕ_0,λ_0) along an S line, is defined by

$$\cos\alpha=\sin\phi\sin\phi_0+\cos\phi\cos\phi_0\cos(\lambda-\lambda_0). \tag{8}$$

The distance (α') between the observation point and the north pole (ϕ'_0,λ'_0) of an eccentric dipole at q_0R from the geocentre ($q_0=0$ to 1) can be calculated from

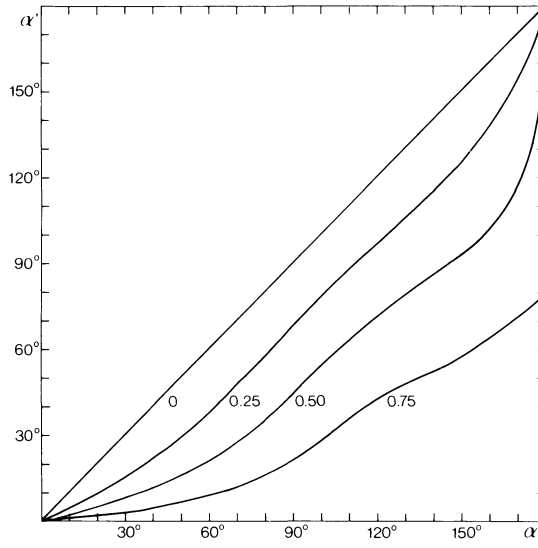


Fig. 8. Distance α' along an S line from the observation point to the pole (ϕ'_0, λ'_0) of an eccentric dipole at $q_0 = 0.0$, 0.25, 0.50, and 0.75, as a function of the distance α to the pole (ϕ_0, λ_0) of a geocentric equivalent dipole

the condition $I_e = I_c$ where I_e is the inclination of the eccentric and I_c that of a geocentric dipole at the point of observation. The equations for I_e and I_c are:

$$\tan I_e = \frac{f_z(\alpha', q_0)}{f_h(\alpha', q_0)} = \frac{\cos\alpha' - 3(\cos\alpha' - q_0)(1 - q_0\cos\alpha')\, r^{-2}}{\sin\alpha'(1 + 3q_0(\cos\alpha' - q_0)\, r^{-2})}$$

$$\tan I_c = \frac{f_z(\alpha, 0)}{f_h(\alpha, 0)} = \frac{-2\cos\alpha}{\sin\alpha}$$

$$r = (1 + q_0^2 - 2q_0\cos\alpha')^{1/2} \tag{9}$$

Figure 8 shows α' versus α when the eccentric dipole is at $q_0 = 0.0$, 0.25, 0.50, and 0.75. It can be seen that α is always greater than α', which means that the 'north pole' of a geocentric dipole, as seen by the observer, is always on the far side of the corresponding eccentric dipole. This far-side effect is similar to that noted by Wilson (1971), who found that paleomagnetic poles in the Tertiary and Quaternary ages tended to lie on the far side of the geographic pole, as seen from the region where the samples were taken. But calculating from an eccentric axial dipole, he obtained poles that were nearer the sampling sites and the geographic pole.

Figure 9 depicts $\dot{Z}_f/\dot{Z}'_f$, the ratio of $\dot{Z}$ at (ϕ_0, λ_0) to $\dot{Z}$ at (ϕ'_0, λ'_0), as a function of α

$$\dot{Z}_f/\dot{Z}'_f = f_z(\alpha', q_0) f_z(\alpha, 0)^{-1} (1 - q_0)^3 r^{-3}. \tag{10}$$

The corresponding ratios of dipole strengths $\dot{M}/\dot{M}'$ defined by Eq. (4d), can be obtained by multiplying the right hand term of Eq. (10) by $(1 - q_0)^{-3}$. From Fig. 9 it can be seen that $\dot{Z}_f/\dot{Z}'_f$ is about 1 and does not depend on q_0 when $\alpha < 60°$, but decreases rapidly when $\alpha > 60°$.

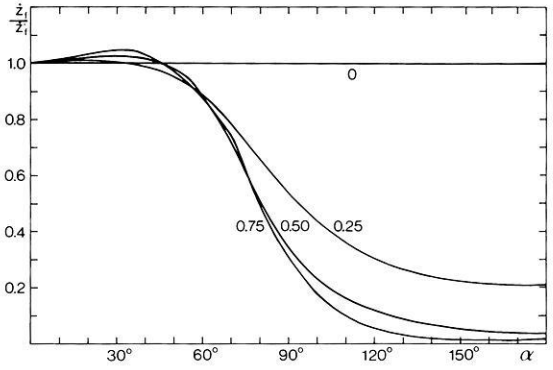


Fig. 9. $\dot{Z}_f/\dot{Z}'_f$ as a function of α, when eccentric equivalent dipole is at $q_0 = 0.0$, 0.25, 0.50, and 0.75. $\dot{Z}_f$ is $\dot{Z}$ at (ϕ_0, λ_0) and $\dot{Z}'_f$ is $\dot{Z}$ at (ϕ'_0, λ'_0)

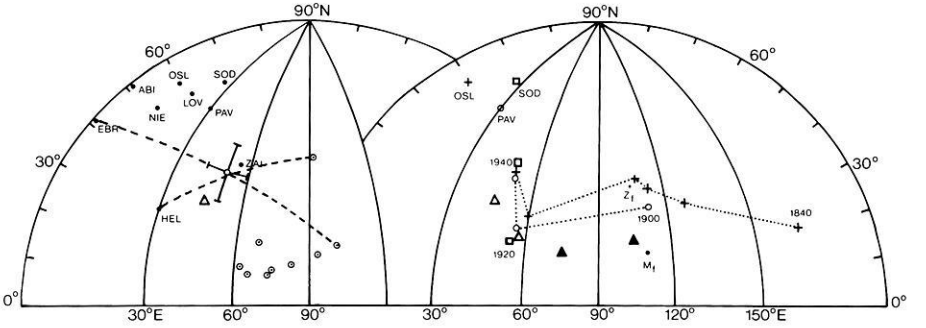


Fig. 10. *Left:* Poles (⊙) of geocentric equivalent dipoles of 10-year mean secular variation observed at nine observatories at 1940. ○ is the mean of the poles (ϕ'_0, λ'_0) of eccentric dipoles at $0.475R$ giving the minimum scatter of poles. Short *solid lines* show the area of 95% confidence around the pole. As examples, *S* lines are given for observatories Helwan *(HEL)* and Ebro *(EBR)*. △ is the focus of $\dot{Z}$ according to *SHA* maps of Cain and Hendricks (1967). *Right:* Path of the poles of eccentric ($q_0 = 0.475$) equivalent dipoles calculated from the secular variation observed at Sodankylä *(SOD)*, Oslo *(OSL)* and Pavlovsk *(PAV)* at 20-year intervals since 1840. △ is focus of $\dot{Z}$ of SHA maps for the epochs 1920 and 1940 by Cain and Hendricks (1967). ▲ is focus of isolines of M according to Gaibar-Puertas (1953) maps for the epochs 1900 and 1920. M_f is focus of the isolines of M from Fig. 5. Z_f is focus of *SHA* nondipole Z from IGRF-1965

The procedure for finding the parameters $(\phi'_0, \lambda'_0, \dot{M}')$ of the eccentric equivalent dipole of a secular variation vector $\dot{\mathbf{B}}$ observed at (ϕ, λ) is as follows:

– Calculate the geocentric equivalent dipole $(\phi_0, \lambda_0, \dot{M})$ and the length (α) of the *S* line using Eqs. (4), (6), and (8).
– Calculate α' for a given value of q_0 from Eq. (9).
– Calculate (ϕ'_0, λ'_0) from α' using Eqs. (7) and (8).
– Calculate $\dot{M}'$ using Eq. (10).

If several observations covering a large area are available, a best fitting dipole can be determined by minimizing the scatter of the poles (ϕ'_0, λ'_0) found for each vector $\dot{\mathbf{B}}$. An example of this is given in Fig. 10, which shows the poles of the geocentric and the best fitting dipoles of the secular variation at nine

Table 3. Fisherian statistics for a best fitting eccentric dipole of the secular variation at the epoch 1940

$\phi_0 = 46.2° N$	$N = 9$
$\lambda_0 = 50.6° E$	$\alpha_{95} = 9.8°$
$q_0 = 0.475$	$k = 205$

European observatories at the epoch 1940. The minimum scatter was found by changing q_0 and calculating Fisher's precision parameter k (Fisher, 1953) for the corresponding values of (ϕ'_0, λ'_0). The scatter is at its minimum when k is at its maximum. The best fitting dipole was found to lie at $46.2° N$, $50.6° E$, and $0.475 R$ from the centre of the Earth. Its strength was $-8.1\, nT$/year; see also Table 3 and Fig. 10. This best fitting dipole gives $\dot{Z}$, $\dot{X}$, and $\dot{Y}$ at the observatories used with r.m.s. errors of $6.4\, nT$/year, $7.3\, nT$/year and $4.0\, nT$/year respectively.

The location of the mean pole found here is fairly close to the focus ($36° N$, $45° E$) of $\dot{Z}$ on the SHA map Cain and Hendricks (1967), which lies almost within the area of 95% confidence for the mean pole. The distance q_0 in Table 3 is near the core-mantle boundary, same as that found by Lowes and Runcorn (1951) for their dipoles.

The relatively small area of scatter of the poles in the left-hand side of Fig. 10 indicates that secular variation in Europe and West Asia has a common source. The right-hand side of Fig. 10 shows the probable movements of the pole of this source since 1840. It is based on eccentric equivalent dipoles ($q_0 = 0.475$) of the secular variation observed at Sodankylä, Oslo and Pavlovsk. The poles appear to have been drifting westwards at a rate of 1°/year. The path of the poles lies in the area of high intensity of the SHA nondipole field, indicating a correlation between the secular variation in North Europe and the anomaly field in Asia. This correlation seems to prevail over a large area of Eurasia, because the focus of the isolines of $\dot{M}$ taken from Gaibar-Puertas (1953) global chart has been moving along a path near the focus of the Asian anomaly, as can be seen in the right-hand side of Fig. 10.

Figure 11 depicts the intensity $\dot{M}$ of the geocentric equivalent dipoles. Striking features in the variations of $\dot{M}$ are the sharp minimum just before 1900, the rapid, almost threefold increase in the subsequent 20 years, and other rapid drop to a minimum around 1960. Observed focal values of $\dot{Z}$ taken from SHA maps (Cain and Hendricks, 1967), reduced to the centre of the Earth show roughly the same decreasing trend as those calculated from the observatory values.

$\dot{M}$ in Fig. 11 oscillates in a cycle of approximately 60 years, which is in phase with the curves of the variation of the Earth's rotation found by Yukutake (1972; 1973) and Jin and Thomas (1977). According to these authors, the variation in the length of the day correlates well with the general 60-year oscillation of the Earth's magnetic field. So, it is suggested here that the variation of the equivalent poles and their intensities, and the change of the SHA $\dot{Z}$ focus in South-west Asia are caused by the 60-year oscillation of the magnetic field in the Asian nondipole anomaly.

The focus of the 60-year oscillation of the secular variation in Europe has also been determined, using other methods, by Barta (1978) and Golovkov and

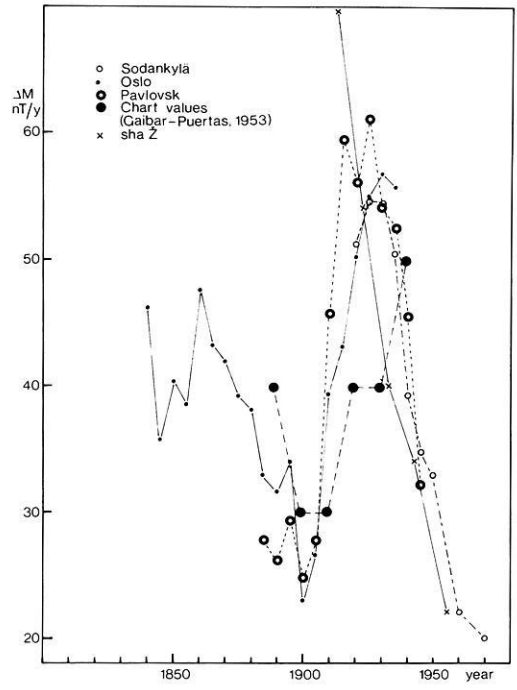


Fig. 11. Ten-year means of the intensity $\dot{M}$ of the geocentric equivalent dipoles calculated from secular variation data

Kolomiytseva (1971), who found foci at (28° N, 68° E) and (27° N, 41° E) respectively. These foci are near the path of the equivalent poles shown in Fig. 10.

Conclusions

The polynomials calculated for the magnetic field in Finland before 1970 reveal only the general trend of the magnetic field and its secular variation. They fail to disclose periods of rapid variations registrated at magnetic observatories between 1890 and 1920.

As could be expected, accuracy was best in the D polynomial, which revealed the declination in Finland with a maximum error of 3° from 1650 to 1850, and less than 1° after 1850.

The rather low value of the westward drift found here means that the drift, though global, is not an important source of secular variation in Finland. The intensification of the SHA nondipole field, especially its Y component, seems to be the major source of secular variation in Finland and nearby countries during the last 150 years.

The equivalent dipole method applied here can be used to locate source regions of secular variation. It was found that a single source has been dominated the secular variation in Europe during the past 150 years.

Because the SHA dipole part of the observed field **B** is usually much stronger than SHA nondipole part, the equivalent dipole method cannot be used

to locate SHA nondipole centers. But if the geocentric equivalent dipole is calculated for large areas at a given epoch, any variation of intensity and direction of the dipole can be assumed to be caused by the SHA nondipole field because the contribution of the SHA dipole field to the equivalent dipole parameters is constant. The method thus reveals the existence of an SHA nondipole anomaly at the point of observation.

Acknowledgement. This study was financially supported by Suomen Tiedeseuran Sohlbergin delegaatio.

References

Abrahamsen, N.: Magnetic secular variation in Denmark, 1500–1970. J. Geomagn. Geoelectr. **25**, 105–111, 1973

Alldredge, L.R., Hurwitz, L.: Radial dipoles as the sources of the Earth's magnetic field. J. Geophys. Res. **69**, 2631–2640, 1964

Alldredge, L.R., Stearns, C.O.: Dipole model of the sources of Earth's magnetic field and secular variation. J. Geophys. Res. **74**, 6583–6593, 1969

As, J.: Past, present and future changes in the Earth's magnetic field. In: Magnetism and Cosmos, W.R. Hindmarsh, F.J. Lowes, P.H. Roberts, S.K. Runcorn, eds.: pp. 29–44. Edinburgh: Oliver and Boyd Ltd. 1967

Barraclough, D.R.: Spherical harmonic analyses of the geomagnetic field for eight epochs between 1600 and 1910. Geophys. J. R. Astron. Soc. **36**, 497–513, 1974

Barta, G.: Secular variation of the terrestrial magnetic field and the internal structure of the Earth. In: Electromagnetic field of the Earth, M. Hvozdara, ed.: pp. 49–58. Bratislava: Publishing House of the Slovak Academy of Sciences 1978

Bauer, L.A.: The local magnetic constant and its variation. Terr. Mag. Atmos. Electr. **19**, 113–125, 1914

Bullard, E.C., Freedman, C., Gellman, H., Nixon, J.: The westward drift of the Earth's magnetic field. Philos. Trans. Soc. London, Ser. A **234**, 67–92, 1950

Cain, C.C., Hendricks, S.: The geomagnetic secular variation, 1900–1965, Greenbelt, Md.: Goddard Space Flight Center September 1967.

Fisher, R.A.: Dispersion on a sphere. Proc. Soc. London Ser. A: **217**, 295–305, 1953

Gaibar-Puertas, C.: Variación secular del campo geomagnético. Observatorio del Ebro Memoria, No. **11**, 1953

Golovkov, V.P., Kolomiytseva, G.I.: Morphology of 60-year geomagnetic field variations in Europe. Geomagn. Aeron. **11**, 571–574, 1971

Ispir, Y., Isikara, A.M., Özden, H.: Variation in the local magnetic constant and seismicity of Turkey. J. Geomagn. Geoelectr. **28**, 137–143, 1976

Jin, R., Thomas, D.M.: Spectral line similarity in the geomagnetic dipole field variations and length of day fluctuations. J. Geophys. Res. **82**, 828–834, 1977

Leaton, B.R.: Geomagnetic secular variation for the epoch 1955. O. R. Obs. Bull. **57**, 245–270, 1962

Lowes, F.J., Runcorn, S.K.: The analysis of the geomagnetic secular variation. Philos. Trans. R. Soc. London, Ser. A: **243**, 525–546, 1951

Lucke, O.: Analyse der Veränderungen des Erdmagnetischen Hauptfeldes, aus den Potentialentwicklungen erschlossen. In: Geomagnetismus und Aeronomie, Band III, G. Fanselau, ed.: pp. 217–312. Berlin: Deutscher Verlag der Wissenschaften 1959

McDonald, K.L., Gunst, R.H.: An analysis of the Earth's magnetic field from 1835 to 1965. ESSA Tech. Rep., IER 46-IES 1, 1967

McElhinny, M.W.: Paleomagnetism and plate tectonics. Cambridge: Univ. Press 1973

Nagata, T., Rikitake, T.: Geomagnetic secular variation during the period 1950–1955. J. Geomagn. Geoelectr. **9**, 42–50, 1957

Nevanlinna, H., Sucksdorff, C.: Magnetic field in Finland, 1800–1970. Finn. Met. Inst. Stud. Earth Magn. **25**, 1976

Sucksdorff, C., Nevanlinna, H., Welling, P.: Magnetic charts of Finland for 1975. 0. Finn. Met. Inst. Stud. Earth Magn. **24**, 1975

Wasserfall, K.F.: A study on the secular variation of magnetic elements based on data for *D*, *I*, and *H* for Oslo, 1820–1938. J. Geophys. Res. **55**, 275–300, 1950

Wilson, R.L.: Dipole offset – The time average paleomagnetic field over the past 25 m. y. Geophys. J. R. Astron. Soc. **22**, 491–504, 1971

Yukutake, T.: The westward drift of the magnetic field of the Earth. Bull. Earthq. Inst. **39**, 427–476, 1962

Yukutake, T.: The westward drift of the Earth's magnetic field in historic times. J. Geomagn. Geoelectr. **7**, 1–14, 1967

Yukutake, T.: The effect of change in the geomagnetic dipole moment on the rate of the Earth's rotation. J. Geomagn. Geoelectr. **24**, 19–47, 1972

Yukutake, T.: Fluctuations in the Earth's rate of rotation related to changes in the geomagnetic dipole field. J. Geomagn. Geoelectr. **25**, 195–212, 1973

Yukutake, T., Tachinaka, H.: The westward drift of the geomagnetic secular variation. Bull. Earthq. Inst. **46**, 1075–1102, 1968

Zmuda, A.J.: World Magnetic Survey. IAGA Bull. **28**, 1971

Received March 22, 1979; Revised Version July 11, 1979; Accepted August 10, 1979

J. Geophys. 46, 217–224, 1979

Journal of Geophysics

Age and Crustal Structure of the Canary Islands

A Discussion

H.-U. Schmincke

Institut für Mineralogie, Ruhr-Universität, Universitätsstr. 150,
D-4630 Bochum, Federal Republic of Germany

Abstract. The postulation of a Mesozoic age for the shield-building basaltic series of Gran Canaria and Tenerife by Storevedt et al. (1978), based upon paleomagnetic data from these islands, is inconsistent with abundant and concordant K/Ar-ages from several laboratories. These latter data leave no doubt that no rocks older than Mid-Miocene have been found on Gran Canaria and Tenerife. Moreover, no volcaniclastic layers older than Miocene were found in Deep Sea Drilling cores near the Canary Islands. Also, the volcanic apron around at least Gran Canaria appears to be Miocene in age judging from seismic reflectors that extend to well-dated drilled sections. There is no evidence for a sialic crustal layer extending beneath all or any of the Canary Islands.

Key words: Canary Islands – Crustal structure – Volcanic Islands – Potassium Argon Ages – Paleomagnetism.

1. Introduction

Storetvedt et al. (1978) have used paleomagnetic data from Gran Canaria and Tenerife to postulate that the Canary Islands were built up during two major volcanic pulses, the first occurring in the Mesozoic (Late Cretaceous or earlier) and the second during Late Tertiary. They also propose that their data, combined with other lines of evidence, favor 'a continental origin of a crustal belt (including the Canary Islands) extending seaward into water depths of at least 4,500 m'.

Storetvedt et al. (1978) appear to have overinterpreted their data and those from the literature, and several relevant papers were not discussed:

A variety of geophysical and geological studies (see below) have been carried out in this part of the Atlantic (off the northwest coast of Africa) during the last fifteen years, this being a key target area in studying the evolution of passive continental margins. The conclusions of Storetvedt et al. (1978) will, therefore, be discussed in detail, using published and some unpublished material.

0340-062X/79/0046/0217/$01.60

2. Age of the Canary Islands

K/Ar-age determinations on rocks from the Canary Islands were published by Rona and Nalwalk (1970), Abdel-Monem et al. (1971; 1972), Lietz and Schmincke (1975), Grunau et al. (1975), Stillman et al. (1975) and McDougall and Schmincke (1976) and a general discussion of these data was given in Schmincke (1976). More than 80 K/Ar-age determinations are available for Gran Canaria alone, making this one of the most thoroughly dated oceanic islands. The age determinations by Abdel-Monem et al. (1971; 1972) are confirmed in a general way by the later workers. The oldest lavas on Gran Canaria, the shield-building alkali basalts, totalling 1,000 m exposed above sea level, were erupted about 13.5–13.7 M.a. ago (McDougall and Schmincke, 1976).

In striking contrast, Storetvedt et al. (1978) concluded from their paleomagnetic measurements on oriented hand specimens that the pole positions for these lavas are similar to *Mesozoic* pole positions in Africa and that the *minimum* age of this series is, therefore, upper Cretaceous and 'it may well have covered a broad time span within the Mesozoic' (loc. cit. p. 328). With the Mesozoic spanning from ca. 65–230 M.a. and the Late Cretaceous from about 65–100 M.a., the difference in age determined by the authors cited and Storetvedt et al. (1978) amounts to some 50 to 200 M.a.

(a) Are the Published K/Ar-Determinations on Rocks From Gran Canaria Reliable?

Storetvedt et al. (1978), in noting the discrepancy, suggested that incomplete retention of radiogenic argon in the older lavas, possibly caused by 'extensive mineral alteration in the rocks and by extensive magmatic activity in late Tertiary times' (p. 329) makes the published K/Ar-age determinations of dubious value. In support of their opinion they quote age discrepancies between the data by Abdel-Monem et al. (1971; 1972) and stratigraphic data of other authors. Indeed, it was the *relatively* large spread of several million years for the shield lavas on Gran Canaria that resulted in the detailed study of McDougall and Schmincke (1976), initiated because geologic evidence suggested rapid build-up of these shields (i.e., within < 1 M.a.) as on other oceanic volcanic islands. Their prediction was confirmed by 20 K/Ar-determinations lying between 13.1 and 13.9 M.a. contrasting with 5 determinations between 10.2 and 16.1 M.a. by Abdel-Monem et al. (1971) for only the upper part of the series. Differences in sampling and/or analytical methods may explain this descrepancy, much emphasis being placed in the study of McDougall and Schmincke (1976) to collect and date only fresh rocks. I suspect (Schmincke, 1976) that careful and detailed study of the shield lavas of other Canary islands will reveal similar differences with the *relatively* large spread in the reconnaissance study of Abdel-Monem et al. (1971; 1972) being reduced to 1 M.a. or less. While such differences, which are common in similar studies of other islands, are relevant for problems of magmatic evolution, they are rather unimportant for the question posed by Storetvedt et al. (1978). I have no doubt that more detailed study will confirm

the Miocene (eastern and central Canaries) or Pliocene (western Canaries) age of the shields. For Gran Canaria, agreement between the studies of Abdel-Monem et al. (1971), Lietz and Schmincke (1975) and McDougall and Schmincke (1976) is excellent for rocks younger than 10 M.a.

(b) Are the Rocks of the Shield-Building Series of Gran Canaria Unsuitable for K/Ar-age Determinations?

There are indeed altered basalts within the shield series – but highly altered rocks can also be found on the top of Recent Pico de Teide and alteration may not, therefore, be related to age. Thin section examination of a number of shield lavas dated by Abdel-Monem et al. (1971) shows some to be somewhat altered, however, the vast majority of the shield-building basalts are rather fresh with regard to K-bearing phases (generally feldspar). Suitable specimens can be found if sampling is done carefully.

Red soils (Storetvedt et al., p. 318) on Gran Canaria have developed (during the Pliocene?) on the northern side of the island, chiefly in Pliocene/Quaternary basaltic pyroclastics. Red horizons are rare in the shield basalts, and close examination shows that most, if not all, existing red layers result from baking by overlying lavas. The further suggestion that later magmatic activity (phases II and III, Schmincke, 1976) has caused radiogenic argon to have leaked from the rocks is, again, unsupported by facts. Examination of the extremely well exposed, up to 1,000 m thick, shield series shows that samples studied by Abdel-Monem et al. (1971, 1972) and McDougall and Schmincke (1976) on Gran Canaria are mostly from areas many kilometers away from any later intrusions. These rocks have generally not been heated appreciably since their solidification, the alteration being of low temperature origin which is common in rocks of this age but is not detrimental for age determinations when only fresh portions are selected. The high degree of consistency within and between laboratories, the agreement between the stratigraphic sequence and the age determinations, and between rock and mineral ages indicate that construction of the *subaerial* part of the shield stage of Gran Canaria took place chiefly between 13 and 14 M.a. B.P.

(c) Biostratigraphic Evidence

During Leg 47A, 'Glomar Challenger' drilled and continuously cored a hole (site 397) about 100 km SE of Gran Canaria penetrating 1,453 m of sediment with a 1,300 m thick section of Quaternary to earliest Miocene (20 M.a.) directly overlying lower Cretaceous sediments. The oldest volcaniclastic debris flow deposits are about 17 m.y. old, based on biostratigraphic evidence, while abundant silicic air fall tephra layers occur throughout the younger sediments (Schmincke and v. Rad, 1979). The only exception is a 19-M.a.-old ash layer. No pyroclastic rocks were found in early Miocene, Paleogene or Cretaceous sediments penetrated at the adjacent slope DSDP Dite 369 (Lancelot et al., 1977) nor in the earliest Miocene or Early Cretaceous of Site 397.

The chemical composition of Miocene glass shards found at Site 397 matches very well that of the Miocene rhyolites of Gran Canaria of the same age, this being a rock group of unique composition in the Canaries. This evidence indicates that the main subaerial volcanic phase of the Canary Islands commenced during the early to middle Miocene. Moreover, Schmincke and v. Rad (1979) have interpreted different types of pyroclastic rocks from Site 397 as representing the submarine, subaerial shield and subaerial post-erosional stages of the island volcanoes.

Age of the Canary Islands (Summary). All available data indicate that Gran Canaria did not rise above sea level before mid-Miocene. The submarine part of the island may not be *appreciably* older because Miocene seismic reflectors traced from Glomar Challenger sites clearly interfinger with the volcanic apron around Gran Canaria (Wissmann, personal communication 1979).

The only published K/Ar-age determinations of Canary Island rocks older than Miocene are two Oligocene ages on plutonic and metamorphic rocks from Fuerteventura (Abdel-Monem et al., 1971); rocks from the same complex were dated as Miocene by Grunau et al. (1975). The intrusions must obviously be younger than the Late Cretaceous sedimentary rocks which they intrude. These rocks are from the Betancuria massif which also contains folded and uplifted upper Cretaceous marine sedimentary rocks and is clearly the most complex area in the Canary Islands (Stillman et al. (1975). According to Robertson and Stillman (1979) the oldest volcanic rocks on Fuerteventura are of submarine origin and were deposited on Albian calcareous pelagic sedimentary rocks, after an hiatus of unknown duration, in Late Cretaceous or early Tertiary time. The top part of the sequence, however, is interbedded with Late Oligocene to early Miocene shallow water sedimentary rocks that predate intrusion of a major dike swarm and later igneous events. Canary Island volcanoes do not appear to have risen above sea level before mid-Miocene. Submarine activity may have started earlier (Eocene?) based on still inconclusive seismic evidence (Uchupi, et al., 1976; Grunau et al., 1975; Watkins et al., in press). A 'mid-Tertiary' age of the Canary Island volcanism is also indicated by the obvious disruption of the Tertiary continental rise prism into a thick western segment (west of the Canaries) and eastern part between Cape Bojador and the Canaries (Uchupi et al., 1976, p. 846, Fig. 26).

Paleomagnetic Data. Storetvedt et al. (1978) base their conclusion on paleomagnetic data from Gran Canaria and Tenerife and the reconstructed pole positions for the shield series which they conclude indicates upper Cretaceous or older Mesozoic age. As a petrologist I am unable to fully judge their evidence but it must be emphasized that Watkins (1973) concluded that *his* pole positions were quite compatible with the Late Tertiary ages. It would have been valuable to compare data from the more detailed treatment by thermal and alternating field demagnetisation by Storetvedt et al. (1978) with those by Watkins (1973) using samples from the same sites. Storetvedt et al. (1978), who give no sampling localities, state that 'it is not known to which extent these sampling localities may duplicate those of earlier studies' (op. cit. p. 319). Earlier authors did

give detailed localities, however, and McDougall and I were able to resample all localities of Abdel-Monem et al. (1971) of interest to us, using published descriptions and still visible drill holes. Watkins, now deceased, unfortunately cannot defend himself, which, no doubt, he would have done with vigor.

3. Nature of the Crust Beneath and Around the Canary Islands

Storetvedt et al. (1978) have combined their speculations on a Mesozoic age for the islands of Gran Canaria and Tenerife with unrelated and, in part unsupported, highly speculative or questionable data to extend their idea to all of the Canary Islands and to the regional crustal structure.

(a) Basement Complexes

Storetvedt et al. (1978) restate the hypothesis of Fuster et al. (1970) that a layered basement complex underlies all of the Canaries and is exposed in some. However, there is little foundation for this theory (Schmincke, 1976). The two large central islands, Tenerife and Gran Canaria, have no such basal complexes exposed. The complex on La Palma is about 2–3-m.y.-old judging from microfossils found in inter-pillow sediments (Berggren, personal communication to H. Staudigel, 1978). The layered basement complex on Fuerteventura is Miocene or, in part, older, while that of La Gomera has not been dated. A much more likely explanation is that these complexes are part of the substructure of an island, being a plexus of submarine volcanics, intrusives and sediments that probably make up much of the submarine part of most oceanic volcanic islands (Schmincke, 1976; Schmincke and Staudigel, 1976).

(b) Sialic Substratum Under the Canaries?

The (rather old) question as to whether the Canary Islands are underlain by continental or oceanic crust has been discussed elsewhere in detail (Schmincke, 1976). Schmincke (1967) suggested that Gran Canaria and probably the Western Canaries were volcanic islands built on oceanic crust and not part of a tectonically broken up, sialic microcontinent. Utilizing geological (Rothe and Schmincke, 1968) and marine geophysical data (Dash and Bosshard, 1968), a 'subdivision' was made between the Eastern Canaries, possibly or probably underlain by continental crust, and an oceanic central and western area. Dietz and Sproll (1970) built on this hypothesis and postulated, based on *their* computer fit on Africa and North America, that the Eastern Canaries block is probably a microcontinent or sialic continental fragment detached from the African margin. Rona and Nalwalk (1970), in reporting some K/Ar-age determinations from Fuerteventura, simply restated views of the earlier authors and concluded 'We *speculate* that the igneous activity was related to possible rifting of the Eastern Islands from Africa' (pp. 2118–2119, italics mine). Storetvedt et al. (1978, p. 318) on the other hand, stating that Rona and Nalwalk (1970)

and Dietz and Sproll (1970) had given '*evidence* for a sialic substratum under the Canaries' (italics mine), omit the major distinction of most authors since 1967 between Eastern and Western Canaries. In three papers by Shell geologists (Beck and Lehner, 1974; Grunau et al., 1975; Lehner and De Ruiter, 1977) (neither mentioned by Storetvedt et al., 1978), apparently using much unpublished information, cross-sections of the area between the Eastern Canaries and Africa are presented. Interestingly, the first published section shows continental crust extending from Africa to the Eastern Canaries (Beck and Lehner, 1974, Fig. 12) while in subsequent, more detailed accounts of the crustal structure in the vicinity of the Canaries, this continental crust has been replaced by a question mark (Grunau et al., 1975; Fig. 5; Lehner and De Ruiter, 1977, Fig. 3). The Concepcion Bank part of the Canary Ridge is interpreted as a tectonic unit of volcanic origin based on refraction seismic investigations (Weigel et al., 1978). The boundary between oceanic and continental crust could even lie near the present Moroccan slope, if recent seismic data, showing a 5.8 km/s layer at 8.5 km depth within (Wissmann et al., 1977) the sedimentary basin between the Eastern Canary Island – Concepcion Bank complex and the shelf of Morocco are interpreted as representing Jurassic deep water carbonate or even oceanic layer 2, as also suggested by the 6 km/s velocity given in Fig. 5 of Grunau et al. (1975) (Weigel et al., 1978).

(c) Erosional Unconformities in the Shield Volcanics on Gran Canaria

Storetvedt et al. (1978) quote Fuster et al. (1968) for the existence of erosional unconformities between the 'various petrographically different units of the older series' on Gran Canaria (p. 318). They have used this evidence to state 'Recalling the existence of erosional discordances within the older lava sequence of Gran Canaria there are in fact reasons to suggest that rifting and lava extrusion may have taken place intermittently (within a subsiding continental belt between Africa and North America) over a time span covering *at least a greater part of the Mesozoic*' (italics mine). Actually, the only major erosional unconformity proposed by Fuster et al. (1968) does not exist. Fuster et al. (1968) opposed the view of Schmincke (1967) that extrusion of a thick series of rhyolite lavas and ignimbrites followed immediately after the shield-building basalts, that this series was associated with caldera collapse, and that there followed intrusion of a large trachytic cone sheet swarm. This group of investigators (Hernán Reguera, 1976) has now accepted both caldera collapse and cone sheet swarm and abandoned the idea of a large erosional unconformity (Hernán Reguera, 1976). McDougall and Schmincke (1976) have documented in detail that there are no significant age differences between the older basalts and the overlying rhyolites, nor is there a significant age difference above or below the only major erosional unconformity found *within* the shield basalt series (i.e., magmatic phase I of Schmincke, 1976, or the 'older lavas' of Storetvedt et al., 1978), this having been described by Schmincke, 1968, but not by Fuster et al., 1968. Moreover, erosional unconformities such as the latter and that between magmatic phase I and the Pliocene magmatic phase 2 on Gran Canaria are commonplace on all volcanic islands.

4. Conclusions

The better known elements of the evolution of the Canary Islands are, judging from presently available data from marine geophysics, deep sea drilling and geological studies of the islands: appearance of volcanic islands – and possibly initiation of melting anomalies in the mantle – during early Miocene, initially in the Eastern Canaries, followed by Gran Canaria (which is rather precisely data) and later the Western Canaries. This migration corresponds approximately with known sea floor spreading rates. Exceptions to this pattern *may be* the basal complexes of Fuerteventura and Gomera but data are still sparse and contradictory. The location of the boundary between the oceanic and continental crust is still unknown but evidence is mounting that it is at least as far east as the Eastern Canaries and is possibly still further east. What initiated the melting anomaly under the Canaries is unknown. Important structural elements in the evolution of the archipelago include vertical movements in some islands with thousands of meters of throw. These are major, tectonic events ranging in age from early Tertiary (?) in Fuerteventura to Pliocene/Quaternary on La Palma, suggesting that the stress field responsible for these movements may still be operative. Other islands like Gran Canaria were remarkably stable with respect to sea level, indicated by the similarity in evelation of Mid-Miocene valley bottoms of these islands to those of today. Regressions and transgressions at the Miocene/Pliocene boundary are an exception. Most islands experienced several major magmatic phases, similar to those known from Hawaii, with a voluminous shield phase of moderately alkalic basalts followed after major erosional intervals by much smaller volumes of more alkalic undersaturated magmas.

Acknowledgements. Work on the Canary Islands was supported by the Deutsche Forschungsgemeinschaft. I thank G. van Kooten, U.v. Rad, G. Wissmann, and I. McDougall for critically reading the manuscript.

References

Abdel-Monem, A., Watkins, N.D., Gast, P.W.: Potassium-argon ages, volcanic stratigraphy and geomagnetic polarity history of the Canary Islands: Lanzarote, Fuerteventura, Gran Canaria and La Gomera. Am. J. Sci. **271**, 490–521, 1971

Abdel-Monem, A., Watkins, N.D., Gast, P.W.: Potassium-argon ages, volcanic stratigraphy, and geomagnetic polarity history of the Canary Islands: Tenerife, La Palma and Hierro. Am. J. Sci. **272**, 805–825, 1972

Beck, R.H., Lehner, P.: Oceans, new frontier in exploration. Am. Assoc. Petrol. Geol. Bull. **58**, 376–395, 1974

Dash, B.P., Bosshard, E.: Crustal studies around the Canary Islands. Prague: Proc. 23rd Intern. Geol. Congr. **1**, 249–260, 1968

Dietz, R.S., Sproll, W.P.: East Canary Islands as a microcontinent within the Africa-North America continental drift fit. Nature **266**, 1043–1945, 1970

Fúster, J.M., Hernandez-Pacheco, A., Muñoz, H., Peoriques Badiola, E., Garcia Cacho, L.: Geologia y volcanologia de las Islas Canarias-Gran Canaria. Madrid: Consejo Sup. Investig. Cient. 243 pp., 1968

Fúster, J.M., Paez, A., Sagredo, J.: Significance of basic and ultramafic rock inclusions in the basalts of the Canary Islands. Bull Volcanol. **33**, 665–691, 1970

Grunau, H.R., Lehner, P., Cleintuar, M.R., Allenbach, P., Bakker, G.: New radiometric ages and seismic data from Fuerteventura (Canary Islands), Maio (Cape Verde Islands) and Sao Tome (Gulf of Guinea). In: Progress in geodynamics. G.J. Borradaille et al., ed.: pp. 90–118. Amsterdam: Koninkl. Nederl. Akad. Wetensch. 1975

Hernán Reguera, F.: Estudio petrológico y estructural del complejo traquiticosienitico de Gran Canaria. Estud. Geol. Madrid **32**, 279–324, 1976

Lancelot, Y., Seibold, E.: Initial Reports of the Deep Sea Drilling Project. Vol. 41, p. 1259. Washington: U.S. Government Printing Office 1977

Lehner, P., Ruiter, P.A.C. de: Structural history of Atlantic margin of Africa. Am. Assoc. Petrol. Geol. Bull. **61**, 961–981, 1977

Lietz, J., Schmincke, H.-U.: Miocene-Pliocene sea-level changes and volcanic phases on Gran Canaria (Canary Islands) in the light of new K/Ar-ages. Palaeogeogr. Palaeoclimatol. Palaeoecol. **18**, 213–239, 1975

McDougall, I., Schmincke, H.-U.: Geochronology of Gran Canary, Canary Islands: Age of shield building volcanism and other magmatic phases. Bull. Volcanol. **40**, 1–21, 1976

Robertson, A.H.F., Stillman, C.J.: Submarine volcanic and associated sedimentary rocks of the Fuerteventura Basal Complex, Canary Islands. Geol. Mag. **116**, 203–214, 1979

Rona, P., Nalwalk, A.J.: Post-early Pliocene unconformity on Fuerteventura, Canary Islands. Geol. Soc. Am. Bull. **81**, 2117–2122, 1970

Rothe, P., Schmincke, H.-U.: Contrasting origins of the eastern and western islands of the Canarian Archipelago. Nature **218**, 1152–1154, 1968

Schmincke, H.-U.: Cone sheet swarm, resurgence of Tejeda Caldera, and the early geologic history of Gran Canaria. Bull. Volcanol. **31**, 153–162, 1967

Schmincke, H.-U.: Faulting versus erosion and the reconstruction of the Mid-Miocene shield volcano of Gran Canaria. Geol. Mitt. **8**, 23–50, 1968

Schmincke, H.-U.: The geology of the Canary Islands. In: Biogeography and Ecology in the Canary Islands, Junk, the Hague, G. Kunkel, ed.: pp. 67–184, 1976

Schmincke, H.-U., Rad, U. v.: Neogene evolution of Canary Island volcanism inferred from ash layers and volcaniclastic sandstones of DSDP site 397 (Leg 47A). In: Initial Reports of the Deep Sea Drilling, Project. Vol. 47(1), U.v. Rad et al., eds.: Washington: U.S. Government Printing Office 1979

Schmincke, H.-U., Staudigel, H.: Pillow lavas on central and eastern Atlantic Islands (La Palma, Gran Canaria, Porto Santo, Santa Maria) (preliminary report). Bull. Soc. Geol. Fr. **18**, 871–883, 1976

Stillman, C.J., Fúster, J.M., Bennell-Baker, M.J., Muñoz, M., Smewing, J.D., Sagredo, J.: Basal complex of Fuerteventura (Canary Islands) is an oceanic intrusive complex with rift-system affinities. Nature **257**, 469–471, 1975

Storetvedt, U.M., Sralestad, S., Thomassen, K., Langlie, Aa., Nergård, A., Gidskehaug, A.: Magnetic discordance in Gran Canaria/Tenerife and its possible relevance to the formation of the N.W. African continental margin. J. Geophys. **44**, 317–332, 1978

Uchupi, E., Emery, K.O., Bowin, C.O., Philipps, J.D.: Continental margin off Western Africa: Senegal to Portugal, AAPG Bull. **60**, 5, 809–878 (1976)

Watkins, N.D.: Palaeomagnetism of the Canary islands and Madeira. Geophys. J. R. Astron. Soc. **32**, 249–267, 1973

Watkins, J.S., Hoppe, K.W.: Reconaissance geology of Atlantic Margin of NW Africa, 25°–34°N. Second Maurice Ewing Memorial Symposium on 'Implications of Deep Sea Drilling Results in the North Atlantic', New York, March 22–25, 1978

Weigel, E., Goldflam, P., Hinz, K.: The crustal structure of Concepcion Bank. Marine Geophys. Res. **3**, 381–392, 1978

Wissmann, F., Hinz, K., Weigel, W.: Geophysical evidence for the nature of the continent-ocean boundary off northwest Africa. Abstract. London: Royal Society Meeting October 19–20, 1977

Received January 17, 1979; Revised Version August 8, 1979; Accepted August 10, 1979

J. Geophys. 46, 225–228, 1979

Journal of Geophysics

Short Communication

Rinneite-Dating of Episodic Events in Potash Salt Deposits

H.J. Lippolt and I. Raczek

Laboratorium für Geochronologie der Universität,
Im Neuenheimer Feld 234, D-6900 Heidelberg, Federal Republic of Germany

Introduction

When reviewing the state of isotope chronometry on salt rocks (Lippolt, 1977) we directed our attention to the salt mineral rinneite (NaK_3FeCl_6) which occurs as metasomatic secondary product in salt deposits. Rinneite is formed as well in the early phases of metamorphism (Kokorsch, 1960; Siemeister, 1969) as during late tectonic disturbances of the salt beds (Hartwig, 1922). Siemeister (1969) distinguished six types of rinneite paragenesis in bed Ronnenberg of the Salzdetfurth mine, Southern Hannover. Though the chemical formula and analyses by Kühn (1972) prove that rinneite contains plenty of potassium and rubidium, this mineral has not been used for isotopic dating so far. When clean, rinneite appears transparent, otherwise pink, yellow or violet. At open air the surface turns brown by oxidation. Rinneite is soluble in water, but as long as it is protected by the salty environment, it was shown to be stable above 26.4° C. Rinneite was discovered in 1908 in the Wolkramhausen mine south of the Harz mountain and investigated by Boeke (1908; 1909). In this mine it occurred in large, meter-sized lenses within carnallites. Later on it was found in all the mining districts around the Harz mountain and in the Hannover plain, occasionally as layer but as well as single crystals or very fine-grained in kieserite – sylvine-halite rocks. Rinneite also occurs fairly commonly in the English potash fields (Stewart, 1956).

The Samples

In order to get some experience with rinneite as dating mineral we asked Dr. R. Kühn (Hannover) for some of the samples which he had used for his review of Rb abundances in salt minerals (Kühn, 1972). Four samples were selected for Rb-Sr analyses, two of them representing vein fillings and the two others secondary components in rocks of the salt beds:

0340-062X/79/0046/0225/$01.00

1. Pure rinneite from a fissure vein in the Grey Salt Clay, 1,050 m floor of field Siegfried in the salt mine Giesen.

2. Rinneite, carnallite and clay from a patchy filling of a vein in the Grey Salt Clay T3, 1,050 m floor, field Rössing-Barnten, mine Giesen.

3. Pure rinneite clusters in seam Staßfurt of the mine Hildesia/Diekholzen (corresponding with sample 1 of Table 5 in Kühn, 1972)

4. Pure rinneite from rinneite patches above the upper part of the carnallite seam below sylvine bed Ronnenberg (corresponding with sample 2 of Table 5 in Kühn, 1972), Salzdetfurth mine.

The salt mine Siegfried-Giesen belongs to the Sarstedt-Lehrte diapir (mining district Central Hannover), Hildesia and Salzdetfurth are situated in the Hildesheim-Wald salt structure (district Southern Hannover). Both districts are about 20 to 40 km SE of the city of Hannover. The geologic and petrologic settings of these salt mines have been described by Middendorf and Kühn (1966) and by Kokorsch (1960) and Siemeister (1969) respectively. Jaritz (1973) compiled the observations concerning the ages of diapir formation in this area.

The Rb-Sr Analyses

The results of our analyses are presented in Table 1. Six Rb-Sr and K analyses have been performed. Rb and Sr were measured by mass spectrometry, potassium by flame photometry. Calibrations were achieved by comparison with international chemical standard samples. The potassium concentrations measured are typical for rinneite. Our Rb-results on rinneite III and IV agree well with the values given by Kühn (1972). Sr had not been measured so far. In three cases its concentration was about 1 ppm, in one case we found

Table 1. Analytical results on rinneite samples

Sample	Salt mine	Concentrations			$^{87}Rb/^{86}Sr$	$^{87}Sr/^{86}Sr$	Model ages[b] [Ma]
		K (%)	Rb (ppm)[a]	Sr (ppm)[a]			
I. Rinneite	Siegfried/ Giesen	28.2	257	0.79	943	1.125	31.0 ± 0.8
II. Rinneite	Siegfried/ Giesen	29.3	218	6.54	96.7	0.735	19.8 ± 0.6
II. Carnallite (+Rinneite)	Siegfried/ Giesen	20.4	207	3.95	152	0.726	8.5 ± 0.3
II. Clay	Siegfried/ Giesen	0.65	12	626	0.054	0.7077	–
III. Rinneite	Hildesia	30.6	131	1.18	322	1.091	83.4 ± 1.8
IV. Rinneite	Salzdetfurth	28.7	128	0.57	645	1.373	72.2 ± 1.8

[a] Analytical errors $\leqq 1\%$

[b] Using $\lambda = 1.42 \cdot 10^{-11} a^{-1}$ [Steiger and Jäger (1977)] and 0.7077 as initial ratio

6.5 ppm. This results in fairly high Rb/Sr ratios of about 30 to 300. For the time being only model ages can be calculated as only for rinneite II an approximative initial $^{87}Sr/^{86}Sr$ ratio could be determined (clay from sample II). We consider this value to be tentative because nothing is known about homogenization of Sr isotopes between salt minerals and clay during salt rock metamorphism. This lack of information however is insignificant as the Rb/Sr ratios and the present $^{87}Sr/^{86}Sr$ ratios are fairly high. The analytical accuracy is given in the table. It seems to be better than necessary in order to come to geologically meaningful conclusions taking in account that the present level of knowledge on the feasibility of salt rock dating is still very low.

Discussion of the Rb-Sr Ages

The model ages of the four rinneites are Cretaceous and Middle Tertiary. The four dates differ significantly from eath other. As rinneite normally is thought to be a replacement mineral we have not expected them to yield Permian ages. Assuming appropriate retentivity we expected their ages to be equal to or lower than the time of the last regional metamorphism of these deposits. After Kokorsch (1960) and Siemeister (1969) this may have happened in Jurassic or Lower Cretaceous times. The isotopic ages of products of this metamorphism which have been determined so far rather date this metamorphism at the boundary Lower to Upper Cretaceous (K-Arges, langbeinite ages by Osterle and Lippolt (1976) and Trümmer-carnallite age by Büchler et al., 1979). Rather high ages which correspond to Upper Cretaceous times (83 and 72 m.y.) were found for the rinneites of the Hildesheim-Wald salt ridge. These ages obviously confirm the view of Kokorsch (1960) and Siemeister (1969) that the rinneite of this paragenesis was formed during the early stages of the late mesozoic metamorphism. Relatively low Rb-Sr ages were found for the rinneites from vein fillings close to the Grey Salt Clay in the Sarstedt salt structure. They correspond to Upper Oligocene (31 m.y.) and Lower Miocene times (20 m.y.). The geological situation from which the rinneites I and II were taken is very different from that of rinneite III and IV. Already Hartwig (1922) and later Kühn (1966) made it clear that the rinneite associated with vein fillings in the neighbourhood of the Grey Salt Clay was formed as late replacements following tectonic disturbances after the rise of the diapirs. We therefore conclude that also in this case the rinneite ages point to a geological event in the history of the salt dome.

For the time being we can only speculate if the age difference between the members of the Cretaceous and the Tertiary rinneite pairs could have a geological meaning. As Oesterle and Lippolt (1976) observed that langbeinite ages of different salt seams within one mine show systematic age differences, it might be possible that the rinneite ages III and IV indicate that the time of metamorphism for the potash seams Staßfurt and Ronnenberg was not the same. The differing ages of rinneites I and II could mean that they date tectonic activities in central Germany just before and after the main volcanic phase in many volcanic centers of the Rhenish shield ($\sim$25–23 m.y.).

There can be no doubt that the feasibility of rinneite dating and the speculations mentioned above have to be checked by further analyses. In order to do this we have already taken more samples from the two salt structures Hildesheimer-Wald and Sarstedt-Lehrte.

The carnallite associated with the rinneite II yielded a much younger Rb-Sr age (8.5 m.y.). At present we do not know whether this means that the carnallite was formed later than the rinneite or that it lost radiogenic ^{87}Sr after its deposition.

Acknowledgement. We thank Professor Kühn (now Wilhelmsfeld/Odenwald) for providing the samples and advice and H. Funke for technical assistance.

References

Boeke, H.E.: Rinneit, ein neugefundenes eisenchlorürhaltiges Salz-Mineral. Kali, Zeitschrift für die Gewinnung, Verarbeitung und Verwertung der Kalisalze, 2. Jahrgang, Heft **23**, 514–515, 1908

Boeke, H.E.: Das Rinneitvorkommen von Wolkramshausen am Südharz, ein Beitrag zur Salzpetrographie. Neues Jahrb. Mineral. Geol. Palaeontol. Band **II**, 19–56, 1909

Büchler, K., Kirsten, T., Ries, D.: Edelgasisotope in Kalisalzen des Zechsteins, Datierung und Eignungsuntersuchung als Target für ein Solar-Neutrino-Experiment. Abstrakt Cl-14, Tagung der Deutschen Geophysikalischen Gesellschaft in Kiel 1979, p. 154, 1979

Hartwig, G.: Stratigraphie und Tektonik des Kaliwerks ‚Siegfried-Giesen' bei Hildesheim. In: Kalifestschrift zum 70. Geburtstage von Heinrich Precht, F. Schöndorf, ed.: pp. 1–83, Hannover 1922

Jaritz, W.: Zur Entstehung der Salzstrukturen Nordwestdeutschlands. Geol. Jahrb. Reihe A: **10**, 77, 1973

Kokorsch, R.: Zur Kenntnis von Genesis, Metamorphose und Faciesverhältnissen des Staßfurtlagers im Grubenfeld Hildesia-Mathildenhall, Diekholzen bei Hildesheim. Beih. Geol. Jahrb. **41**, 140, 1960

Kühn, R.: Zu den mineralogisch-petrographischen Verhältnissen des Kalibergwerks Siegfried-Giesen, Groß-Giesen. Fortschr. Mineral. **43**, 153–187, 1966

Kühn, R.: Zur Kenntnis der Rubidiumgehalte von Kalisalzen ozeanischer Salzlagerstätten nebst einigen lagerstättenkundlichen Ausdeutungen. Geol. Jahrb. **90**, 127–220, 1972

Lippolt, H.J.: Isotopische Salz-Datierung: Deutung und Bedeutung. Der Aufschluß **28**, 369–389, 1977

Middendorf, E., Kühn, R.: Befahrung des Kalibergwerks Siegfried-Giesen, Groß-Giesen bei Hildesheim, 8. September 1965. Fortschr. Mineral. **43**, 145–187, 1966

Oesterle, F.-P., Lippolt, H.J.: Langbeinitdatierungen und Kalisalzmetamorphosen. Nachr. Dtsch. Geol. Ges. **15**, 22–23, 1976

Siemeister, G.: Primärparagenese und Metamorphose des Ronnenberg-Lagers nach Untersuchungen im Grubenfeld Salzdetfurth. Beih. Geol. Jahrb. **62**, 122 pp. 1969

Steiger, R.H., Jäger, E.: Subcommision on Geochronology: Convention on the use of decay constants in geo- and cosmochronology. Earth Planet. Sci. Lett. **36**, 359–362, 1977

Stewart, F.H.: Replacements involving early carnallite in the potassium-bearing evaporites of Yorkshire. Mineral. Mag. **31**, 127–135, 1956

Received May 22, 1979; Revised Version July 17, 1979; Accepted August 10, 1979

J. Geophys. 46, 229–236, 1979

Journal of Geophysics

Magnetic ULF-Waves in the Vicinity of Active Auroral Forms

N. Klöcker and B. Theile

Institut für Geophysik und Meteorologie der Technischen Universität Braunschweig, Mendelssohnstraße 1A, D-3300 Braunschweig, Federal Republic of Germany

Abstract. Magnetic field variations in the frequency range one to 5 Hz were measured with a rocket-borne magnetometer over an auroral arc during a substrom event. There is evidence that these fluctuations were produced by magnetohydrodynamic waves generated by ion gyroresonance. The payload was launched on October 13, 1977, from Andenes/Norway around magnetic midnight.

Key words: ULF-waves – Magnetohydrodynamic waves – Gyroresonance.

1. Introduction

In the frequency range 0.2–5 Hz of the electromagnetic spectrum a multitude of phenomena is observed on and near earth, e.g., pearls in geomagnetic micropulsations, auroral flickering, and microbursts in particle precipitation. Electric ULF waves in this frequency band have been detected within and above the ionospheric F-layer (Kintner and Cahill, 1978; Petelski et al., 1978). Recently published observations by the team of S-300 experimenters (1979) of the GEOS satellite show evidence of simultaneous magnetic and electric field fluctuations in the ULF frequency domain. An appropriate approach to explain these waves in terms of magnetohydrodynamics is to identify them as being either ion cyclotron or fast magnetosonic waves. These wave modes are believed to play an important role in wave-particle interaction (Cornwall, 1965; Cornwall, 1966; Kennel and Petschek, 1966; Cornwall et al., 1970).

Measurements of the magnetic field of these waves are still rare. The reason is twofold: (1) The spin frequency of sounding rockets and satellites is very often in the range of ULF band. Thus the waves can only be detected after very accurate restitution of the spacecraft's attitude in an earth-fixed frame of reference. (2) Most magnetometers flown on sounding rockets and near-earth satellites are not well suited to measure waves in the ULF band, e.g., due to limited dynamical range and limited time and amplitude resolution.

0340-062X/79/0046/0229/$01.60

We will present data which were measured with a three-component fluxgate magnetometer aboard a sounding rocket payload within the German IMS sounding rocket campaign 'substorm-phenomena'. The payload was equipped with a gyro based attitude control system. After a short outline of our instrument we describe the observations and rocket trajectory. The measurements will be introduced by an amplitude-time diagramme. We will then deal with the analysis of the waveforms and finally discuss some theoretical aspects.

2. Instrumentation

A three-component fluxgate magnetometer (Förstersonde type) is used. Full details of the instrument and the test philosophy were published elsewhere (Theile and Lühr, 1976). The instrument consists of an orthogonal sensor triple, an analogue and a digital electronic. The analogue electronic drives the sensor and transforms the returning signal into a voltage proportional to the field strength in the direction of the respective sensor axis. The dynamic range of the instrument's analogue output is 120,000 nT, the noise level is about 0.1 nT. A 16 bit analog to digital converter (ADC) is used to transform the output voltage into a digital telemetry compatible word. The input voltage of the ADC may only correspond to $\pm 30{,}000\,nT$ for the radially measured field components and 0–$60{,}000\,nT$ for the longitudinal (i.e., parallel to the spin axis) field component. This instrumental configuration allows the measurement of magnetic fields with a resolution of about 1 nT. The instrument responds to frequencies up to 500 Hz. There are no phase shifts and linearity errors for signals below 15 Hz. This instrumental property facilitates the analysis of the ULF waves.

The sampling rate amounts to 800 samples per channel per second. This high data rate is needed to compute highly accurate angles for the transformation of the magnetic field data into an earth fixed coordinate system.

3. Observations

The payload was launched on October 13, 1977, at 21 : 26 : 00 UT by a Skylark 12 from Andoya Rocket Range/Norway. A magnetospheric substorm had started at about 20 : 25 UT and continued until about 23 : 30 UT. The magnetic north component of the range magnetometer fluctuated around $-300\,nT$ during the payload's flight (21 : 26 – 21 : 38 UT). The payload reached an apogee of 536 km and splashed down at 540 km ground range distance.

A faint homogeneous arc was observed north of the range prior to launch. The launch decision was based on increasing micropulsation activity and the brightening of the arc. The intensity of the 557.7 nm auroral emission intensified to more than 50 kR shortly after launch. An auroral break-up occured at 21 : 31 UT followed by a rapid northward expansion of the aurora. Micropulsation activity prevailed throughout the flight. However, the pulsation magnetometer was not sensitive in the pc 1 range. The wave phenomena which will be treated here occurred around 21 : 35 UT. At that time the payload traversed an

Table 1. Velocities with respect to local geomagnetic field and spin frequencies for four selected times

Time after launch (s)	$V_\perp$ (ms^{-1})	$V_\parallel$ (ms^{-1})	Spinfrequency (Hz)
534.9	525	1,282	3.0033
538.1	519	1,307	3.0012
545.1	504	1,365	2.9921
549.1	496	1,398	2.9838

intensity maximum of an active auroral arc. More than 20 kR were measured for the λ557.7 nm oxygen line.

The measurements discussed here were made at 450 km altitude. The payload's velocity components parallel and perpendicular with respect to the earth's magnetic field as well as the spin frequency are listed in Table 1 for four selected times. The decrease of the spin frequency coincides with an increase of the payload's coning angle. This change of attitude parameters is caused by energy dissipation by long flexible booms tipped with probes for the electric field measurement. An activation of the attitude control system was necessary every 60 s.

Figure 1 shows a section of the original magnetometer data after transformation into an earth fixed coordinate system and deduction of a reference magnetic field as a function of time. The indicated times are seconds after lift-off. B_x points towards the magnetic north, B_y indicates the eastern direction, whereas B_z is parallel to the measured undisturbed geomagnetic field, i.e., pointing downwards with an angle of about 12 degrees with respect to the vertical.

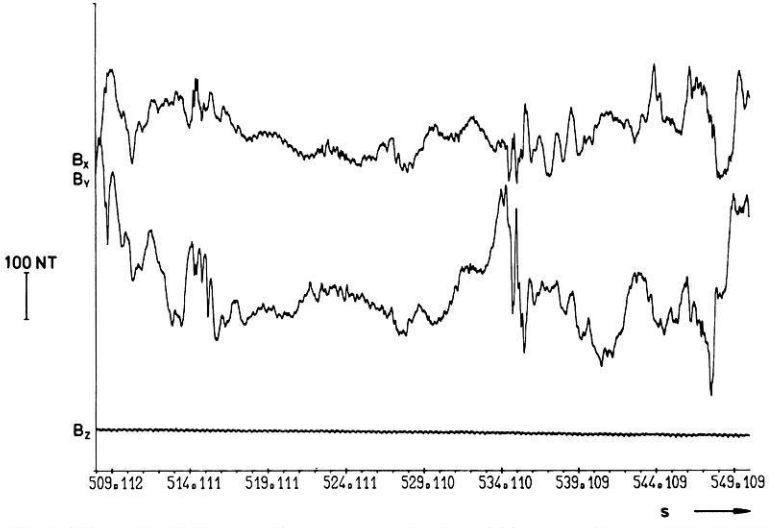


Fig. 1. Magnetic field variations measured aboard the sounding rocket payload (Substormphenomena F3A) on October 13, 1977. B_x-north, B_y-east, B_z-direction of earth's magnetic field vector. The abscissa gives seconds after launch which took place at 21:26:00 UT

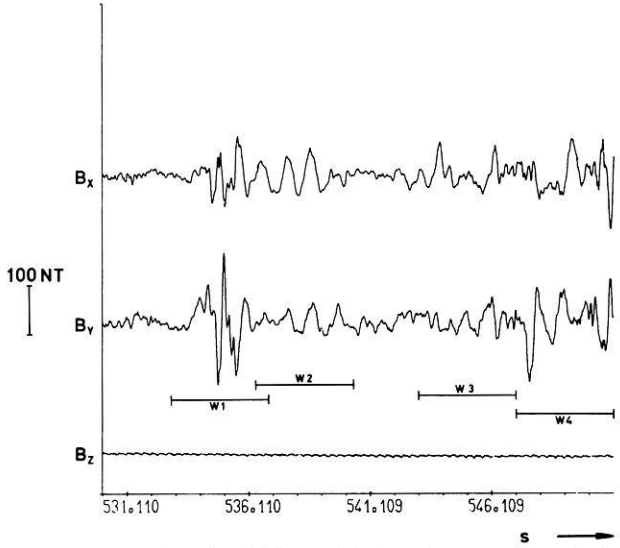


Fig. 2. Data section after high-pass filtering with cut-off at 0.65 Hz was applied. The windows $W1$–$W4$ mark those intervalls which were Fourier-transformed for spectral analysis

The transformation is carried out by a mere rotation of the sensor-based coordinate system. The reference directions for the transformation angles are established by the measured geomagnetic field and the direction of the payload's total angular momentum. The latter is computed using attitude data of the gyro based attitude control system. The payload's rotational motions are being represented by a set of time dependent analytical functions. These functions are coupled by gyroscopic laws. A new set of initial conditions must be determined after each activation of the attitude control system. Thus each interval between two attitude control system activations requires new parameters for the analytical description of the rotational motions. Consequently only magnetic field variations with periods below 60 s (i.e., the time interval between two attitude control system activations) can be analysed.

Field variations with different amplitudes and periods can be identified in B_x and B_y. Variations in B_z cannot be observed. The conclusion of this observational evidence is that the longitudinal (i.e., parallel to the earth's magnetic field) component of the field variation must be below $1\,nT$. With an order of magnitude of $100\,nT$ for the transverse component we deduce by statistical analysis that the angle of the field variation vector must be between 89.5 and 90.5 degrees with respect to the local geometric field. The fluctuations with longer periods ($T \sim 40$ s) seen in Fig. 1 may be of temporal or spatial nature. They will not be discussed here.

Further analysis and discussion will be dedicated to variations at frequencies of more than 1 Hz. Magnetic variations transverse to the main field may be due to either field-aligned currents or to magnetohydrodynamic waves. The current density of a field-aligned current sheet can be computed by using the first Maxwell equation. Assuming an east-west current sheet of infinite extent and a payload displacement towards north (x-direction), we find for the current

density $j=1/\mu_0 \cdot dB_y/dx$. The numerical evaluation results in a current density of $10^{-3}\,\mathrm{Am}^{-2}$, a value of unrealistic high magnitude. Thus we consider these variations as being due to waves.

The data are prepared for further analysis by being high pass filtered. The cut-off frequency is 0.65 Hz. The result is shown in Fig. 2. Some periodic structures can be identified.

4. Analysis

Further analysis will be carried out by Fourier transformation of selected time intervals. These intervals are marked in Fig. 2 by bars and are identified by the designations $W1$ through $W4$. The waves can be dissolved into two circularly polarized waves with opposite senses of rotation.
Considering only the transverse field components we write

$$B_x = B_{x_0} e^{i(\omega t + \phi_x)} \tag{1}$$

$$B_y = B_{y_0} e^{i(\omega t + \phi_y)} \tag{2}$$

The right hand polarization is then given by

$$B_R = \tfrac{1}{2}(B_x + iB_y) \tag{3}$$

and the left hand mode by

$$B_L = \tfrac{1}{2}(B_x^* + iB_y^*) \tag{4}$$

where the asterisk stands for the complex conjugate.

The windows $W1$ through $W4$ are bilaterally smoothed by a cosine function:

$$\omega(t) = \begin{cases} \dfrac{1}{2T}\left(1+\cos\dfrac{\pi}{(1-a)T}t\right) & aT \leqq |t| \leqq T \\ \dfrac{1}{T} & -aT < t < aT \end{cases} \tag{5}$$

The main lobe of the transformed window $W(t)$ has an effective width of 0.4 Hz ($T=2$ s, $a=0.4$). The first side lobe shows a maximum transfer ratio of 0.3 at 0.48 Hz.

The instrument's noise is $1\,nT$. The spectral estimates were optimized by varying the window length. Spectral peaks protruding by more than $2\,nT$ must be considered to be real signals. However, there is one exemption in the left-hand polarized spectrum of Fig. 4: the peak at 3 Hz is caused by a not completely corrected misalignment of a sensor axis.

Figure 3 shows the spectra of the right-hand (r-mode) and the left-hand (1-mode) polarized waves belonging to window $W1$. The r-mode spectrum is characterized by two maxima at 0.5 and 1.3 Hz which do not appear in the l-mode spectrum. At frequencies higher than 2 Hz the l- and r-mode spectra are almost the same which means that the wave is linearly polarized.

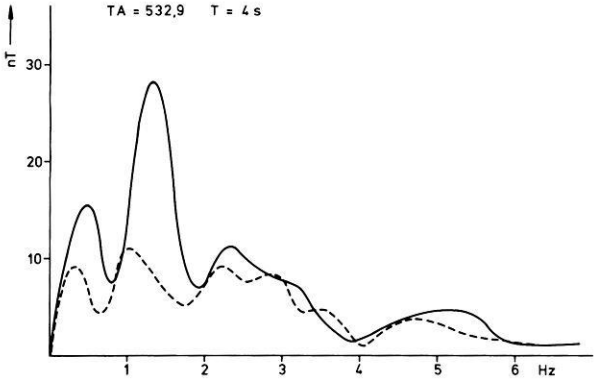


Fig. 3. Spectrum of window $W1$ for the right-hand polarized (*solid line*) and the left-hand polarized (*dashed line*) wave. The statistical significance is discussed in Sect. 4

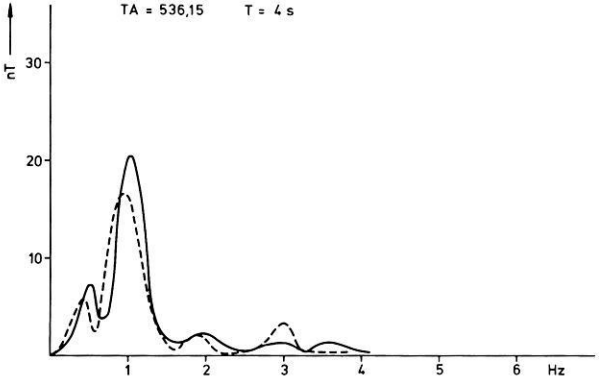


Fig. 4. Spectrum of window $W2$

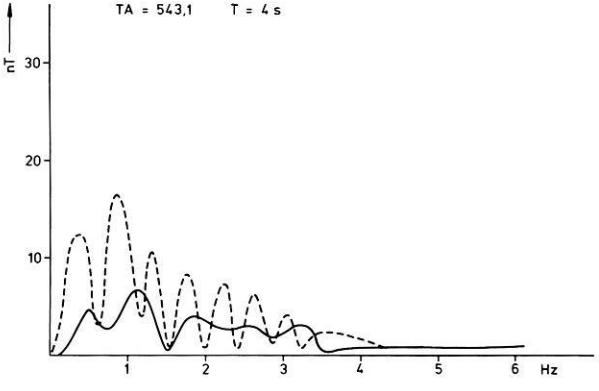


Fig. 5. Spectrum of windows $W3$

It should be mentioned, that a field-aligned proton flux at 65 keV was observed between 534 and 536 s after launch (W. Stüdemann, private communication), that is within this window. The continuation is shown by the second window $W2$ in Fig. 4. The wave becomes nearly linearly polarized and almost monochromatic at 1 Hz.

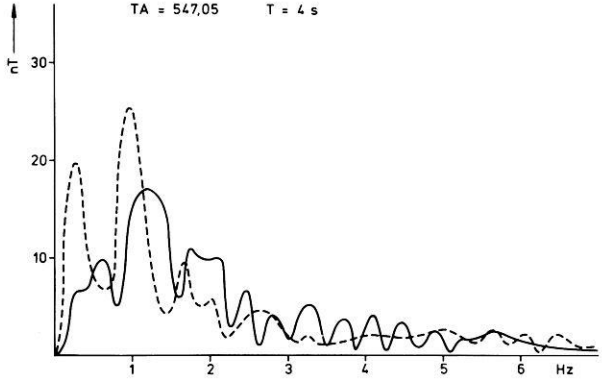


Fig. 6. Spectrum of window $W4$

The spectra of windows $W3$ and $W4$ are shown in Figs. 5 and 6. The l-mode-spectrum of Fig. 5 is characterised by peaks at integer multiples of 0.437 Hz. The r-mode-spectrum is not well defined. Figure 6 shows multiples of 0.4 Hz in the r-mode-spectrum. These multiples can be clearly identified above 2 Hz.

A common feature of all spectra is a shift of corresponding maxima in the l- and r-mode-spectra. The spectrum of $W2$ (Fig. 4) shows a shift of 0.1 Hz with all three peaks. The right-mode maximum is always shifted towards the higher frequency. This phenomena can be explained by a 0.05 Hz clockwise rotation of the ambient plasma with respect to an earth fixed coordinate system.

5. Discussion

The plasma of the inner magnetosphere is characterized by a very low ratio of the thermal velocity to the Alfén velocity. Thus, with $v_{th}^2/v_A^2=\beta\ll 1$ we find two wave modes being candidates to explain our observations: The anisotropic Alfvén mode for ions (left-hand polarization) and the fast magnetosonic mode (right-hand polarization). The instable growth mechanism of these modes has been investigated by Cornwall (1965) as well as by Kennel and Petschek (1966).

In order to explain the waves measured between 532 and 540 s after launch, we assume a cause and effect relationship between the above quoted field-aligned ion beam at 536 s and the wave phenomena.

The observations within windows $W1$ and $W2$ together with the field-aligned protons are consistent with Kennel and Petschek's (1966) conditions for the magnetosonic mode near ion gyroresonance. The resonant frequency of a gyroresonance interaction is determined by the ratio of the parallel energy (E_R) to the magnetic energy (E_C) per particle. This ratio is determined by

$$\frac{E_R}{E_C}=\left(\frac{\omega_i}{\omega}\right)^2\left(1+\frac{\omega}{\omega_i}\right)^3 \tag{6}$$

ω_i – Ion gyrofrequency
ω – Wave frequency

$E_R = \frac{1}{2} m_i v_R^2$ parallel energy
$E_C = B^2/2\mu_0 N$ magnetic energy
m_i – Ion mass
v_R – Resonant velocity
B – Magnitude of magnetic field
N – Number density.

According to Kennel and Petschek (1966) the waves can only be generated near the equator, because E_C grows fast away from the equator. We find for $L \gtrsim 7$ where the measurements were taken $\omega_i = 9\,\mathrm{rad \cdot s^{-1}}$ and $E_C \approx 10\,\mathrm{keV}$. From our measurements we get $\omega = 2\pi f = 8\,\mathrm{rad\,s^{-1}}$. Thus, $E_R/E_C = 8$, which is consistent with the particle observations (Stüdemann, internal report, 1979). Protons are the only candidates for this resonance interaction as for heavier ions $\omega_i > \omega$ would be no longer valid.

We do not have an adequate explanation for a possible relationship of the spectra shown in Figs. 5 and 6. Higher harmonics can be excited, if the wave vector is inclined against the magnetic field (Stix, 1962). However, these waves ought to be elliptically polarized, which is not observed in our case.

Acknowledgements. We thank Professor Dr. W. Kertz, Direktor des Institutes für Geophysik und Meteorologie der TU Braunschweig, for many fruitful discussions and steady support of the project. H. Lühr performed most of the hardware tests. Dr. Wilhelm, Max-Planck-Institut für Aeronomie in Lindau, acted as project scientist. The project was financed by Bundesministerium für Forschung und Technologie and managed by Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt.

References

Cornwall, J.M.: Cyclotron instabilities and electromagnetic emission in the ultra low frequency and very low frequency ranges. J. Geophys. Res. **70**, 61–69, 1965

Cornwall, J.M.: Micropulsations and the outer radiation zone. J. Geophys. Res. **71**, 2185–2199, 1966

Cornwall, J.M., Coroniti, F.V., Thorne, R.M.: Turbulent loss of ring current protons. J. Geophys. Res. **75**, 4699–4709, 1970

Kennel, C.F., Petschek, H.E.: Limit on stably trapped particle fluxes. J. Geophys. Res. **71**, 1–28, 1966

Kintner, P.M., Cahill, L.J.: Electric field oscillations measured near an auroral arc. Planet. Space Sci. **26**, 555–558, 1978

Petelski, E.F., Fahleson, U., Shawhan, S.D.: Models for quasiperiodic electric fields and associated electron precipitation in the auroral zone. J. Geophys. Res. **83**, 2489–2498, 1978

S-300 Experimenters: Measurements of electric and magnetic wave fields and of cold plasma parameters on-bord GEOS-1. Preliminary results. Planet. Space Sci. **27**, 317–339, 1979

Stix, T.H.: The theory of plasma waves. New York, San Francisco, Toronto, London: McGraw-Hill Book Company, Inc. 1962

Theile, B., Lühr, H.: Magnetfeldmessungen an Bord von Höhenforschungsraketen. Raumfahrtforschung **20**, 301–305, 1976

Received May 29, 1979; Revised Version September 19, 1979; Accepted September 22, 1979

J. Geophys. 46, 237–247, 1979

Journal of Geophysics

Bursts of Irregular Magnetic Pulsations During the Substorm

J. Kangas[1], T. Pikkarainen[1], Yu. Golikov[2], L. Baransky[2], V.A. Troitskaya[2], and V. Sterlikova[2]

[1] University of Oulu, Department of Physcis, SF-90570 Oulu 57, Finland
[2] Academy of Sciences, Institute of Physics of the Earth, Moscow, USSR

Abstract. New characteristics of PiB magnetic pulsations which are associated with the substorm in the local midnight sector are revealed from the recordings of a meridional network of ground stations. PiB's are observed during distortions of the auroral arc which enclose a local region of upward directed field-aligned current. An enhancement at about 0.3–0.4 Hz is shown to be a permanent feature of PiB's. In addition, a secondary amplitude maximum at about 1 Hz sometimes appears in the recordings far from the auroral zone. The generation of PiB pulsations in the Pi1 period range is located at auroral latitudes at a low altitude. Low-altitude electric fields and acceleration of auroral particles, low-altitude resonance cavities and isotropic propagation of waves is discussed in the interpretation of PiB pulsations.

Key words: Magnetospheric substorm – PiB magnetic pulsations.

Introduction

Magnetic pulsations which are observed during the magnetospheric substorm near the midnight sector are termed Pi pulsations. Pulsations with a period less than 40 s are called Pi1 pulsations and those of a longer period Pi2 pulsations. For irregular pulsations $T > 150$ s a term Pi3 has been adopted recently (see general reviews by Saito, 1969; and Nishida, 1978).

Pi pulsations are characterized by a wide range a frequencies in their dynamic spectra. However, it appears useful to have a further classification, and the terms like PiB (Pi burst) and PiC (Pi continuous) have been adopted (Heacock, 1967). IPDP pulsations (interval of pulsations of diminishing periods) should also be discussed together with Pi pulsations due to their close association with the substorm although the dynamic spectrum of an IPDP event in some respects resembles the spectra of more regular Pc pulsations (see Saito, 1969; Nishida, 1978).

During the development of a typical substorm a PiB event appears at the onset of the substorm around the midnight sector. This burst lasts only some

0340-062X/79/0046/0237/$02.20

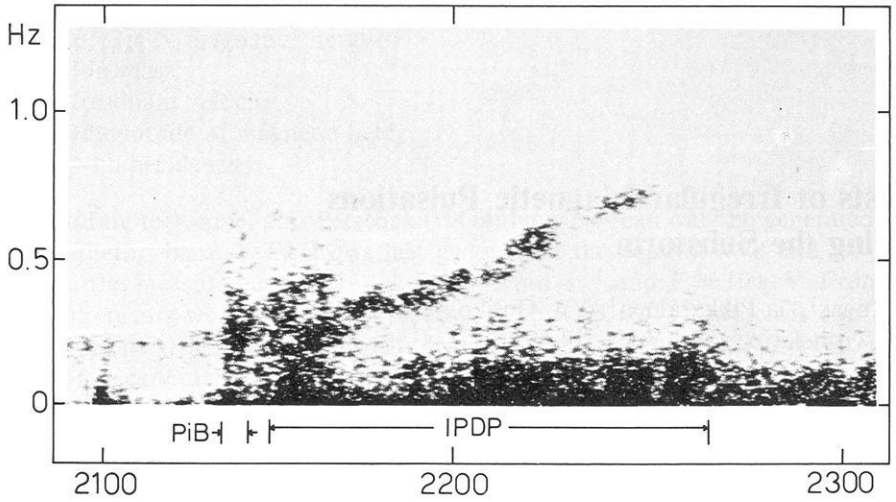


Fig. 1. Dynamic spectrum of magnetic pulsations at Sodankylä on October 5, 1974, illustrating the appearance of PiB and IPDP events

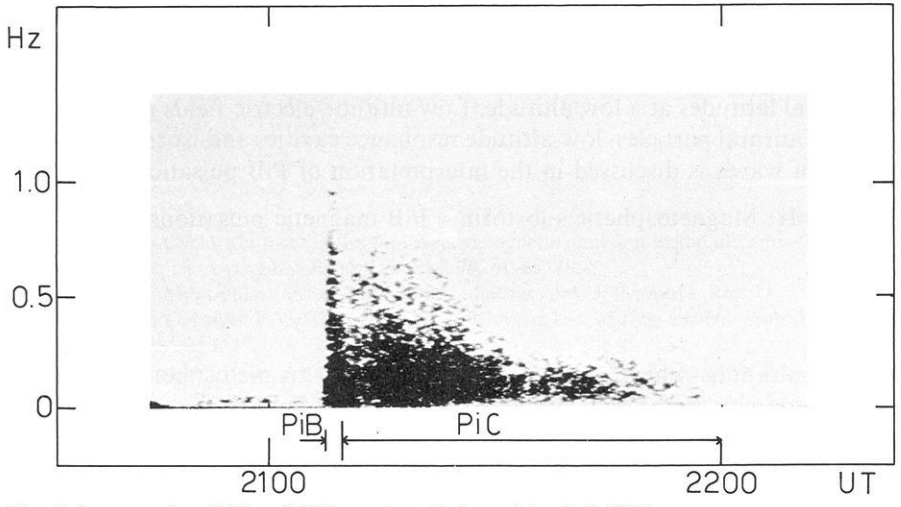


Fig. 2. An example of PiB and PiC events at Oulu on March 2, 1974

minutes and is followed by an IPDP event on the evening side of the auroral zone and by a PiC event on the morning side. Near the local midnight the dynamic spectrum is usually characterized by the combinations PiB+IPDP or PiB+PiC as shown in Figs. 1 and 2.

A PiB event appears in a dynamic spectrum as a short vertical line as shown in Figs. 1 and 2, extending sometimes up to several Hz. This means that a PiB contains both Pi1 and Pi2 pulsations. Heacock (1967) reports that these ranges may be separated partly especially in summer time when an enhance-

ment at about 0.3 Hz is often seen. Such an enhancement has also been observed at the ATS-1 geostationary location (Coleman and McPherron, 1976) but at a little higher frequency than simultaneously on the ground.

In this paper, data from several stations are used in order to study the latitudinal profile of PiB events in the Pi1 period range. Some other simultaneous observations have also been analysed in order to relate the generation of PiB magnetic pulsations with the magnetospheric processes during the substorm.

Observations and Results

Spectral Analysis of PiB Events

In October 1974, several pulsation magnetometers have been operated in the Soviet Union at about the same meridian with the Finnish stations. Later during some IMS periods similar measurements have been made. The locations of these stations are given in Table 1. Core type induction magnetometers with similar frequency characteristics in the Pi1 period range at all stations have been used.

PiB pulsation events have been identified from dynamic spectra (sonagrams) which give the frequency of pulsations. To get more information of the wave amplitudes, band-pass filters have been used. The frequency band from about 0.2 Hz to about 1.5 Hz has been analysed in this way. Some spectra will be shown below.

All the events presented here have been measured during strong magnetic activity. Due to difficulties to separate overlapping substorms no attempt has been made to correlate PiB's more exactly with the development of the substorm. On October 18 and 20, 1974, several PiB's occurred as shown by the dynamic spectra from Sodankylä in Figs. 3 and 4. Amplitude spectra have been analysed for the events at 17.55, 18.35, and 18.45 UT on October 18 and at 17.20 UT on October 20. The maximum amplitude at about 15 frequency bands has been determined. For comparison, spectra from one auroral zone station (Sodankylä) and from one midlatitude station (Nurmijärvi) have been presented in Figs. 5–8. Some main characteristics can be summarized as follows.

Table 1. Locations of Finnish and Russian pulsation magnetometer stations

	L-value	Geomagn. coord. (corrected)	
		Latitude	Longitude
Kevo	6.0	66.2°	111°
Sodankylä	5.1	63.9°	109°
Oulu	4.3	61.8°	107°
Kostamuska	4.1	60.5°	110°
Suckozero	3.8	59.0°	111°
Nurmijärvi	3.3	57.0°	103°

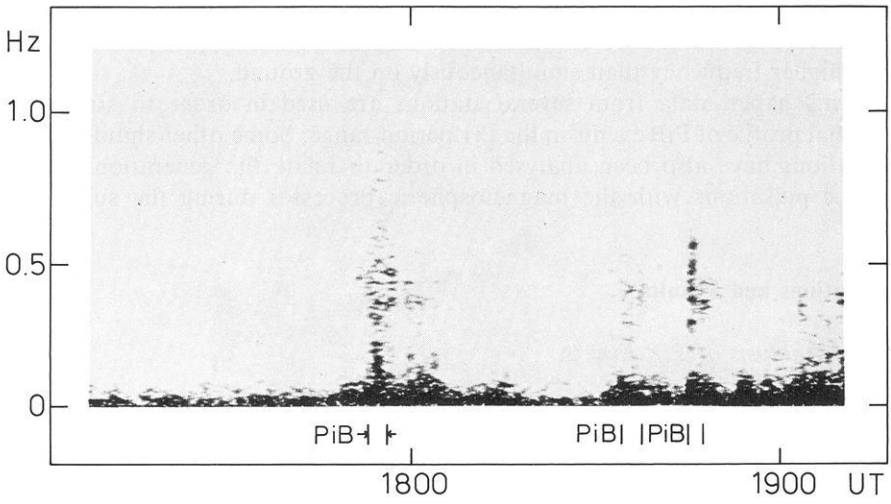


Fig. 3. Dynamic spectrum of Pi pulsations at Sodankylä on October 18, 1974. PiB events at 17.55, 18.35, and 18.45 UT are denoted

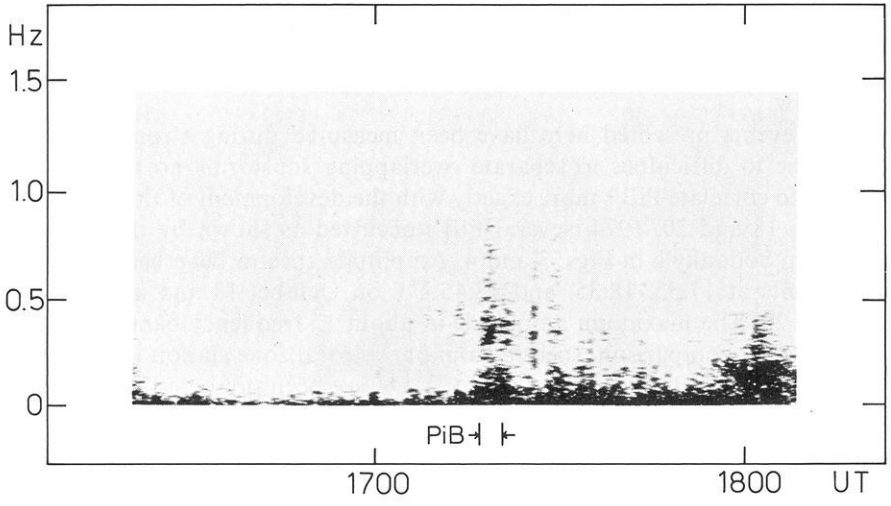


Fig. 4. Dynamic spectrum of Pi pulsations at Sodankylä on October 20, 1974. A PiB event at 17.20 UT is denoted

1. An intense peak at about 0.3–0.4 Hz appears in the high-latitude spectra.
2. A smaller peak at about the same frequency is always seen in the low-latitude spectra.
3. A secondary maximum at about 1 Hz sometimes appears in the low-latitude spectra. The amplitude of these high-frequency pulsations may be even greater than that in the auroral zone.

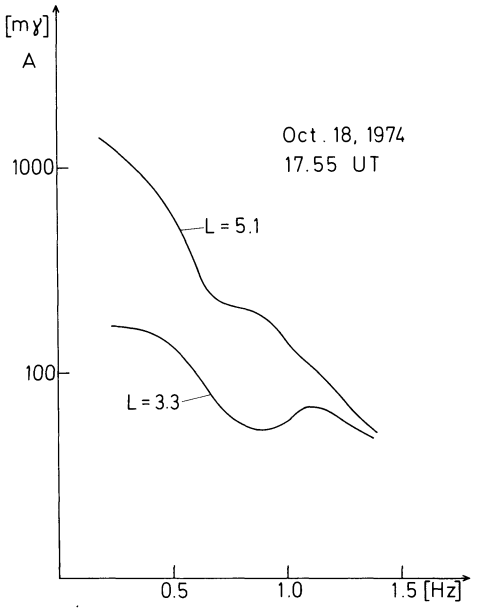


Fig. 5. Amplitude spectra of the PiB event on October 18, 1974 at 17.55 UT for the station Sodankylä ($L \sim 5.1$) and Nurmijärvi ($L \sim 3.3$)

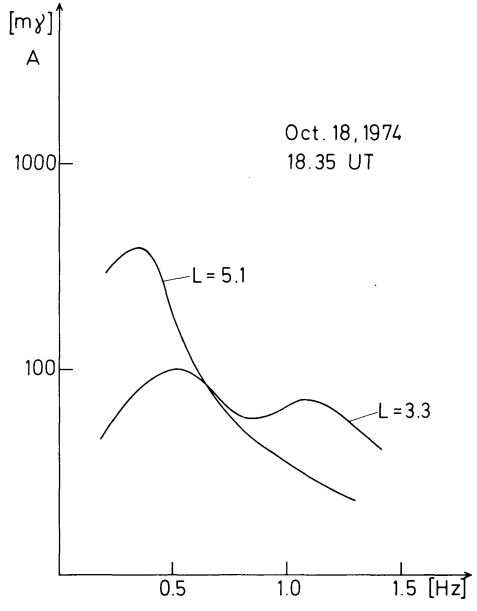


Fig. 6. Same as Fig. 5, but for the PiB event on October 18, 1974, at 18.35 UT

Dynamic spectra show a minimum of the intensity of pulsations at about 0.2 Hz. At lower frequencies Pi2 activity is dominating. This enhancement is seen clearly only in one amplitude spectrum (Fig. 7) because the method of analysis used here does not permit to continue the analysis down to low frequencies with a sufficient resolution.

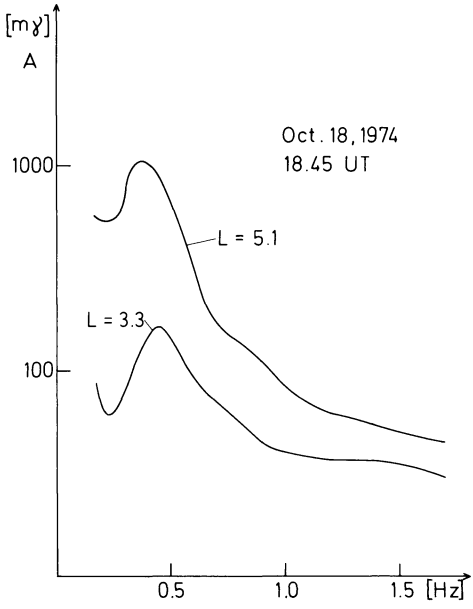


Fig. 7. Same as Fig. 5, but for the PiB event on October 18, 1974, at 18.45 UT

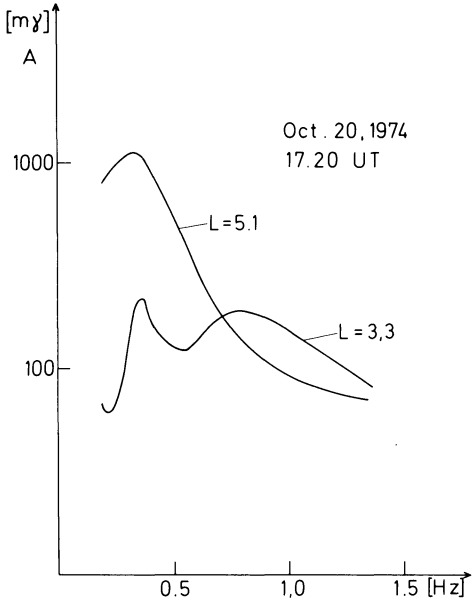


Fig. 8. Same as Fig. 5, but for the PiB event on October 20, 1974 at 17.20 UT

As a further example of the latitudinal behaviour of PiB's an event on December 1, 1977, at 20.30 UT (see Fig. 9) is presented in Fig. 10. The amplitude of magnetic pulsations at six frequency bands characterized by the center frequency has been plotted as a function of latitude. Frequencies from 0.2 to 1 Hz have been used. The same features as above can easily be noted. The maximum amplitude at low frequencies has been observed at high-latitude station

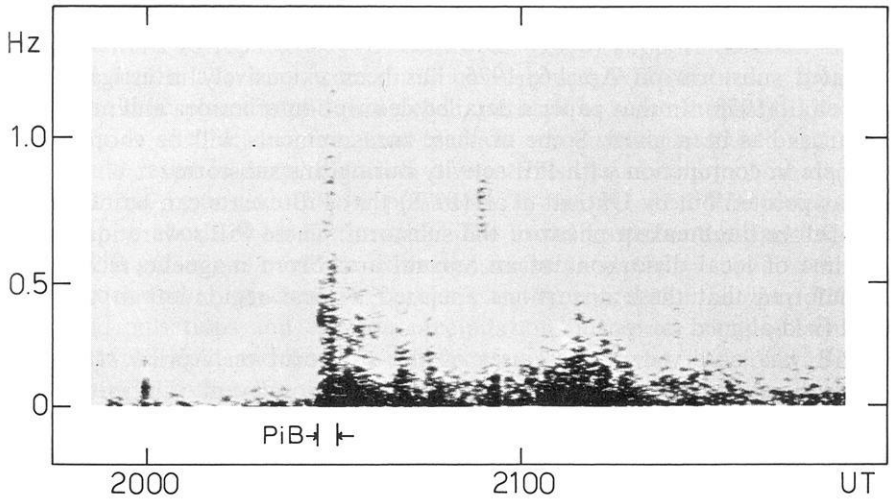


Fig. 9. Dynamic spectrum of Pi pulsations at Sodankylä on December 1, 1977. A PiB event at 20.30 UT is denoted

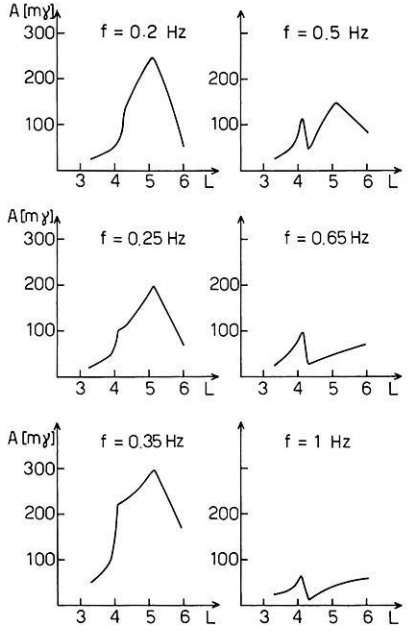


Fig. 10. Latitudinal profiles of the amplitude of PiB pulsations on December 1, 1977 at 20.30 UT at six frequencies

but at higher frequencies a secondary maximum developes at the more southern stations.

Two more characteristics of PiB events can be identified in dynamic spectra. In all cases presented here no clear dispersion of waves is observed. On the other hand, often several PiB's occur as a series of bursts. This is seen in Fig. 4 where at least three PiB's can be identified within about 10 min.

PiB Events and Auroral Structures

An isolated substorm on April 6, 1975, has been extensively investigated by Untiedt et al. (1978). In that paper a detailed description of aurora and magnetic disturbances has been given. Some of these measurements will be shortly discussed here in conjunction with PiB activity during this substorm.

It was pointed out by Untiedt et al. (1978) that PiB events can be observed already before the breakup phase of the substorm. These PiB's were identified at the time of local distortions of an auroral arc. From magnetic recordings it was inferred that these distortions enclosed a local region of an upward directed field-aligned current.

A PiB was observed at the onset of the substorm on April 6, 1975, at the same time with an impulsive riometer absorption event (Untiedt et al., 1978). This is in accordance with Heacock's result (1967) that a Pi burst always appears when an impulsive electron precipitation event is observed. The reverse is not true but this seems to be due to lack of suitable riometer stations in Heacock's analysis. Later Heacock and Hunsucker (1977) concluded, that PiB events seem to correspond to inverted-V precipitation events.

Discussion and Conclusions

PiB magnetic pulsations are intimately associated with the development of the magnetospheric disturbances. They occur at the onset of the substorm but also during the pre-breakup phase of the substorm as was shown by Untiedt et al. (1978). There are now strong indications that they should be related to the processes which result in the generation and development of discrete auroral arcs.

Strong electric fields, both horizontal and vertical, and magnetic field aligned currents characterize the discrete aurora (Davis, 1978). Observations of field-aligned currents have been reviewed recently by Potemra (1978) and the existence of horizontal and vertical electric fields at the height of some thousand kilometers has now been confirmed (see reviews by Block, 1978, and Swift, 1978). The observations of the inverted-V electron precipitation events associated with the discrete aurora seem to agree with the models for the electric field configuration above auroras (Frank and Ackerson, 1971; Ackerson and Frank, 1972; Swift, 1978).

Several proposals have been made for the generation of the electric field along the geomagnetic field (Block, 1978; Swift, 1978). The potential double layer is of a special interest in the present context as it has been used recently by Petelski et al. (1978) in modelling of some auroral pulsation phenomena.

A double layer represents an instability of the field-aligned current. According to Carlqvist (1972), the Birkeland current must be strong enough in order to initiate a magnetospheric double layer. The generation and characteristics of the double layer have been much investigated in the laboratory plasma (Torven and Babic, 1975; Quong and Wong, 1976; Torven and Andersson, 1978).

We assume that local changes in the field-aligned current at the time of PiB pulsations as reported by Untiedt et al. (1978) support the view that the potential double layer may occur simultaneously with PiB pulsations. The double layer should be placed at a low altitude, perhaps at some thousand kilometers according to the electric field and particle measurements mentioned above. This would result in an acceleration of electrons. According to Nishida (1964) such electron beams can excite pulsations in the Pi1 and Pi2 period range. Waves up to about 10 Hz are possible. The broad spectrum and lack of dispersion of PiB pulsations could be explained in such a model.

Heacock and Hunsucker (1977) report a close correlation between PiB magnetic field pulsations and electron precipitation pulsations. Only a small time delay between field and particle pulsations was observed. Heacock and Hunsucker conclude that the waves are generated by the precipitation pulsations. They locate the source of PiB pulsations also at a relatively low altitude.

After being generated at a height of some thousand kilometers waves are propagating isotropically (*R*-mode) or anisotropically (*L*-mode). As was pointed out by Petelski et al. (1978) the steep density gradients within the double layer facilitate mode coupling between *L*- and *R*-mode pulsations. Toward the earth two resonance cavities are formed below the source region: one between the maximum of the Alfvén velocity (about 2,000 km) and the lower ionosphere and another between the double layer and the level of 2,000 km. Using typical values for the Alfvén velocity ($v_A \sim 10^6$ m/s in the lower cavity and $v_A \sim 6 \cdot 10^6$ m/s in the higher cavity) we arrive at resonance frequencies of about 0.3 Hz and 0.8 Hz, respectively. The altitude of the double layer is estimated to be 6,000 km.

Due to the cavity resonance intense pulsations at about 0.3 Hz should be expected below the source region. Other parts of the spectrum are attenuated. At the ionospheric level horizontal propagation of waves is possible and the waves around 0.3 Hz can be observed far from the source region. Some attenuation occurs, and the amplitude of these waves at low latitude stations is smaller.

Another band around the higher resonance frequency is only partly penetrating to the ground below the source region. However, it is able to propagate isotropically. These high-frequency waves can be guided to the ground at the plasmapause region. This may explain our observations of the secondary maximum in certain low-latitude recordings.

Laboratory plasma experiments show that the double layer may become unstable at certain conditions (Torven and Babic, 1975; Quong and Wong, 1976) and that the plasma begins to radiate electromagnetic emissions with a broad spectrum. However, these observations refer to non-magnetic plasmas. It is not clear how they can be extrapolated to the magnetospheric plasma. Moreover, we feel that there is one serious difficulty with this model where the electric current is flowing against the direction of the magnetic field. In such a case the emission of waves should be confined in a cone directed upwards, and it would be difficult to explain the pulsations observed on the ground.

In this paper, the emphasis has been put on Pi1 pulsations. However, it is important to note that according to Stuart (1974), Lanzerotti and Fukunishi (1974), and Saito et al. (1976) the amplitude of Pi2 waves shows the main maximum in the auroral zone and a secondary maximum, at a higher frequency,

at the plasmapause region. This indicates also that the plasmapause has an important effect on the propagation of hydromagnetic waves.

Summary

Some new characteristics of PiB magnetic pulsations for the Pi1 period range have been revealed in the present paper. These results are summarized in the following way:

1. PiB's are observed simultaneosly with the brightening and local distortions of the auroral arc.
2. PiB dynamic spectrum does not indicate any pronounced dispersion of waves. Sometimes several PiB's occur as a series of bursts.
3. An intense amplitude maximum of waves appears at auroral latidues at about 0.3–0.4 Hz; a less entense maximum at the same frequency is observed at lower latitudes.
4. A secondary amplitude maximum at about 1 Hz is sometimes seen in the PiB spectrum at certain lower latitude stations.

We think that the generation of PiB pulsations in the Pi1 period range should be placed at auroral latitudes at a low altitude where also the acceleration of auroral electrons seems to take place. Some of our observations indicate the existence of low-altitude resonance cavities. Also the isotropic propagation seems to contribute in an important way to the characteristics of PiB magnetic pulsations on the ground.

PiB events have several well-defined characteristics. It is expected that they could give useful information about the low-altitude electric fields and acceleration of auroral particles. More details about the relationship between PiB's, field-aligned currents and auroral structures are being collected during the present IMS. Also the relation between Pi1 and Pi2 pulsations during the PiB event needs to be studied.

References

Ackerson, K.L., Frank, L.A.: Correlated satellite measurements of low-energy electron precipitation and groundbased observations of a visible auroral arc. J. Geophys. Res. **77**, 1128–1136, 1972

Block, L.: A double layer review. Astrophys. Space Sci. **55**, 59–83, 1978

Carlqvist, P.: On the formation of double layers in plasmas. Cosmic Electrodynamics **3**, 377–388, 1972

Coleman, P.J., McPherron, R.L.: Substorm observations of magnetic perturbations and ULF waves at synchronous orbit by ATS-1 and ATS-6. In: The Scientific Satellite Programme during the International Magnetospheric Study, Knott and Battrick, eds., pp. 345–365. Dordrecht-Holland: D. Reidel Publ. Co. 1976

Davis, T.N.: Observed characteristics of auroral forms. Space Sci. Rev. **22**, 77–113, 1978

Frank, L.A., Ackerson, K.L.: Observations of charged particle precipitation into the auroral zone. J. Geophys. Res. **76**, 3612–3643, 1971

Heacock, R.R.: Two subtypes of type Pi micropulsations. J. Geophys. Res. **72**, 3905–3917, 1967

Heacock, R.R., Hunsucker, R.D.: A study of concurrent magnetic field and particle precipitation pulsations, 0.005 to 0.5 Hz, recorded near College, Alaska. J. Atmos. Terr. Phys. **39**, 487–501, 1977

Lanzerotti, L.J., Fukunishi, H.: Modes of magnetohydrodynamic waves in the magnetosphere. Rev. Geophys. Space Phys. **12**, 724–729, 1974

Nishida, A.: Theory of irregular magnetic micropulsations associated with a magnetic bay. J. Geophys. Res. **69**, 947–954, 1964

Nishida, A.: Geomagnetic diagnosis of the Magnetosphere. Berlin, Heidelberg, New York: Springer 1978

Petelski, E.F., Fahleson, U., Shawhan, S.D.: Models for quasi-periodic electric fields and associated electron precipitation in the auroral zone. J. Geophys. Res. **83**, 2489–2498, 1978

Potemra, T.A.: Observation of Birkeland currents with the Triad satellite. Astrophys. Space Sci. **58**, 207–226, 1978

Quong, B.H., Wong, A.Y.: Formation of potential double layers in plasmas. Phys. Rev. Lett. **37**, 1393–1396, 1976

Saito, T.: Geomagnetic pulsations. Space Sci. Rev. **10**, 319–412, 1969

Saito, T., Sakurai, T., Koyama, Y.: Mechanism of association between Pi2 pulsation and magnetospheric substorm. J. Atmos. Terr. Phys. **38**, 1265–1277, 1976

Stuart, W.F.: A mechanism of selective enhancement of Pi2's by the plasmasphere. J. Atmos. Terr. Phys. **36**, 851–859, 1974

Swift, D.W.: Mechanisms for the discrete aurora. A review. Space Sci. Rev. **22**, 35–75, 1978

Torven, S., Babic, M.: Current chopping space charge layers in a low pressure arc plasma. In: Proc. 12th International Conference on Phenomena in ionized gases. New York: American Elsevier Publ. Co. 1975

Torven, S., Andersson, D.: Observations of electric double layers in a magnetized plasma column. TRITA-EPP-78-12. Royal Institute of Technology, Stockholm, 1978

Untiedt, J., Pellinen, R., Küppers, F., Opgenoorth, H.J., Pelster, W.D., Baumjohann, W., Ranta, H., Kangas, J., Czechowsky, P., Heikkilä, W.J.: Observations of the initial development of an auroral and magnetic substorm at magnetic midnight. J. Geophys. **45**, 41–65, 1978

Received May 28, 1979; Revised Version August 22, 1979; Accepted August 31, 1979

J. Geophys. 46, 249–259, 1979

Journal of Geophysics

Cosmic Noise Absorption Events and Magnetic Pulsation Activity During Substorms

U. Wedeken[1], O. Hillebrand[1], E. Krenzien[1], A. Ranta[2], H. Ranta[2], and H. Voelker[1]

[1] Institut für Geophysik der Universität Göttingen, Herzberger Landstraße 180, D-3400 Göttingen, Federal Republic of Germany
[2] Geophysical Observatory, SF-99600 Sodankylä, Finland

Abstract. Meridional profiles of cosmic noise absorption and magnetic pulsations have been analyzed for three substorms observed over Scandinavia during disturbed (December 2, 1977), moderately disturbed (March 1, 1978) and slightly disturbed (March 2, 1978) conditions. Localized absorption events in the afternoon sector, a growth phase of a substorm, breakup phases of substorms and pulsating absorption events are discussed. The activity of Pi2,3 pulsations is enhanced near the position of the polar electrojet and coincides with the location of maximum cosmic noise absorption. Largest pulsation amplitudes occur at times of steeply increasing cosmic noise absorption. Two onsets on December 2 show a similar time delay from south to north in the riometer and the pulsation activity data.

Key words: Cosmic noise absorption CNA – Magnetic pulsations Pi1,2,3 – Polar electrojet – Substorm.

1. Introduction

During a magnetospheric substorm one observed phenomenon is precipitation of electrons into the atmosphere. Auroral absorption events are caused by electrons whose energy is greater than 40 keV.

The initiation of an auroral absorption substorm most frequently occurs near magnetic midnight (The term auroral absorption substorm is used because it is the heavy ionization during the substorm which can be studied by means of riometers in and near the auroral zone). The onset of an auroral absorption substorm in the longitudinal sector where the substorm breaks up is very sharp and a weak equatorwards moving absorption precedes the sharp onset (Ranta, 1978). Latitudinally very localized absorption events (Lukkari et al., 1977) are found at L-values between 3 and 6 in the afternoon sector during a substorm. Absorption events observed at the same time with IPDP type pulsations (Lukkari et al., 1977) are interpreted by Thorne (1974) to be due to drifting protons which are injected towards the earth during the initiation of a substorm and which stimulate instabilities to precipitate energetic electrons from the magnetosphere.

0340-062X/79/0046/0249/$02.20

Some of the electrons injected during a substorm are not precipitated immediately, but are first captured by the closed field lines. In the later phase of a substorm these electrons drift to the morning side and are precipitated there (Arnoldy and Chan, 1969). The longitudinal expansion velocities are in the range 0.7–7 km/s, except for the extreme values which exceed 20 km/s (Berkey et al., 1974). On the morning side well defined absorption pulsations appear occasionally. Absorption pulsations with periods of several minutes and 5–20 s are observed (Hargreaves, 1969).

For magnetic pulsations the onset of a substorm is marked by a Pi2 event (formerly pt) occurring simultaneously at stations in auroral and middle latitudes in the evening sector (Saito et al., 1976). During the expansion phase there is strong activity in the Pi 1,2,3 band preferentially in the vicinity of the auroral zone. These pulsations are assumed to be partly due to hydromagnetic waves in the magnetosphere and to fluctuations of the substorm current systems (Kuwashima, 1975). Theoretical examinations of the magnetospheric origin of magnetic pulsations are developed by Chen and Hasegawa (1974). The propagation of the hydromagnetic waves through the ionosphere down to the ground is treated by Hughes and Southwood (1976a and b).

In this paper the intensity of magnetic pulsation activity for two different period ranges is compared with the occurrence of cosmic noise absorption on a meridional profile in the region of the auroral zone. The data were selected from three different time intervals when several auroral breakups took place over Scandinavia.

2. Measurements

A north-south chain of seven riometers is operated by the Geophysical Observatory in Sodankylä on a routine basis at 27.6 MHz. All riometers use a 3-element Yagi antenna and the region sampled is, at 100 km altitude, about 200 km in east-west extent and about 90 km in north-south extent as shown by the ellipses in Fig. 1. The data are recorded with a chart speed of 60 mm/h allowing 1 min time resolution.

During the IMS (International Magnetospheric Study) the University of Göttingen is operating a meridional chain of 3 component Grenet type induction magnetometers in northern Scandinavia. The data are recorded at a rate of 1 sample/s, and the period range of the instruments extends from 2 s to approximately 600 s, thus covering the range of magnetic pulsations from Pc2 to Pc5 and Pi1 to Pi3. The temporary stations are marked in Fig. 1. The precise locations of the riometers are listed by Untiedt et al. (1978). The geographic and geomagnetic coordinates of the pulsation magnetometers, which are operating at the same sites as the IMS magnetometers of the TU Braunschweig, are given by Maurer and Theile (1978).

3. Results

From the existing data material of the IMS special interval December 1977 and the Auroral Breakup Campaign (ABC) in February/March 1978 three sub-

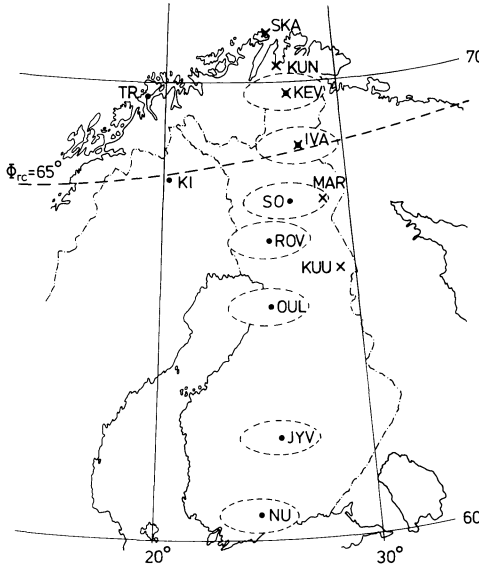


Fig. 1. Map of stations: The seven 27.6 MHz riometers are at *KEV:* Kevo; *IVA:* Ivalo; *SO:* Sodankylä; *ROV:* Rovaniemi; *OUL:* Oulu; *JYV:* Jyväskylä; *NU:* Nurmijärvi. *L*-values of these sites vary between 3.3 and 6.0. Ellipses give the antenna patterns of the riometers projected to the 100 km level in the ionosphere. The locations of the six pulsation magnetometers are *SKA:* Skarsvåg; *KUN:* Kunes; *KEV:* Kevo; *IVA:* Ivalo; *MAR:* Martti; *KUU:* Kuusamo with corresponding *L*-values between 6.8 and 4.6. A curve of constant revised corrected geomagnetic latitude (Gustafsson, 1970), i.e., $\Phi_{rc}=65°$, is denoted by a broken line

storm events have been selected in order to study the possible relation between cosmic noise absorption and magnetic pulsations for times of different magnetic activity:

(a) disturbed conditions on December 2, 1977 (Kp(1200–2400 UT): 4+ 6+ 7o 5−),

(b) moderately disturbed conditions on March 1, 1978 (Kp(1200–2400 UT): 5− 5o 5+ 5o),

(c) slightly disturbed conditions on March 2, 1978 (Kp(1200–2400 UT): 3+ 3o 4+ 3o).

Cosmic noise absorption measurements are analyzed with special emphasis on the events seen in the afternoon sector and around magnetic midnight. For all stations the local magnetic time is about UT+3 h.

The magnetic pulsation data were treated by dynamic spectral analysis with an output of 60 period bands between 5 and 300 s. In this paper the results are compressed into two bands 'Pi1' (5 to 40 s) and 'Pi2,3' (45 to 300 s) with one intensity value every four minutes giving the maximum of the peak to peak amplitude during this time interval. This procedure takes into account the different behaviour of Pi1 and Pi2 pulsations and the maximum time resolution of the method.

3.1. Event of December 2, 1977

There is strong pulsation activity during the whole day. In Fig. 2 riometer and pulsation data of December 2, 1977, 1600–2300 UT are displayed. The magnetic pulsation activity in the 'Pi2,3' band estimated by the spectral analysis described above is shown in a contour map of the total horizontal component ($T=\sqrt{H^2+D^2}$) in a time versus latitude display. For the 'Pi1' band the analogous contour map contains similar results.

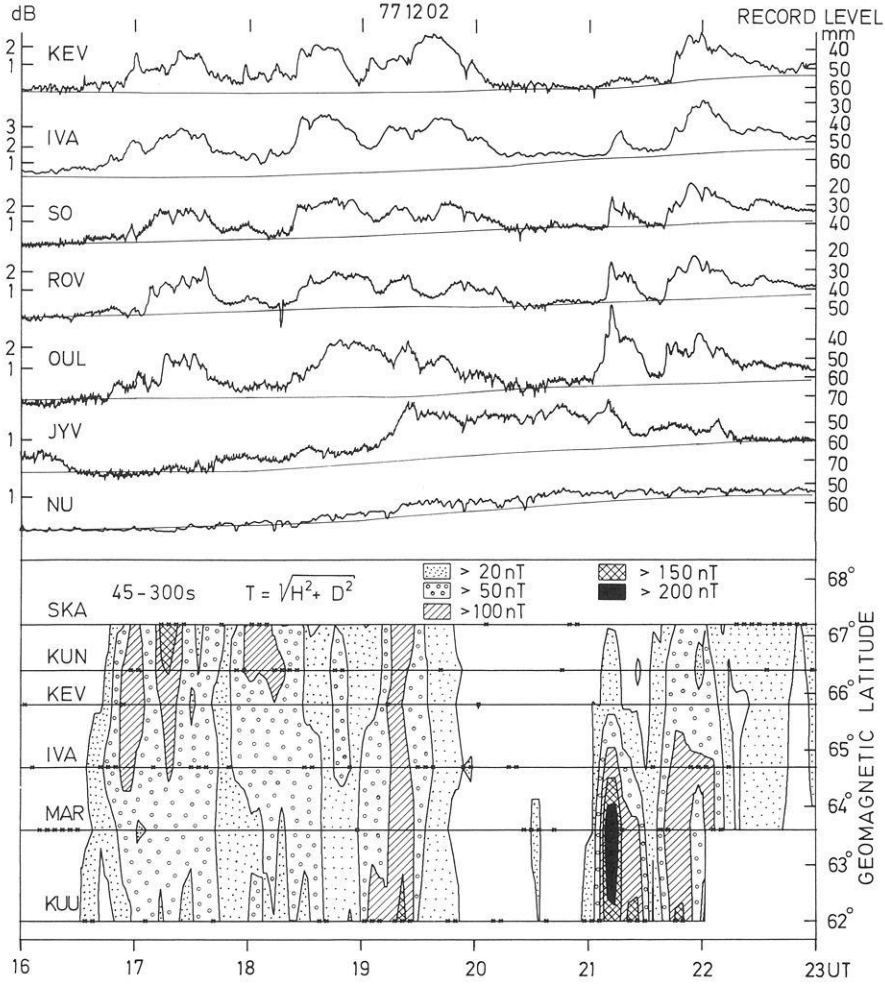


Fig. 2. Cosmic noise absorption recorded by the Finnish riometer chain and data of the Göttingen pulsation magnetometer chain for December 2, 1977, 1600–2300 UT. *Upper panel:* Riometer recordings, undisturbed levels are indicated. *Lower panel:* Magnetic pulsations presented by contour lines of equal amplitude vs. time and geomagnetic latitude for the total horizontal component. The contour lines (20/50/100/150/200 *nT*) indicate the double amplitude envelope of the 'Pi2,3' band (45–300 s). Every four minutes the station with maximum amplitude is marked. No data of *KUU* exist after 2200 UT

Between 1630 and 2000 UT pulsating absorption is observed north of OUL, i.e., the absorption level varies with periods of approximately 5–10 min. No sharp onset is seen during these hours indicating that observed substorms do not break up above Finland but west of it. Pulsating absorption appears most clearly at ROV. The maxima occur at 1708, 1715, 1724, 1730, and 1736 UT.

The magnetic pulsation activity on average increases after 1645 UT, approximately, with a maximum at 1659 UT at KEV. In the original data a time delay in the maximum of activity between the southern and northern stations

can be recognized (2 min in the *H*-components of MAR and KUN). At this time the IMS magnetometer chains of Braunschweig and Münster observe a westward electrojet in northern Scandinavia.

Around 1718 UT, predominantly the northern stations from IVA to SKA record a second Pi2. Simultaneously the *H*-component of the magnetic field increases, especially at KEV and KUN. Near 1800 UT another maximum of activity is observed in the north with two distinct maxima in the original data of KEV, KUN, and SKA at 1756 and 1812 UT respectively. They coincide with increasing absorption at KEV.

In contrast to the previous events the maximum of 'Pi2,3' amplitudes around 1920 UT is found in the south. A detailed analysis reveals increasing *H* and decreasing *D* amplitudes from south to north. The displacement of the general pulsation activity to the south is accompanied by a southward movement of the minimum of *H* in the standard magnetograms, i.e., the westward polar electrojet is shifting to the south.

After 2020 UT only small absorption is seen at JYV. Two clear breakups happen after 2100 UT. Between 2103 and 2110 UT a small absorption event can be recognized at OUL and JYV. Then, the first onset starts at these two stations. Later, at 2112 UT it occurs at IVA. At KEV this substorm cannot be observed clearly. The second onset which starts at 2141 UT at OUL, begins later in the north, at 2145 UT at KEV. After 2200 UT the absorption decreases smoothly. Also during these substorms some pulsating absorption is observed. At ROV, for example, the maxima are seen at 2113, 2120, 2123, 2143, 2148, 2158, 2204, 2208, and 2222 UT, with a distinct minimum between 2130 and 2141 UT.

Simultaneous with the first absorption onset and auroral breakup an increase of the magnetic pulsation amplitudes and of the westward electrojet occurs at 2110 UT in the south. The modulation of the absorption signal described before can also be detected in the magnetic pulsation activity, but because of the resolution of 4 min, however, only four distinct maxima at the southern pulsation stations are seen, at 2113, 2123, 2148, and 2204 UT. The time delays of the absorption onsets for station pairs OUL – IVA and OUL – KEV mentioned above can also be seen in the intensity level of the original magnetic pulsations. Nevertheless the Pi2 onsets at 2110 UT and 2141 UT occur simultaneously at all stations (with small amplitudes in the north). For the second onset the time shift from south to north is obvious too in the negative slope of the *H*-component of the IMS magnetometers between IVA and SKA. Around 2200 UT the region of maximum 'Pi2,3' activity starts migrating to the north with slowly decreasing amplitudes. This is coincident with a northward movement and decrease of the westward polar electrojet.

3.2. Event March 1, 1978

During the Auroral Breakup Campaign (ABC) in February/March 1978 only the two pulsation stations KEV and MAR were operating. In Fig. 3 the original

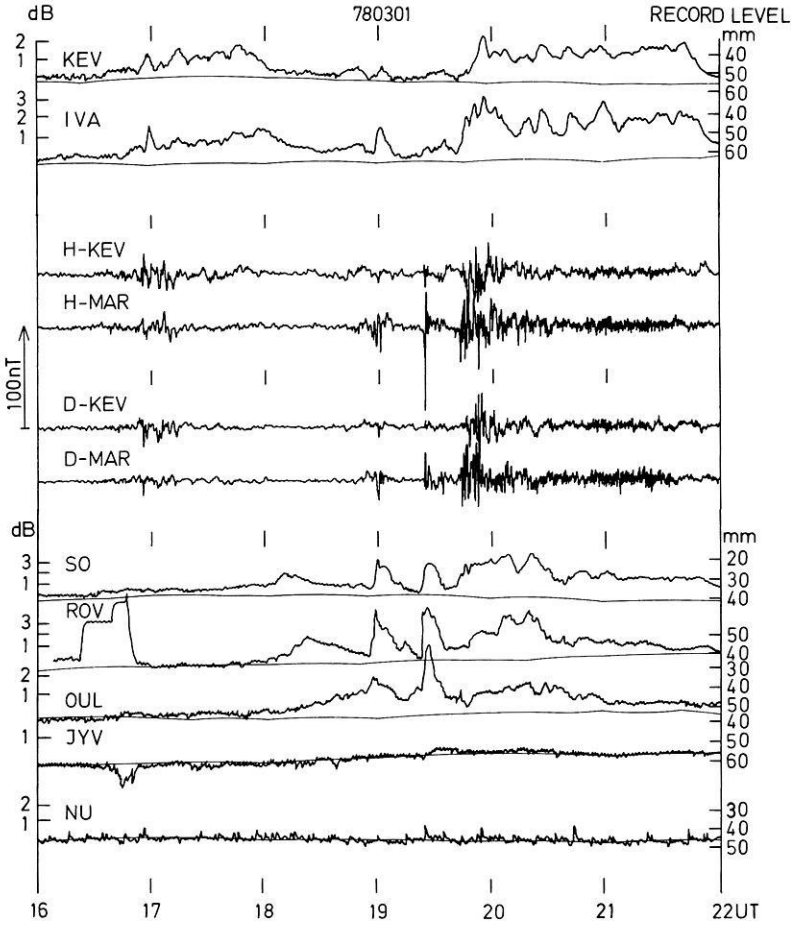


Fig. 3. Riometer and pulsation recordings for March 1, 1978, 1600–2200 UT. To facilitate the comparison the pulsation data of the horizontal components of *KEV* and *MAR* have been inserted in the northern riometer stations. The indicated scale value of 100 *nT* exactly belongs to a period of 20 s

data of the *H*- and *D*-component of both stations are shown together with the riometer data for March 1, 1600–2200 UT.

From 1635 to 1850 UT some absorption is seen at IVA and KEV with a peak around 1657 UT. The sharp onset of a substorm occurs at ROV at 1856 UT and at IVA at 1859 UT. At OUL the absorption begins at about 1855 UT, at KEV at about 1900 UT. A small absorption event preceding the onset is seen north of OUL; it propagates equatorwards. The intensification factor and the propagation speed can be evaluated by comparing the absorption values of SO and ROV. In this event the factor is 1.2 and the speed 130 m/s.

Another onset with preceding absorption is seen at OUL, ROV, and SO at 1924 UT. But the riometers at IVA and KEV do not show this sharp onset. The small preceding absorption is disturbed by the previous substorm. Therefore

this equatorwards moving absorption can only be distinguished at OUL and ROV.

After these two onsets pulsating auroral absorption is observed between 1940 and 2200 UT. The pulsation of electron precipitation is best seen at IVA. At first the period is about 4 min with maxima at 1948, 1952, and 1957 UT. After 2000 UT the period of pulsating absorption varies between 10 and 20 min (maxima at 2004, 2022, 2028, 2044, 2101, 2120, 2133, and 2143 UT). From 2200 to 2300 UT the absorption is rather small, and it begins to increase smoothly after 2300 UT.

Strong magnetic pulsation activity starts at 1654 UT with larger amplitudes at KEV, coinciding with absorption peaks at KEV and IVA. The equatorward travelling absorption between 1800 and 1900 UT is not accompanied by any significant magnetic pulsation structure.

During the first absorption onset the magnetic pulsation activity increases at 1856 UT and reaches a maximum around 1859 UT at MAR. The second onset at 1924 UT is accompanied by a very distinct Pi2 event at MAR, coinciding with the maximum of ionospheric absorption at SO, ROV, and OUL. This Pi2 shows much smaller amplitudes at KEV. Between both onsets the magnetometer chain observes a southward moving eastward electrojet; after the second onset a westward electrojet appears around IVA. The fine structure of the absorption at IVA cannot distinctly be seen at both pulsation stations. MAR, however, has maxima of intensity at 1949 UT and 1959 UT.

The interval from 2000 to 2200 UT shows short period storm associated pulsations, especially at MAR, while long period pulsations of comparable amplitude occur at both stations. The general position of the westward electrojet as derived from the magnetometer data is predominantly in the south, between IVA and MAR.

3.3. Event March 2/3, 1978

In Fig. 4 the pulsation and riometer data of this slightly disturbed event are shown for the time interval 1800 to 0100 UT. From 1800 to 2100 UT small absorption is observed north of ROV. At 1800 UT a series of Pi2 events starts with larger amplitudes at KEV. Simultaneously an eastward electrojet is centered over KEV moving equatorwards. The absorption peak at KEV at about 2016 UT coincides with an auroral arc and high Pi2 activity at this station. The records of the IMS magnetometer chains show the pattern of a Harang discontinuity for this time. Hereafter the activity decreases slowly.

At 2223 UT an absorption onset occurs simultaneously at SO and KI (Kiruna) which is only visible at stations north of OUL. The extremly sharp onset at KI indicates that the auroral breakup happens quite near KI. It is accompanied by strong pulsation activity starting with a Pi2. Worth noting are the different patterns of the pulsation records from KEV and MAR; the larger amplitudes prevail at MAR and at the onset the Pi2 in the *H*-component shows nearly opposite phase behaviour. During this event the existing westward electrojet suddenly moves from IVA to MAR and returns to IVA at about

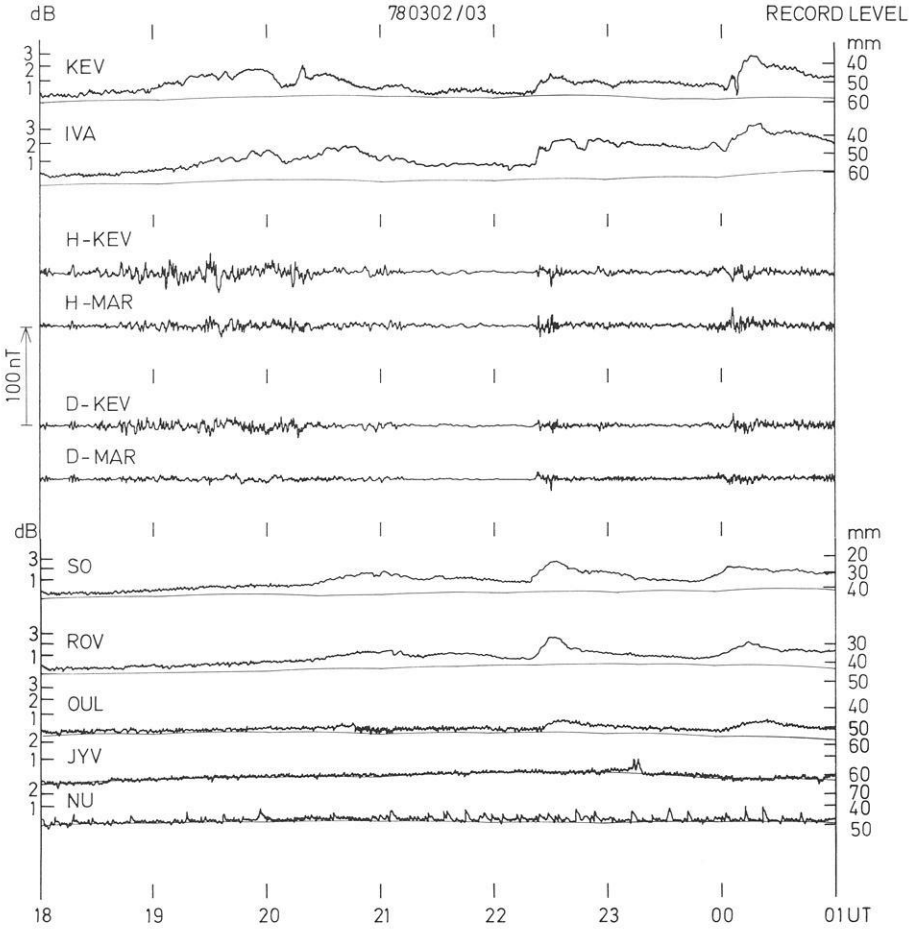


Fig. 4. The same as Fig. 3 for the event March 2/3, 1978, 1800–0100 UT

2250 UT. A further Pi2 onset at 2230 UT behaves in the same manner as far as the phase of H is concerned. This event is correlated with a sharp increase of absorption at IVA. At this time two auroral arcs are observed, one north of KEV and a weaker one south of MAR.

The activity decreases after 2300 UT and starts again shortly before 0000 UT, lasting up to about 0300 UT. Between 0000 and 0300 UT slowly varying absorption is seen north of OUL. Maxima are found at IVA at 2358, 0023, 0110, and 0122 UT, at ROV at 0017 and 0113 UT.

4. Discussion

In the preceding chapter the behaviour of magnetic pulsation activity and of the absorption of cosmic noise for three different disturbed intervals has been

described. Generally a rising CNA (cosmic noise absorption) is coupled with an intensification of Pi activity in both studied period bands (from 5 to 300 s). In detail the largest amplitudes of these pulsations occur just after the onset of absorption events and sometimes remain for the time of maximum absorption. The absolute level of CNA is not coupled to the degree of pulsation activity. However, during each substorm event maximum CNA and maximum pulsation activity occur at the same region of the meridional profile. This is in agreement with results of Novikov et al. (1979), which, different from this investigation, were found for the plasmapause region.

In the event of preceding absorption during March 1 the magnetic pulsations do not show any significant alteration. When the magnetosphere is considered to be the only source of pulsation activity this observation means that there is an influence neither on the generation process of the magnetic pulsations nor on the transmission qualities of the ionosphere during the time of preceding absorption.

According to Wilhelm et al. (1977) the magnetic pulsations in the auroral latitudes are composed of oscillations of magnetospheric origin and irregular fluctuations of ionospheric currents (mainly polar electrojet) and Birkeland currents. Saito et al. (1976) identify substorm onsets by low latitude Pi2. All absorption onsets described in Sect. 3 fall together with clear Pi2 events at the midlatitude station Göttingen. The amplitudes of these Pi2 reach their maximum at auroral latitudes (Baranskiy et al., 1974). This explains the strong pulsation activity just after the onset. The event with the most pronounced Pi2 amplitudes at the onset occurred on March 1 at 1924 UT (see Fig. 3). At this time the auroral breakup happened to be just over the profile. It is coupled to an impulsive Pi2 event in the *H*-component at MAR, the station closest to the breakup.

The time delays between southern and northern stations on the meridional profile during December 2, 1977, around 2112 and 2145 UT, are recognizable in the absorption and the magnetic pulsation intensity data thus giving an apparent south-north velocity of roughly 2–3 km/s. The magnetograms just indicate the second time delay because only the four northern stations were operating at this time and the first event occurred too far south. The pulsation event around 1658 UT gives a similar poleward velocity. The same observation of poleward travelling Pi2 intensity together with a parallel movement of the electrojet was made by Olson and Rostoker (1975). In contrast to their result the Pi2 onsets (December 2 event) occurred simultaneously at stations from midlatitudes to far north of the electrojet; however, the amplitudes of the Pi2 at the onset times are very small in the north. The velocity mentioned above agrees with the speeds of plasma drift motions in the vicinity of active auroral arcs (Whalen et al., 1975).

The clear events of pulsating CNA are correlated to variations found in the standard magnetograms of the corresponding stations. Especially the riometer data and the records of the magnetic *H*-component show similar behaviour. The sensitivity of the pulsation instruments for this period range (about 400 s) is low, so these events are masked by shorter period activity. On the other hand the time resolution of the riometer is not good enough to allow a detailed comparison with magnetic pulsations of periods lower than 300 s. However the

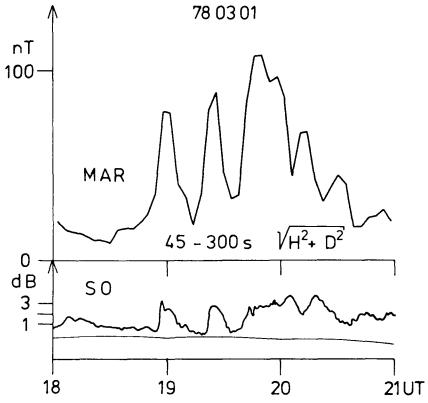


Fig. 5. Display of the pulsation activity at *MAR* and the cosmic noise absorption at *SO* for March 1, 1978, 1800–2100 UT. The maximum double amplitude of the total horizontal component of *MAR* taken every four minutes is plotted

pulsation activity and the CNA usually show a parallel development. For a 3-h interval Fig. 5 demonstrates this correlation. It displays the riometer record of SO and the magnetic pulsation activity in the 'Pi2,3' band at MAR which is situated about 80 km east of SO. At the beginning both curves show an identical behaviour. During the main phase of the substorm (after 2000 UT) the general features of the two plots are similar, but it is difficult to relate single absorption peaks to pulsation activity peaks. Hargreaves (1974) distinguishes between two classes of absorption peaks, those occurring within 10 min of the onset and those 30 min after substorm beginning.

A direct comparison between the magnetic pulsation periods and the riometer data was not possible. According to Stuart et al. (1977) the modulation of the intensity of precipitating electrons in the auroral zone is correlated to coincident Pi2's. After Heacock and Hunsucker (1977), using a high time resolution riometer at College, Alaska, Pi1,2 magnetic pulsations are highly correlated with the pulsating component of the riometer data. Therefore the dominant periods of magnetic pulsations should also be present in the CNA.

Acknowledgements. We wish to thank Dr. H. Lauche, MPAe Lindau, for providing all-sky-camera data of Abisko and Mr. H. Maurer, TU Braunschweig, who kindly sent us plots of his magnetometer data.

The University of Göttingen part of this work has been supported by the Deutsche Forschungsgemeinschaft.

References

Arnoldy, R.L., Chan, K.W.: Particle substorms observed at the geostationary orbit. J. Geophys. Res. **74**, 5019–5028, 1969

Baranskiy, L.N., Shchepetnov, R.V., Afanas'yeva, L.T., Zybin, K.Yu., Hillebrand, O., Sanker Harayan, P.V.: Intensity distribution of Pi2 pulsations along the geomagnetic meridian and on the night side of the earth. Geomagn. Aeron. **14**, 743–746, 1974

Berkey, F.T., Driatsky, V.M., Henriksen, K., Hultquist, B., Jelly, D.H., Schuka, T.I., Theander, A., Yliniemi, J.: A synoptic investigation of particle precipitation dynamics for 60 substorms in IQSY (1964–1965) and IASY (1969). Planet. Space Sci. **22**, 255–307, 1974

Chen, L., Hasegawa, A.: A unified theory of long period micropulsations 1, Steady state approach. J. Geophys. Res. **79**, 1024–1032, 1974

Gustafsson, G.: A revised corrected geomagnetic coordinate system. Ark. Geofys. **5**, 595–617, 1970

Hargreaves, J.K.: Auroral absorption of HF radio waves in the ionosphere: A review of results from the first decade of riometry. Proc. IEEE **57**, 1348–1373, 1969

Hargreaves, J.K.: Dynamic of auroral absorption in the midnight sector – the movement of absorption peaks in relation to the substorm onset. Planet. Space Sci. **22**, 1427–1441, 1974

Heacock, R.R., Hunsucker, R.D.: A study of concurrent magnetic field and particle precipitation pulsations, 0.005 to 0.5 Hz, recorded near College, Alaska. J. Atmos. Terr. Phys. **39**, 487–501, 1977

Hughes, W.J., Southwood, D.J.: The screening of micropulsation signals by the atmosphere and ionosphere. J. Geophys. Res. **81**, 3234–3240, 1976

Hughes, W.J., Southwood, D.J.: An illustration of modification of geomagnetic pulsation structure by the ionosphere. J. Geophys. Res. **81**, 3241–3247, 1976

Kuwashima, M.: Some characters of substorm-associated geomagnetic phenomena in the southern polar region (1). Mem. Kakioka Mag. Obs. **16**, 95–110, 1975

Lukkari, L., Kangas, J., Ranta, H.: Correlated electron precipitation and magnetic IPDP events near the plasmapause. J. Geophys. Res. **82**, 4750–4756, 1977

Maurer, H., Theile, B.: Parameters of the auroral electrojet from magnetic variations along a meridian. J. Geophys. **44**, 415–426, 1978

Novikov, Yu.P., Kopytenko, Yu.A., Raspopov, O.M., Bösinger, T., Kangas, J., Ranta, H.: Observations of the maximum of the magnetic pulsation amplitude at the plasmapause on the afternoon side of the auroral zone. J. Geophys. in press, 1979

Olson, J.V., Rostoker, G.: Pi2 pulsations and the auroral electrojet. Planet. Space Sci. **23**, 1129–1139, 1975

Ranta, H.: The onset of an auroral absorption substorm. J. Geophys. Res. **83**, 3893–3899, 1978

Saito, T., Yumoto, K., Koyama, Y.: Magnetic pulsation Pi2 as a sensitive indicator of magnetospheric substorm. Planet. Space Sci. **24**, 1025–1029, 1976

Stuart, W.F., Green, C.A., Harris, T.J.: Correlation between modulation of the intensity of precipitating electrons in the auroral zone and a coincident Pi2. J. Atmos. Terr. Phys. **39**, 631–635, 1977

Thorne, R.M.: A possible cause of dayside relativistic electron precipitation events. J. Atmos. Terr. Phys. **36**, 635–645, 1974

Untiedt, J., Pellinen, R., Küppers, F., Opgenoorth, H.J., Pelster, W.D., Baumjohann, W., Ranta, H., Kangas, J., Czechowsky, P., Heikkila, W.J.: Observations of the initial development of an auroral and magnetic substorm at magnetic midnight. J. Geophys. **45**, 41–65, 1978

Whalen, B.A., Verschell, H.J., McDiarmid, I.B.: Correlations of ionospheric electric fields and energetic particle precipitation. J. Geophys. Res. **80**, 2137–2145, 1975

Wilhelm, K., Münch, J.W., Kremser, G.: Fluctuations of the auroral zone current system and geomagnetic pulsations. J. Geophys. Res. **82**, 2705–2716, 1977

Received May 29, 1979; Revised Version August 7, 1979; Accepted August 10, 1979

J. Geophys. 46, 261–271, 1979

Journal of Geophysics

On the Sources of the 12-Month Wave in the An and As Geomagnetic Activity Indices

P.N. Mayaud

Institut de Physique du Globe de Paris, Université Pierre et Marie Curie, 4, Place Jussieu, F-75230 Paris Cedex 05, France

Abstract. Damaske (1977; 1978b) and Mayaud (1977b) suggest two greatly different interpretations for the cause of the 12-month wave existing in the an and as indices (or Kn and Ks indices) or in the indices of the five northern and three southern observatory groups from which the former are derived. The present paper discusses the relative validity of these two interpretations. First, the methods of analysis are compared. It is noticeable that the amplitude modulation of the 24-h UT wave, the main argument of the 1977 Damaske'work, is not studied in his 1978b paper because his method does not allow for a separation of the LT component and of its UT modulation. On the other hand, the inverse problem method used by Mayaud does not suffer from this limitation and is applied to much narrower time intervals over the year (60 periods instead of only 8 as considered by Damaske). Furthermore phases and amplitudes are considered by Mayaud whereas Damaske's discussion is mainly based on the amplitudes. Secondly, Mayaud considers morphological features which clearly appear in the records themselves, and which are related to the DP2 fluctuations (Mayaud, 1978a). Damaske neglects them in his analysis, which questions his result. Thirdly, the observed annual and daily variations of the magnetic activity obviously depend on a geometrical factor related to the angle between the solar wind and the dipole axis (with or without an additional tilt). Regarding the question if this geometrical factor is due to an excitation mechanism or to a modulation mechanism, Damaske's interpretation corresponds to the first assumption. Mayaud's interpretation seems to be interesting in that he notices that the second assumption requires the existence of a 12-month wave, whose phase depends on the longitude, and whose interaction with the 12-month wave due to the DP2 fluctuations fully allows for the large scatter in the amplitude of the 12-month wave observed in the various observatory groups.

Key words: Geomagnetic activity indices – An and as indices – Annual wave of geomagnetic activity.

0340-062X/79/0046/0261/$02.20

1. Introduction

According to Damaske (1978b), changes in the 12-month wave of the geomagnetic activity, as they are observed in the indices of the five northern and three southern observatory groups from which the planetary indices an and as (Mayaud, 1968) are derived, do not invalidate the description of hemispheric activity modulation by the function $\sin^2(\beta+\beta_0)$. This function was introduced by him (1977) in order to interpret the systematic amplitude modulation (with opposite sign in both hemispheres) of the 24-h UT wave existing in the Kn and Ks indices, which necessarily induces a 12-month wave. Indeed, the function $\sin^2\beta$ where β is the time dependent angle between the solar wind and the dipole axis allows for both the 6-month wave and the varying UT daily variation existing in the am index (see, for instance, Mayaud, 1977a, where the symbol ψ_M is used instead of β), but does not allow for any difference of the average amplitude between solstices. Then, Damaske assumes that there exists an additional tilt of the effective dipole axis for each hemisphere in the direction away from the sun (one would have $\beta_0=11°$); such a tilt is equivalent to adding dipoles of constant momentum each upon the revolving main dipole. This interpretation of the varying 24-h wave and of the associated 12-month wave greatly differs from the one proposed by Mayaud (1977b), who refers to a special type of disturbances, much larger in local summer during the local afternoon, and to a modulation of the LT auroral disturbances by $\sin^2\beta$. The present paper intends to evaluate the relative validity of these two interpretations.

2. Annual and Daily Variations of the Indices An and As

Figure 1 displays the daily and annual variations of the indices an and as for the years 1959–1974 (a sample identical to those used by Damaske and Mayaud in their respective analyses) in the following way (see also Mayaud 1977a and b): For every 3-h UT interval, each one of the indices has been averaged over the years mentioned and over the nth interval

$$(n\text{–}1)\times 6° < \Lambda < n\times 6°$$

where Λ denotes solar longitude ($\Lambda=0°$ coresponding to the vernal equinox) and n runs from 1 to 60. The resulting eight average values of an and as per interval of Λ thus represent the average daily variation over a period of about six consecutive days at the corresponding time of the year (as given by Λ). The 60 sets of eight values per day and per index (an or as) are then drawn side by side yielding the curves shown within Fig. 1. A modulating function such as the function $\sin^2\beta$ causes a 6-month wave culminating at the equinoxes ($\Lambda=0°$ or 180°) and daily variations made up of a predominantly 24-h wave at the solstices (out of phase from one solstice to the other) and of a predominantly 12-h wave at the equinoxes (see the top-curve of Fig. 2, discussed below and displaying the variation of $\sin^2\beta$). Such features partly appear in Fig. 1 (a 6-month wave and a phase reversal of the daily variation

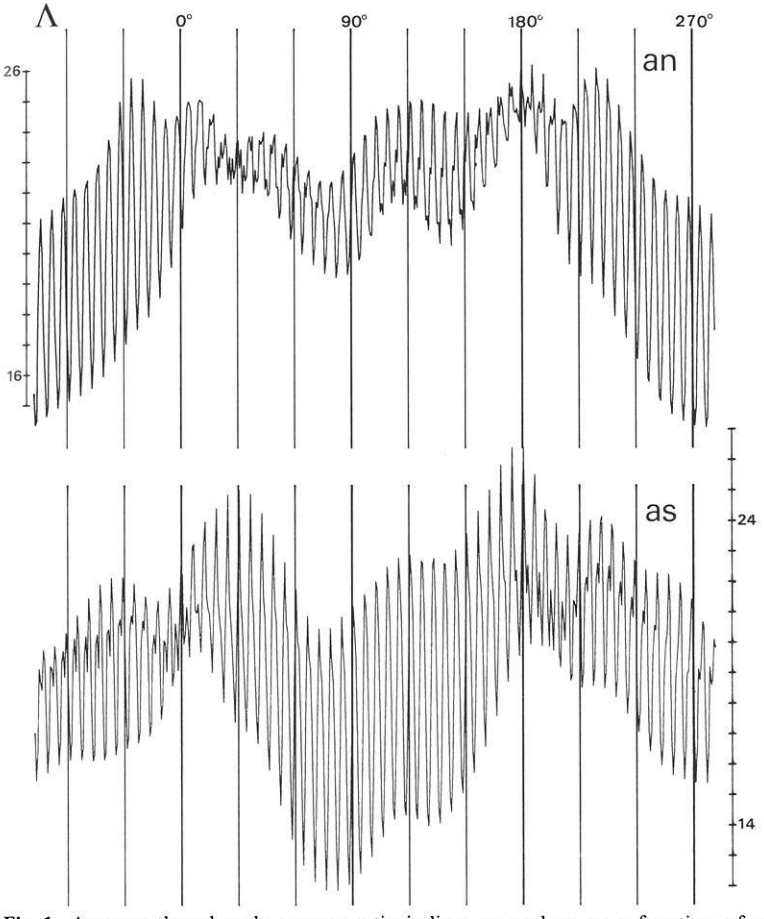


Fig. 1. Average three-hourly geomagnetic indices an and as as a function of solar longitude Λ ($\Lambda = 0°$ corresponding to the vernal equinox) and of the universal time h for the years 1959–1974. The Λ-scale shown indicates some of the lower boundary values of the 60 non-overlapping 6° wide Λ intervals within which the original indices have been averaged for every 3-h UT interval of the day. Within every Λ interval the curves contain 8 average an or as values as a function of h (0–3, 3–6, ..., 21–24 UT; the corresponding 60 UT scales are not shown, because of their smallness). Essentially, each curve gives 60 daily variations in the course of the year, drawn as a continuous function. For more explanation, see text

from one solstice to the other), but two other features appear: (a) a significant daily variation is still present at the equinoxes, and its amplitude is greater in the southern hemisphere (Mayaud, 1979, for a discussion of this feature); (b) there exists a large asymmetry in the amplitude of the daily variation from one solstice to the other in each hemisphere (it is what Damaske calls the amplitude modulation or the diurnal UT wave), which is associated with a 12-month wave culminating in local summer. This is the feature which Damaske and Mayaud interpret in different ways.

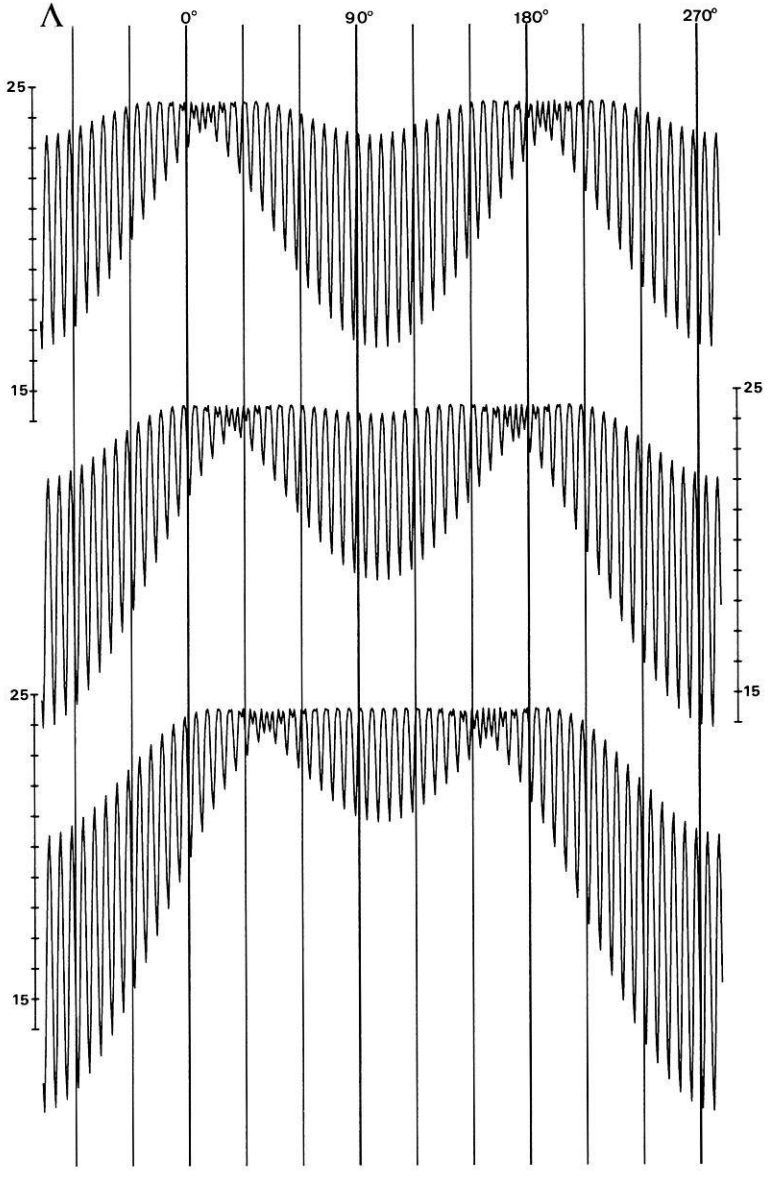


Fig. 2. Annual and daily variations of the function $\sin^2(\beta^+\beta_0)$, as a function (in the manner applied within Fig. 1) of the solar longitude Λ and of the universal time h, for various values of β_0. Top curve: $\beta_0=0°$, mid-curve: $\beta_0=6°$, bottom-curve: $\beta_0=12°$. Note that all these curves are shifted by 10° with respect to the solar longitude $\Lambda=0°$ for taking into account the phase shift of 10° found by Mayaud (1977a) for the index am=(an+as)/2 or the similar shift of 11 days found by Damaske (1977) in the indices an and as

3. Capability of the Function $\sin^2(\beta+\beta_0)$ to Fit the Data

For comparison to the experimental values shown within Fig. 1, Fig. 2 in the same way displays the synthetic function

$$A=A_0+A_1 \sin^2(\beta+\beta_0)$$

for different values of β_0. Here, the above-mentioned angle β is considered to be a function of Λ and h (daily hour, in UT) according to the relation (see Mayaud 1977a and b)

$$\cos\beta=\cos(\pi/2+\delta\sin(\Lambda-\Lambda_0))\cos\Phi$$
$$+\sin(\pi/2+\delta\sin(\Lambda-\Lambda_0))\sin\Phi\cos(h-h_0)$$

where $\delta=23°\ 27'$ denotes the obliquity of the ecliptic and $\Phi=11.5°$ is the earth's dipolar offset angle. Λ_0 and h_0 have been chosen according to the results given by Mayaud (1977a, Table 1) for the am index. A_0 and A_1 have been chosen arbitrarily in such a way that the curves in magnitude roughly correspond to the curves shown within Fig. 1. The sign chosen for β_0 corresponds to the additional tilt in the northern hemisphere; similar curves for the southern hemisphere can be easily imagined. The fact that the bottom-curve is drawn with $\beta_0=12°$ instead of the value 11° evaluated by Damaske (1977) does not invalidate the following remarks. Furthermore, one must note that, if the above formula used for $\cos\beta$ is approximate, it is easy to check that differences with the exact formula are negligible. They are largest for $\Lambda=45°+n\,90°$ (n being an integer) and are of about 1% only. Now, it is obvious that the main feature in the $\beta_0=12°$ curve is similar to the feature (b) mentioned in section 2: The amplitude of the daily variation is graeter in local winter, and the activity level is higher in local summer. Consequently, it might be concluded that the modulating function $\sin^2(\beta+\beta_0)$ is capable of fitting the experimental data.

However, let us first consider a point on which Damaske (1978b) strongly insists: The reliability of the statistical harmonic analysis method used by him, which allows a clear judgement of reality through an 'unobjectionable' test of significance. Such harmonic analyses are performed for individual days (amplitudes and phases of the 24-h and 12-h UT waves) and for sequences of 27-day averages over a fundamental interval of two years (amplitudes and phases of the 12-month and 6-month waves are obtained as the second and fourth harmonics, respectively). In the first case, days are ranked into 8 sectors of 43 days per year, which yields a cloud of 602 points per sector with the 14-year sample used; in the second case, one has a cloud of 7 points (the 7 sequences of 2 years). The limit 2.92 (corresponding to a 0.27% exceeding probability) chosen for the ratio between the vector amplitude and the radius of the probable error circle may be used only at the end of the harmonic analysis, but not to the same extent in the subsequent treatment of the data. Furthermore, non significant vector amplitudes are used in the latter, and when rebuilding the 12-month modulation of the 24-h amplitude wave, phase deviations are ignored. Now, if one computes the value of β, from the ratios g/f given by Damaske (1977) in his analysis of the Kn and Ks indices, as such a value is derived from the 24-h amplitude modulation, one finds that the corresponding values of β_0 are respectively 11°9 and 10°6 in the northern and southern hemispheres. The resulting theoretical ratios between the amplitudes of the 6-month and

12-month waves are equal to 0.49 and 0.54 respectively, while observed values for these ratios, according to his Figs. 19 and 20, are equal to 0.55 and 0.69 hence, they deviate by 12% and 27% in a direction which indicates that the 12-month wave amplitude is too small with respect to the one of the 6-month wave. May one say that the independent check (a full accordance of the observed 12-month and 6-month waves with expectation deduced from the annual amplitude modulation of the 24-h UT wave), considered as the basis of interpretation, is satisfied? The question is all the more sensible since one must remember that, as it will be pointed out in the next section, there exists another source for the 12-month wave observed in the indices, which cannot be ascribed to effects of the additional tilt β_0 of the dipole axis.

Now, a comparison of Figs. 1 and 2 raises several questions. Firstly, the envelopes of the maxima and minima in the curve $\beta_0 = 12°$ of Fig. 2 do not resemble those of the experimental curves of Fig. 1. One has however to be aware that the latter correspond to the linear an and as indices. From Fig. 15 of Damaske (1978a), it is easy to evaluate the values of β_0 for these indices: they are equal to 9.7° and 8.2° respectively. The difference from the result for the Kn and Ks indices is quite significant, and might be questioned. But, whatever be its source (see Sect. 6), it is obvious that the envelope of the maxima, at local summer time, of the an or as indices in Fig. 1 presents a trough which is quite pronounced, while it already hardly marked in the mid-curve of Fig. 2 (that is, with $\beta_0 = 6°$). Again, the 12-month wave amplitude in the indices is smaller than it is in the function $\sin^2(\beta + \beta_0)$, and this is consistent with the deviations mentioned above between the observed and expected ratios (6-month/12-month amplitude wave). Secondly, the solar longitudes at which the daily variation becomes predominantly a 12-h wave are at a distance of about four times 30° in the $\beta_0 = 12°$ curve of Fig. 2. But in the experimental curves of Fig. 1, this distance is apparently of about five times 30° with an, and probably more with as. Now, Damaske's analysis of the diurnal waves is, to some extent, rather crude since he uses only 8 sectors per year, each of them ocrresponding to a solar longitude sector of approximately 45°. A division by 6° wide sectors, as made in Fig. 2 (and as carried out by Mayaud in all his analyses), shows that it is hardly sufficient for following the extremely rapid deformations of the daily variation at some epochs of the year. Thirdly, the morphological aspect of the experimental curves greatly differs from the one of the theoretical functions. We already said that a significant daily variation appears at the equinoxes (the feature a mentioned in Sect. 2). But, the feature b has to be completed by this additional observation: while, during the local winter in each hemisphere, the daily variation of the indices appears to be rather regular (namely, an almost pure 24-h wave) as the daily variation of $\sin^2 \beta$ [or $\sin^2(\beta + \beta_0)$] is, it is no longer true during the local summer. At these epochs, the daily variation is greatly distorted, which indicates that another phenomenon is superimposed, and the additional tilt β_0 of the main dipole does not allow for it since the regularity of the daily variation of the function $\sin^2(\beta + \beta_0)$ is as great at a given solstice as at the other. This additional phenomenon corresponds to a first component of the 12-month wave in the an and as indices, as described by Mayaud (1977b).

4. A Local Time Source of the 12-Month Wave

Bartels et al. (1940) already noticed a summer-winter difference in the activity level from subauroral stations. Mayaud (1956) detected it as being due to a local afternoon phenomenon, but misunderstood its interpretation (see Mayaud, 1978a, on that error). Mc Intosh (1959) referred to this phenomenon, and Mayaud (1965) gave a new and clear illustration of it in displaying statistical daily curves derived from K indices at a chain of European stations. Mayaud (1977b) made a new and extensive study of this particular wave of activity: his Fig. 2 (a 103-year sample of data at two subauroral stations), 3 (a 3-year sample at the European chain of stations mentioned above) and 4 (an 11-year sample at another subauroral station) are an undeniable proof that a particular phenomenon, at work in the H component during the local afternoon, is much larger during the local summer but is almost non existent during the local winter. Furthermore, the phenomenon does not depend on the $\sin^2\beta$ modulating function, and was interpreted by Mayaud (1978a) as being due to the DP2 fluctuations. Now, such a phenomenon necessarily induces a 12-month wave in the an and as indices, which culminates at the local summer of each hemisphere. If its amplitude was identical at any geographical longitude, the local time daily variation of the phenomenon would be averaged out in the planetary indices derived from an ideal longitudinal network distribution, and one would have only the 12-month wave. In other words, the envelopes of the maxima and minima of the daily variation in the top curve of Fig. 2 would present a 12-month wave without any change in the range of the UT daily variation due to $\sin^2\beta$. In fact, the analysis of Mayaud (1977b) for each of the five northern and three southern observatory groups from which an and as indices are derived shows that the amplitude of this particular phenomenon varies from one group to another, that is with longitude; hence, a UT pseudo-component is brought about in the UT daily variation of the an and as indices, which is the cause of the irregularities in the daily variation at other seasons than the local winter (see Fig. 1). Furthermore, it is clear in this Figure that the phenomenon thus superimposed appears to be more intense in the northern hemisphere (an) than in the southern hemisphere (as). This is consistent with Fig. 12 of Mayaud (1977b), which indicates that the amplitude of the phenomenon is, on the average, greater in the five northern groups than in the three southern (it is also the reason for which Damaske, 1977) obtains a smaller deviation in the northern hemisphere than in the southern one when comparing the observed and expected ratios of the 6-month and 12-month waves). Finally, it is obvious that, since Damaske (1977) ignores the existence of this particular phenomenon in his analysis, his results may be questioned and, at least, values obtained for β_0 are certainly greatly overestimated.

5. A Universal Time Source of the 12-Month Wave

We have now to compare the respective interpretations of Damaske (1978b) and Mayaud (1977b) concerning the very large scatter in the amplitudes of the 12-month waves observed in the eight observatory groups from which an

and as indices are derived. Damaske judges that a superimposed systematic amplitude variation cannot be excluded, and might be described in connection with the asymmetry of the polar oval and associated with fields and processes in the magnetospheric tail. Mayaud accounts for it by the interaction between the 12-month wave due to the DP2 fluctuations and another 12-month wave due to the modulating function $\sin^2\beta$ without any interference of an additional tilt β_0.

A first remark has to be made concerning the analyses themselves. Damaske deals with the single annual variation and neglects entirely the daily variation. Then Mayaud's analysis is more comprehensive since it deals with both, and makes that by grouping the data into 60 solar longitude sectors; this guarantees that deformation on the daily variation may be followed with detail. Furthermore, the inverse problem method used allows for a direct and coherent computation of the parameters of both the daily and annual variations; smallness of, and coherence between the residues obtained for the various observatory groups are considered as being the test of the reliability of the results.

A crucial question in evaluating the respective interpretations is as follows: are the geometrical factors $\sin^2\beta$ or $\sin^2(\beta+\beta_0)$ an excitation mechanism or only a modulation mechanism? First, one knows (Mayaud, 1978b) that the ring current variations do not depend on $\sin^2\beta$ since the corresponding UT daily wave does not exist but only the 6-month wave. The DP2 fluctuations are no more dependent on this geometrical factor. We are then left with the auroral disturbances only, and they must be sensitive to $\sin^2\beta$ since they provide the main contribution to the an and as indices. Now, the main daily modulation in the auroral disturbances is a LT variation, due to the configuration of the magnetosphere, and their excitation mechanism is not the geometrical factor $\sin^2\beta$ (it is recognized nowadays that the interplanetary magnetic field plays a determining role). Hence, let us assume that the LT auroral variation is modulated by $\sin^2\beta$: the modulation of a LT daily variation by the UT wave depending on $\sin^2\beta$ must bring about a 12-month wave whose amplitude and phase vary similarly with longitude in both hemispheres.

Table 1 gives the values of $\sin^2\beta$ at the times of its maximum or minimum around the days close to $\Lambda=90°$ (June solstice) or $\Lambda=270°$ (December solstice). If the longitude of the station (or of a given observatory group from which an or as are derived) is such as the LT maximum of the auroral disturbances occurs at 1030 UT (or 2230 UT), the effect of $\sin^2\beta$ would be symmetrical from one solstice to the other, and no 12-month wave is brought about. It is no longer true at longitudes where this maximum occurs at 0430 UT or 1630 UT. In these cases, the intensity of the maximum will be more reduced at the December solstice than the other solstice. Thus, a 12-month wave is brought about which is out of phase from one longitude (LT maximum at 0430 UT) to the opposite longitude (LT maximum at 1630 UT). And, at intermediate longitudes, its amplitude varies according to a sinusoidal law, passing through a zero value where the LT maximum occurs at 1030 UT or 2230 UT. Obviously, the effect is the same, at a given longitude, in both hemispheres. This fact seems to be misunderstood by Damaske (1978b), who refers to this 12-month wave in saying that it corresponds to a greater activity around LT

Table 1

	0430 UT	1630 UT
June solstice	0.96	0.67
December solstice	0.67	0.96

Values of $\sin^2\beta$ for two epochs

midnight during the winter solstice than during the summer solstice. In fact, the modulation is effective at any time of the day and, at a given longitude, one can have in both hemispheres an increase (or a decrease) either at the June solstice or at the December solstice. Furthermore, the resulting effects of this modulation by $\sin^2\beta$ of the auroral disturbances must be well understood at the level of the planetary indices. Assuming (1) that the network of stations is well distributed with longitude and (2) that the LT auroral daily variation has the same range at any longitude, the 12-month waves existing at each longitude would be averaged out. If not, a UT pseudo-component will exist in the planetary indices, and also a 12-month wave. Let us note (and this is very important in order to understand the scatter of the amplitudes of the 12-month wave in the 8 observatory groups) that, at some longitudes in each hemisphere, the two components of the 12-month wave (the one due to the DP2 fluctuations and the one, just described, due to the modulation of the auroral disturbances by $\sin^2\beta$) may have the same phase or opposite phases.

Now, the model used by Mayaud (1977b) allows for these two components when analysing, by the inverse problem method, the 3-h average values of the 60 solar longitude sectors for the indices of the five northern and three southern observatory groups. His Fig. 9 and 10 show that the consistency of the residues is quite good (any 12-month wave culminating at a given solstice has disappeared, and only the 12-month wave culminating at fall and discussed by Damaske (1978b) is still present because the model does not allow for it), and that the amplitude of the 12-month wave due to the $\sin^2\beta$ modulation varies with longitude as expected; furthermore, the times of the LT auroral variation maximum, when corrected from the $\sin^2\beta$ effect, are quite consistent between the 8 observatory groups, since their average is $22.44\,\mathrm{h} \pm 0.34$ in corrected geomagnetic time. Such results are, to some extent, quite impressive since the data analyzed constitute eight series of independent data, and they are obtained without any use of an additional tilt β_0 of the dipole axis. On the other hand, in his analysis of these same data, Damaske (1978b) cannot deal with the method used for the Kn and Ks indices in its totality because a harmonic analysis of the daily variation would not permit him to discriminate between the LT and the UT daily variation at each observatory group. Then, only harmonic analyses of the 27-day averages are performed for each observatory group, which provide amplitude vectors of the 12-month and 6-month waves; this is made both for the quasi-logarithmic indices and the linear indices but amplitude vectors of the 6-month wave are not considered in his paper. With the quasi-logarithmic indices, a large scattering is observed for the 12-

month wave but not for the 6-month wave. Damaske states that 'the average amplitude of the 6-month wave amounts to about two-thirds of the one of the 12-month wave; within the scope of statistical accuracy, this ratio agrees with the theoretical ratio of 0.53 deduced from a value $\beta_0 = \pm 11°$'. In fact, a ratio of $^2/_3$ corresponds to $\beta_0 = \pm 8°7$. Furthermore, ratios of the 6-month and 12-month amplitude vectors for each observatory group fluctuate between 1.28 and 0.44, which corresponds to values of β_0 equal to 4.6° and 13° respectively. May so important variations of β_0 be ascribed to 'the asymmetry of the polar oval which is certainly associated with fields and processes in the magnetospheric tail'? As said above, the non-availability of the 6-month amplitude vectors for the linear indices in Damaske's paper does not permit us to compute the ratios for each observatory group in this case, but the scattering of the 12-month amplitude vectors, as they are illustrated (his Fig. 5, when compared with his Fig. 1), is still larger. In our own line of interpretation, such a scatter is easily interpreted and taken into account by the model because, at some of the observatory groups, the two components of the 12-month wave are in phase while they have opposite phases at others. Let us note that the latter case is the reason why Damaske gets 12-month vector amplitudes below the significance level in some of the observatory groups.

6. Quasi-Logarithmic Indices or Linear Indices?

Damaske (1978b), in his conclusion, states that the quasi-logarithmic indices are more advantageous, when analysing geomagnetic activity modulation, because they are less affected, if at all, by the activity level and then yield the more accurate results thanks to a smaller scattering. This is partly true but calls for the following remarks. Firstly, quasi-logarithmic indices are based on a scale which is greatly distorted with respect to a true logarithmic scale, and one would know the effect of this distortion. In our opinion, the only correct method, if one wishes to use logarithmic values, is to convert the amplitude indices into true logarithms. Anyway, this would accentuate the effect looked for by Damaske. But, if the meaning of an arithmetic average of amplitudes which do not have a gaussian distribution is not always clear, what is the meaning of a geometric average (which is what one uses when taking the average of logarithms)? Secondly, in the present case, it is easy, for instance, to understand the difference found by Damaske (1978b) for the 12-month vector amplitudes of the observatory groups N2 and S8 when one compares those obtained with the quasilogarithmic indices and with the linear indices. These two groups are those where the two components of the 12-month wave (according to our interpretation) have opposite phases. With the quasi-logarithmic indices, the resulting vectors culminate at local summer, while they are greatly reduced in amplitude (with the linear indices) and turn by about or more than 90° (see his Figs. 1 and 5: we do not understand however the phase of the vector **S8** in this Fig. 5 when we look at the annual variation of this group as illustrated in Fig. 8 of Mayaud, 1977b). Now, Fig. 4 of Mayaud (1977b) shows that the effect of the DP2 fluctuations is more important at low activity levels than at high activity levels (these fluctuations are merged into the auroral disturbances in the latter case). This means that the quasi-

logarithmic indices are fully sensitive to the DP2 fluctuations while the effect of the large auroral disturbances is lessened; hence, the effect of the 12-month wave component due to $\sin^2\beta$ is also lessened. When using any index, one has not to forget that the frequency distributions of the individual indices, possibly different for one or the other phenomenon, can play a significant role.

7. Conclusions

Finally, any evaluation of the relative validity of Damaske's and Mayaud's interpretations of the 12-month wave in the indices of the observatory groups has to consider the following facts. Firstly, Mayaud's analysis is more comprehensive since not only annual variations but also daily variations are taken into account. Secondly, the existence of a 12-month wave due to the DP2 fluctuations which are not modulated by the geometrical factor $\sin^2\beta$, is certain. Thirdly, the modulation by $\sin^2\beta$ [or $\sin^2(\beta+\beta_0)$] may be conceived in two ways: it would be either an additive effect (that is, an excitation effect) or a multiplicative effect (that is, a time modulation). What is the most sensible assumption?

References

Bartels, J., Heck, N.H., Johnston, H.F.: Geomagnetic three-hour-range indices for the years 1938 and 1939. Terr. Magn. Atmos. Electr. **45**, 309–337, 1940

Damaske, D.: Magnetospheric modulation of geomagnetic activity. I. Harmonic analysis of quasi-logarithmic indices Km, Kn, and Ks. Ann. Géophys. **33**, 461–478, 1977

Damaske, D.: Magnetospheric modulation of geomagnetic activity. II. Harmonic analysis of linear indices am, an and as. Ann. Géophys. **34**, 231–241, 1978a

Damaske, D.: On the annual wave of hemispheric geomagnetic activity. J. Geophys. **45**, 81–90, 1978b

Mayaud, P.N.: Activité magnétique dans les régions polaires. Ann. Géophys. **12**, 84–191, 1956

Mayaud, P.N.: A propos de la contamination de la variation journalière Sq du champ magnétique terrestre par la variation journalière SD et d'un type spécial de perturbations contribuant au SD d'été. Ann. Géophys. **21**, 219–227, 1965

Mayaud, P.N.: Indices Kn, Ks et Km, 1964–1967. Paris: C.N.R.S. 1968

Mayaud, P.N.: The Boller-Stolov mechanism and the semi-annual and daily McIntosh effects in geomagnetic activity. J. Geophys. Res. **82**, 1266–1270, 1977a

Mayaud, P.N.: Analyse d'une série centenaire d'indices d'activité magnétique, IV. Les diverses composantes de l'onde annuelle aux latitudes subaurorales. Ann. Géophys. **33**, 479–502, 1977b

Mayaud, P.N.: Morphology of the transient irregular variations of the terrestrial magnetic field, and their main statistical laws. Ann. Géophys. **34**, 243–276, 1978a

Mayaud, P.N.: The annual and daily variations of the Dst index. Geophys. J. R. Astron. Soc. **55**, 193–201, 1978b

Mayaud, P.N.: Derivation, meaning and use of geomagnetic indices. To be published as AGU Geophysical Monograph **22**, 1979

McIntosh, D.: On the annual variation of magnetic disturbances. Philos. Trans. R. Soc. London, Ser. A: **251**, 525–552, 1959

Received January 22, 1979; Revised Version August 23, 1979; Accepted August 31, 1979

J. Geophys. 46, 273–289, 1979

Journal of Geophysics

Calculation of the Effect of the Oceans on Geomagnetic Variations With an Application to the Sq Field During the IGY

B.A. Hobbs and G.J.K. Dawes

Department of Geophysics, University of Edinburgh, James Clerk Maxwell Building, Mayfield Road, Edinburgh EH9 3JZ, Great Britain

Abstract. A thin non-uniformly conducting shell at the Earth's surface is taken to represent the distribution of the world's oceans. The continents are represented by zero conductance. The Earth model is completed by a perfectly conducting conductosphere electrically insulated from the surface shell. The theory for electromagnetic induction in such a model has been given by Hobbs and Brignall (1976); this paper presents the detailed method of calculating solutions, using that theory. Examples of induction by various spherical harmonic terms are presented. A synthesis of such solutions is used to approximate the observed external Sq field during the IGY, and the corresponding calculated induced field in the model Earth is compared to that observed.

Key words: Electromagnetic induction – Oceans – Sq – Analytic continuation – Numerical calculations – Vertical component induction.

Introduction

The classical paper of Lahiri and Price (1939) showed that to model the response of the Earth to Sq and Dst variations requires the presence of a conductosphere, in which the conductivity increases steeply with depth, together with a thin conducting shell at the Earth's surface. One possibility is that this shell, taken to be uniform in their model, represents an approximation to the effect of the oceans. However, the oceans are distributed irregularly over the Earth's surface. To obtain a more detailed analysis of their effect on geomagnetic variations therefore requires solution to the problem of electromagnetic induction in a non-uniformly conducting thin shell. Price (1949) gave the appropriate theory and a slight extension, involving analytic continuation (Hobbs and Brignall, 1976) enables his basic theory to be used for general problems. The next step is to make numerical calculations for a conductosphere and oceanic shell in which the conductance of the shell is more representative of the

0340-062X/79/0046/0273/$03.40

ocean/continent distribution. That is the task of this paper. The model for the surface shell is shown in Fig. 1, and is described below. In this first investigation the conductosphere is assumed perfectly conducting. The region between the shell and conductosphere has zero conductivity.

The general method of solution and details of the numerical calculations are described in this paper and examples are given of induction by elementary spherical harmonic terms. Such harmonic terms can be superposed to obtain any inducing field and the corresponding solution may be determined. As an example, a study is made of induction by the Sq field during the IGY, as determined by Malin and Gupta (1977).

Mathematical Method

Since the surface shell is assumed infinitesimally thin the currents induced therein may be described by a current streamline function ψ. The boundary condition to be satisfied at the surface $r=a$ (Price, 1949) is

$$\operatorname{div}(\rho \operatorname{grad} \psi) = -\mu(\partial Z^e/\partial t) - \mu(\partial Z^i/\partial t) \tag{1}$$

where ρ is the reciprocal of the conductance of the shell, μ is the freespace permeability and $(\partial/\partial t)$ denotes differentiation w.r.t. time. For a conductosphere-shell model, Z^e and Z^i, both vertical fields measured positive down, have the following significance. A primary magnetic field external to the shell may be written as the gradient

$$\mathbf{H} = \operatorname{grad} \Omega^p$$

where Ω^p is the scalar magnetic potential. Only the radial component of this field has any induction effect since the model confines induced current flow to concentric shells. The vertical component of this primary inducing field is

$$Z^p = (\partial \Omega^p/\partial r). \tag{2}$$

This primary field induces current flow directly in the conductosphere and the magnetic field of these currents has a vertical component at $r=a$ which we denote by Z^c. The initial vertical field causing induction in the shell is then

$$Z^e = Z^p + Z^c. \tag{3}$$

Currents induced in the surface shell (and described by ψ) have a magnetic field the vertical component of which causes further induction in the shell (the self induction effect). In addition, the magnetic field due to ψ induces currents in the conductosphere, the magnetic field of these currents induces further current flow in the shell (the mutual induction effect). The vertical component of the magnetic field at the surface $r=a$ due to both self and mutual induction is denoted by Z^i. Thus the total vertical component at $r=a$ is

$$Z^{\text{tot}} = Z^e + Z^i, \tag{4}$$

the induced vertical component is

$$Z^{\text{ind}} = Z^c + Z^i. \tag{5}$$

For time-harmonic fields varying as $\exp(i\omega t)$ the operator $(\partial/\partial t)$ may be replaced by $i\omega$ where ω is angular frequency. Price (1949), shows that Eq. (1) may be solved by iteration, where for the first step Z^i is ignored, and for subsequent steps successive approximations to Z^i are used. This method yields a series solution for ψ of the form

$$\psi = \psi_0 + \psi_1 + \psi_2 + \ldots \tag{6}$$

which depends on frequency ω and converges for ω small.

It is convenient to define

$$\phi(\omega) = \frac{\psi}{\omega} = \sum_{j=0}^{\infty} \omega^j \phi_j \tag{7}$$

where $\omega^{j+1}\phi_j = \psi_j$ and ϕ_j is independent of frequency. For a given configuration of conductors, (7) is a Taylor series having a certain radius of convergence. Hobbs and Brignall (1976) show that analytic continuation may be used to determine a series for ϕ convergent for all real ω. The transformation required is

$$p = \omega/(\gamma - \omega) \tag{8}$$

for some constant γ. This yields the solution

$$\phi(\omega) = (\gamma/(\gamma - \omega)) \sum_{l=0}^{\infty} p^l \chi_l \tag{9}$$

where the χ_l's are the linear combinations

$$\chi_l = \sum_{j=0}^{l} \binom{l}{j} \gamma^j \phi_j. \tag{10}$$

The constant γ has an optimum value given by

$$\gamma = 1.14\, \Omega_1\, i \tag{11}$$

where Ω_1, the smallest eigenvvalue of (1), is

$$\Omega_1 = \lim_{j \to \infty} |\phi_{j-1}|/|\phi_j|. \tag{12}$$

$|\phi_j|$ is some norm of ϕ_j (Hobbs and Bringall, 1976, p. 535). With the functions $\{\phi_j\}$ known, series (9) is the solution of Eq. (1) for any real frequency ω.

The frequency independent functions ϕ_j are determined by the following 2-part iterative procedure. The first part of the j'th iterative step is the solution of

$$\operatorname{div}(\rho \operatorname{grad} \phi_j) = R_j \tag{13}$$

for ϕ_j, where R_j is known. The second part is the determination of R_{j+1} from ϕ_j. We define

$$\begin{aligned} R_j &= -i\mu Z^e \quad j=0 \\ &= -i\mu Z_j \quad j \geqq 1 \end{aligned} \tag{14}$$

where

$$Z^i = \sum_{j=1}^{\infty} \omega^j Z_j. \tag{15}$$

This second part is then accomplished by surface integral formulae for Z_j in terms of ϕ_{j-1} given by Hobbs and Price (1970) representing both self and mutual induction. Methods of numerical quadrature for these integrals is given in Hobbs (1971). It remains to solve, at each step, the differential Eq. (13).

A Finite Difference Approximation and Method of Solution

Equation (13) can be solved readily only for very simple distributions of ρ. We determine a numerical solution by replacing the differential equation with a finite difference form and solving the resulting linear system of equations. The discrete points at which values of ϕ_j are to be determined are defined as the intersections (nodes) of latitude and longitude lines, with some given spacing, on the surface $r=a$. In the present application, lines spaced every 5° in latitude and longitude are used, giving 2522 nodes. Each node of this grid is specified as an oceanic or a continental point, these being chosen by reference to world maps (taking into account continental shelves) with the constraint that, apart from Antarctica, there should be one continuous land mass. The conductance at the nodes corresponding to this land mass, and to Antarctica, are set to zero. Figure 1 is a Plate Carré projection of the surface shell in which adjacent nodes corresponding to zero conductance at the edge of land masses have been joined by straight lines. It roughly approximates the distribution of continents and oceans. Nodes for oceanic points are assigned an appropriate value of ρ.

In order to evaluate the r.hs. of Eq. (12) and to measure the accuracy of solutions, a norm $|\phi_j|$ has to be defined. For any function f having numerical values at the oceanic nodes, we choose

$$|f| = \{\Sigma'(f(\theta, \lambda))^2\}^{\frac{1}{2}} \tag{16}$$

where θ and λ are the co-latitude and longitude of a node respectively, and the sum Σ' extends over the 20 nodes indicated by circles in Fig. 1.

For each iteration, ϕ_j is chosen as zero on the continental mass and is a prescribed constant on Antarctica. (The method of determining this constant is given below under 'the island condition.') To find the remaining values of ϕ_j, Eq. (13) has to be discretized at each oceanic point and this is achieved by replacing 1st and 2nd derivatives of ϕ_j, and 1st derivatives of ρ, by their 3-point finite-difference approximations. For points adjacent to a boundary, these

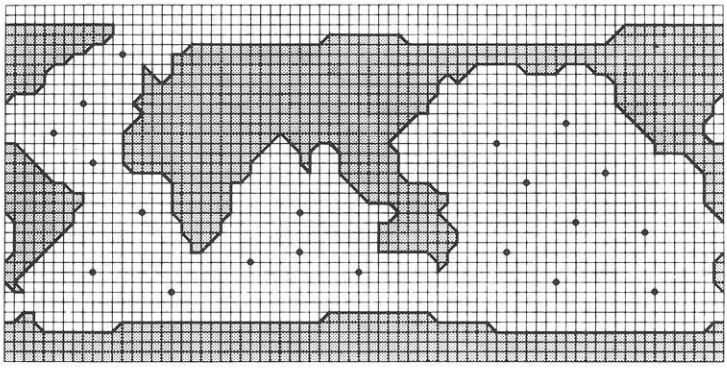


Fig. 1. Plate Carré projection of the surface shell representing the oceans and continents on a $5° \times 5°$ grid. Latitude span is from North to South poles. Longitude span is from 75° W *(left margin)* to 70° W *(right margin)*, an overlap of 5°. *Full circles* indicate nodes from which norms of solutions are calculated. The surface shell contains one mainland and one island

formulae apply directly to ϕ_j, whereas one-sided differences are used for the derivatives of ρ.

There are 1406 oceanic points in the approximation given by Fig. 1. A direct solution therefore requires inversion of a 1406×1406 matrix. The matrix is sparse, may be partitioned, and the partitioned matrices are either zero or striped. It is possible to invert such a matrix, but it is also amenable to iterative methods of solution. An adaptation of the block iterative method of Cuthill and Varga (1959) was used. A natural method of defining blocks is to consider them composed of all points on a given line of latitude between two coastlines. Inspection of Fig. 1 shows that the oceanic points may be represented by 76 such blocks, ranging in length from 1 to 72. To update the values of ϕ_j in a given block, the finite-difference approximation is applied, with the values of ϕ_j on adjacent lines of latitude and boundaries assumed known. The only inversion required is that of a simple tri-diagonal matrix. To avoid bias, the iterations commence with $\phi_j = 0$ in all blocks. Updating the 76 blocks in prescribed order corresponds to 1 iteration. The process is repeated until a given accuracy is achieved. In our numerical examples, the iterations were terminated when the norm $|\phi_j|$ changed from one iteration to the next by less than 0.1 %. (In some cases of slow convergence, 0.25 % was used as the terminating value).

The Island Condition

The current streamline function ψ is arbitrary to within a constant. We are at liberty therefore to choose the value of ψ on one coastline (and it will be the same value over that land mass). We choose the value $\psi = 0$ on the joined continents, but have to determine the constant value of ψ on Antarctica. For each 1st part of our 2-part iterative scheme, this means finding the constant value of $\phi_j = \phi_j^c$ on Antarctica and on its boundary. The value ϕ_j^c must be such

that the corresponding solution ϕ_j satisfies Faraday's law (in our frequency independent form)

$$\oint_\Gamma \rho \hat{\mathbf{r}} \wedge \nabla \phi_j \cdot \mathbf{dl} - i \iint_S \mathbf{B}_j \cdot d\mathbf{S} = 0 \tag{17}$$

where S is a surface that includes Antarctica and which is bounded by a curve Γ lying within the ocean. $\mathbf{B}_j \cdot d\mathbf{S}$ is the flux of magnetic induction passing through an element of the surface $d\mathbf{S}$ at the j'th iterative step. If S is part of the surface $r=a$ then

$$\mathbf{B}_j = \mu Z_j \hat{\mathbf{r}} \tag{18}$$

To find ϕ_j^c we solve Eq. (13) twice with two different boundary values ϕ_j^a and ϕ_j^b, and use the resulting solutions to determine $\mathscr{L}(\phi_j^a)$, $\mathscr{L}(\phi_j^b)$ where $\mathscr{L}$ is the value of the l.h.s. of Eq. (17). The value ϕ_j^c such that $\mathscr{L}(\phi_j^c)=0$ is found by the linear interpolation

$$\phi_j^c = \{\phi_j^a \mathscr{L}(\phi_j^b) - \phi_j^b \mathscr{L}(\phi_j^a)\} / \{\mathscr{L}(\phi_j^a) - \mathscr{L}(\phi_j^b)\}. \tag{19}$$

In the following calculations the boundary values $\phi_j^a = 1$, $\phi_j^b = -1$ were used. Solutions for the many island problem may be similarly constructed as outlined by Bullard and Parker (1970).

Final Solution and Accuracy

For a given configuration of conductors and for a given inducing field (usually in the form of a spherical harmonic of degree n, order m) the functions $\{\phi_j\}$ may be determined and stored. The solution for any frequency is then obtained using the series (9).

To measure the accuracy of a solution for a given inducing field and frequency, and thereby to determine the number of terms required in the sum (9), we re-write Eq. (1) as

$$L = R \tag{20}$$

where

$$L = \operatorname{div}(\rho \operatorname{grad} \psi) + i\omega\mu Z^i \tag{21}$$

and

$$R = -i\omega\mu Z^e. \tag{22}$$

R is given whereas L is determined by the solution. A measure of the accuracy of the solution is afforded by calculating

$$\varepsilon = (|L-R|/|R|) \times 100\,\% \tag{23}$$

where the norms are defined by Eq. (16). We continued adding terms in the sum (9) until ε no longer decreased (presumably because of round-off errors). The

solutions were generally accurate to 1% or less in about 8 iterations, although the last 3 or 4 of these iterations produced little improvement. The ratio of successive norms $|\phi_{j-1}|/|\phi_j|$ had reached a fairly steady value by this stage, so that γ (Eq. 11) was well determined. (In any case, the value of γ is not critical). Thus, as in the simple example considered by Hobbs and Brignall (1976), the series (9) is rapidly convergent and only a few terms, say 4 or 5, need be calculated for accurate solutions. Some examples of convergence are shown in Table 1 for the cases described below.

Numerical Values

The value of ρ for each oceanic point was taken as $0.625\ 10^{-4}$ ohm. This corresponds to an ocean of conductivity $4\,\mathrm{Sm}^{-1}$ and of depth 4 km. The conductosphere was assumed perfectly conducting and of radius $0.9\,a$, the Earth's radius a being $6.37\ 10^3$ km.

A General Inducing Field

Under the usual geomagnetic assumptions, an external varying magnetic field may be represented as the gradient of a scalar magnetic potential function Ω^p satisfying Laplace's equation. The general solution for Ω^p therefore consists of an infinite sum of spherical harmonics each of the form

$$\Omega_n^m(\omega) = a \left(\frac{r}{a}\right)^n \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} P_n^m(\cos\theta) \exp(i\omega t). \tag{24}$$

Here the sphere of reference is the Earth's surface $r=a$, (r, θ, λ) are spherical polar coordinates and $P_n^m(\cos\theta)$ is the Schmidt partially normalised associated Lengendre polynomial of degree n, oder m. Any inducing field may be synthesised by an appropriate sum of functions (24) over n, m, and ω. As in Hobbs (1971) the normal component Z^e of the initial inducing field for the oceanic shell corresponding to the harmonic (24) is

$$Z^e = n(1-\zeta^{2n+1}) \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} P_n^m(\cos\theta) \exp(i\omega t) \tag{25}$$

where $\zeta = b/a$ and b is the radius of the conductosphere.

We have solved the induction equation, and stored the solutions, for a large number of harmonics and are thus able to synthesise many inducing fields. For the Sq field, which is considered below, we have determined solutions corresponding to 16 harmonics, each solution having an in-phase and a quadrature part. The solutions are best displayed by contours on the surface $r=a$ of the current streamline function ψ, the vertical component Z^i and the total vertical component $Z^e + Z^i$. A complete description would thus require 80 diagrams. We give a few representative examples from this set. As the order n of the spherical harmonic increases, so does the value of Z^e in Eq. (25). The relevant coefficient

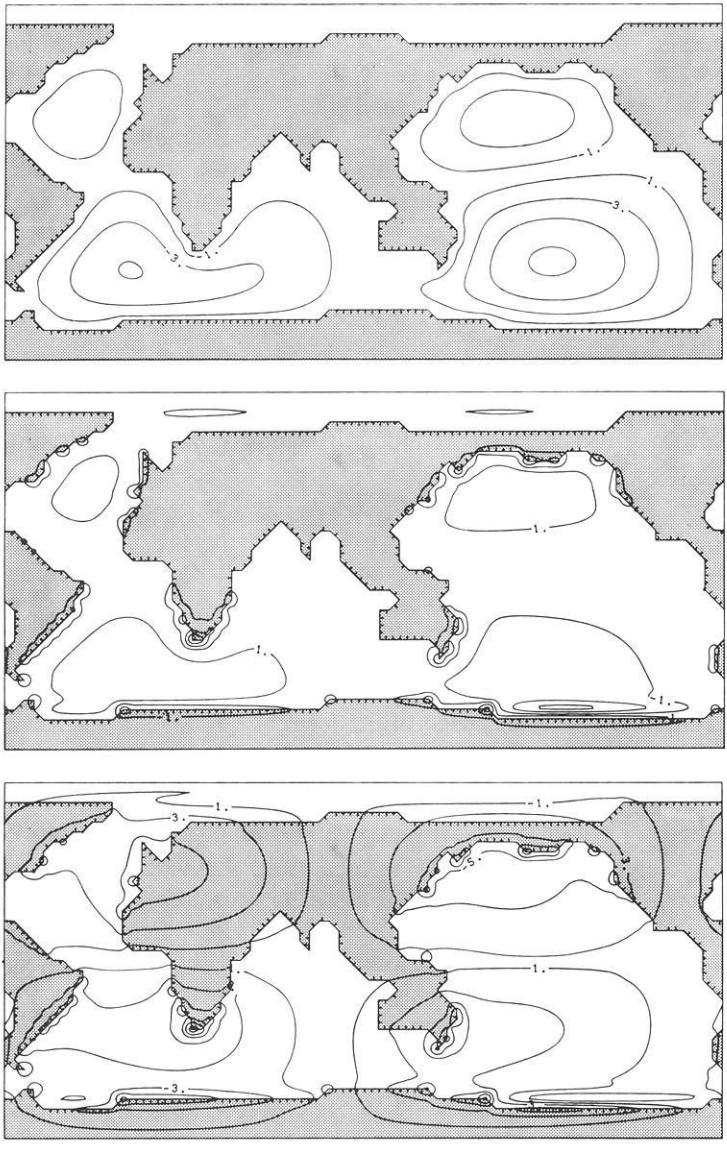


Fig. 2a–c. The in-phase part of the solution for induction by the unit amplitude harmonic $\cos\lambda P_2^1(\cos\theta)$ of period 24h. **a** Streamline function. 500A flows between adjacent contours, positive values indicate anti-clockwise flow. **b** Induced vertical field (positive down) in units of 0.125nT. Contour intervals are 0.25nT. **c** Total vertical field, units and intervals as in **(b)**

being $n(1-\zeta^{2n+1})$ times the normalisation coefficient for the harmonic. This increase is due partly to the spatial variation of the harmonic and partly to the influence of the conductosphere. The increase is notable in Figs. 2, 4, and 5. In specifying an inducing field, care must therefore be taken to ensure that all relevant harmonic terms, especially in the sum for Z^e, are accounted for.

Table 1. Details of the intermediate calculations for 3 spherical harmonic inducing fields

Harmonic Period	$\cos\lambda P_2^1(\cos\theta)$ 24h	$\sin 2\lambda P_3^2(\cos\theta)$ 12h	$\cos 3\lambda P_4^3(\cos\theta)$ 8h
(a) The ratio of successive norms $\lvert\phi_{j-1}\rvert/\lvert\phi_j\rvert$ *in units of* 10^{-4}Hz			
Iteration No.			
3	1.045	1.048	1.073
4	1.044	1.040	1.072
5	1.049	1.038	1.070
6	1.056	1.041	1.068
7	1.066	1.050	1.067
8	1.079	1.062	1.066
9	1.089	1.075	1.068
(b) The parameter ε *(an indication of* % *accuracy achieved)*			
Iteration No.			
3	1.952	1.562	1.613
4	1.313	1.180	0.677
5	0.967	1.021	0.452
6	0.860	1.003	0.399
7	0.824	1.000	0.384
8	0.693	0.983	0.376
9	0.414	0.964	0.373

Example: $\cos\lambda P_2^1(\cos\theta)$; In-Phase; Period 24 h

Figure 2a shows the current streamline function ψ for the above case. The time origin is at $\lambda = 180°$. In a uniform ocean the above harmonic would induce 4 current vortices with symmetry about $\lambda = 180°$ and anti-symmetry about $\theta = 90°$. Our model attempts to respond in that way, but the vortices are greatly distorted by the land masses. The vertical component Z^i is shown contoured in Fig. 2b. The field is practically zero over the continents, is small over oceans and has its largest values near coastlines. Physically, the induced currents are trying to flow so as to oppose the magnetic field over the oceans. The current gradient is steepest near the ocean edges, and this gives rise to large magnetic fields there. In Fig. 2c the induced and inducing fields are combined. Basically the pattern is that of 4 vortices, but these are somewhat distorted by the presence of the oceans. The effect of the non-uniform surface shell is to reduce the vertical field over the oceans and to distort it near ocean/continent boundaries. If the oceans were perfectly conducting (which is mathematically the same as the limit $\omega \to \infty$) the total vertical field over the oceans would be zero. For the real oceanic conductance and for a period of 24 h the solution is far from this limit. (Diagrams 4c and 5c show the effect of increasing frequency, Z^{tot} over the oceans being successively reduced compared to that over continents. The overall strength is increased however, owing to the increased spatial variation). It should be remembered that this simple cylindrical projection gives undue emphasis to points near the north and south poles, the field distortion near Antarctica is not as predominant as it appears in the diagrams.

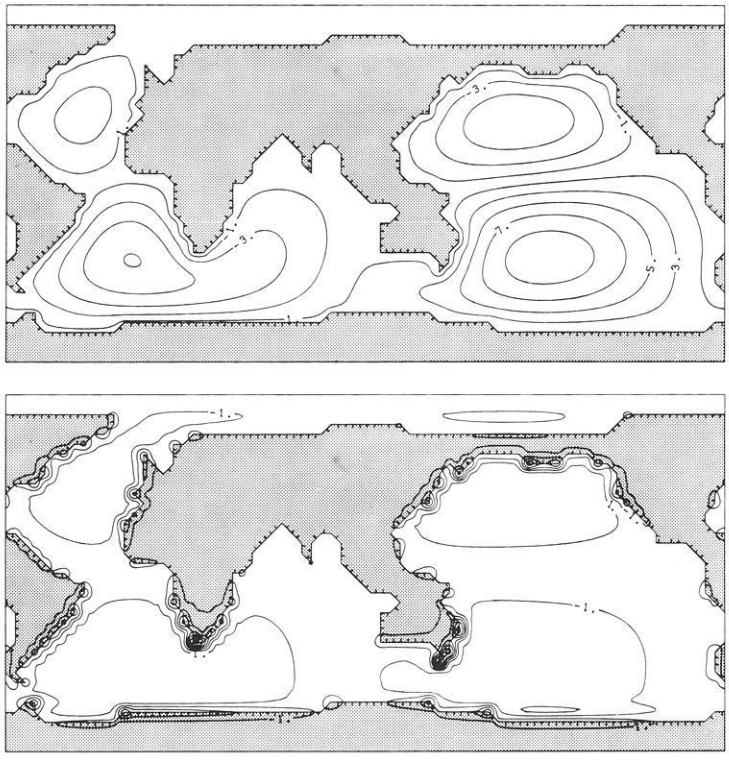


Fig. 3a and b. The quadrature part of the solution for induction by the unit amplitude harmonic $\cos\lambda P_2^1(\cos\theta)$ of period 24h. Contour values as in Fig. 2. **a** Streamline function, **b** induced vertical field

This solution was obtained in 9 iterations to within 0.414% as measured by ε in Eq. (23). Details of the intermediate calculations, in the form of the ratio of successive norms $|\phi_{j-1}|/|\phi_j|$ and the value of ε after each iteration are given in Table 1a and b for this harmonic and for the following examples. This solution could be obtained to within 1% in only 5 iterations.

In all the examples, the smallest eigenvalue is approximately $1.08\ 10^{-4}$ Hz. The implication is that the low frequency method resulting in series (6) is convergent for frequencies less than this value. In terms of period P series (6) is convergent for

$P > 16.2$ h.

This result is predicted approximately by the simple formula of Hobbs (1971) (inequality 20) which is dependent on spherical harmonic order n and which gives $P > 18.3$, 16.7, 15.2 h for orders $n = 2$, 3, 4 respectively.

Example: $\cos\lambda P_2^1(\cos\theta)$; Quadrature; Period 24 h

To complete the description of induction by the harmonic $\cos\lambda P_2^1(\cos\theta)$ the 2 diagrams for the imaginary parts of ψ and Z^i are given in Fig. 3a and b. The induced field in quadrature is seen to be more intense than that in-phase. Again

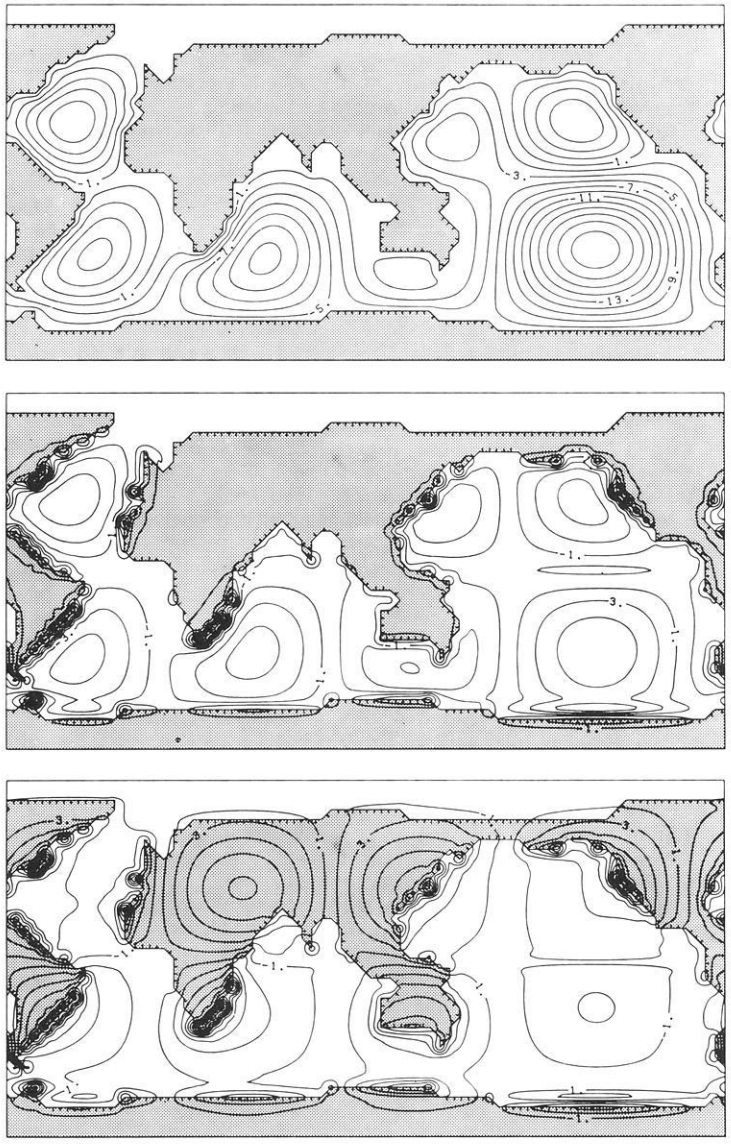


Fig. 4a–c. The in-phase part of the solution for induction by the unit amplitude harmonic $\sin 2\lambda P_3^2(\cos\theta)$ of period 12 h. Contour values as in Fig. 2. **a** Streamline function; **b** induced vertical field; **c** total vertical field

this is due to the low frequency of the inducing field and to the finite conductance of the oceans. It implies that all three terms in Eq. (1) are important. For high frequency (and/or high conductance) the term on the l.h.s. of Eq. (1) becomes relatively small and Z^e (which is real) is balanced by Z^i over the oceans. The part of Z^i in quadrature would then be small. Clearly a period of 24 h is not one of these 'high frequency' problems, but we begin to see these effects for periods of 12 and 8 h.

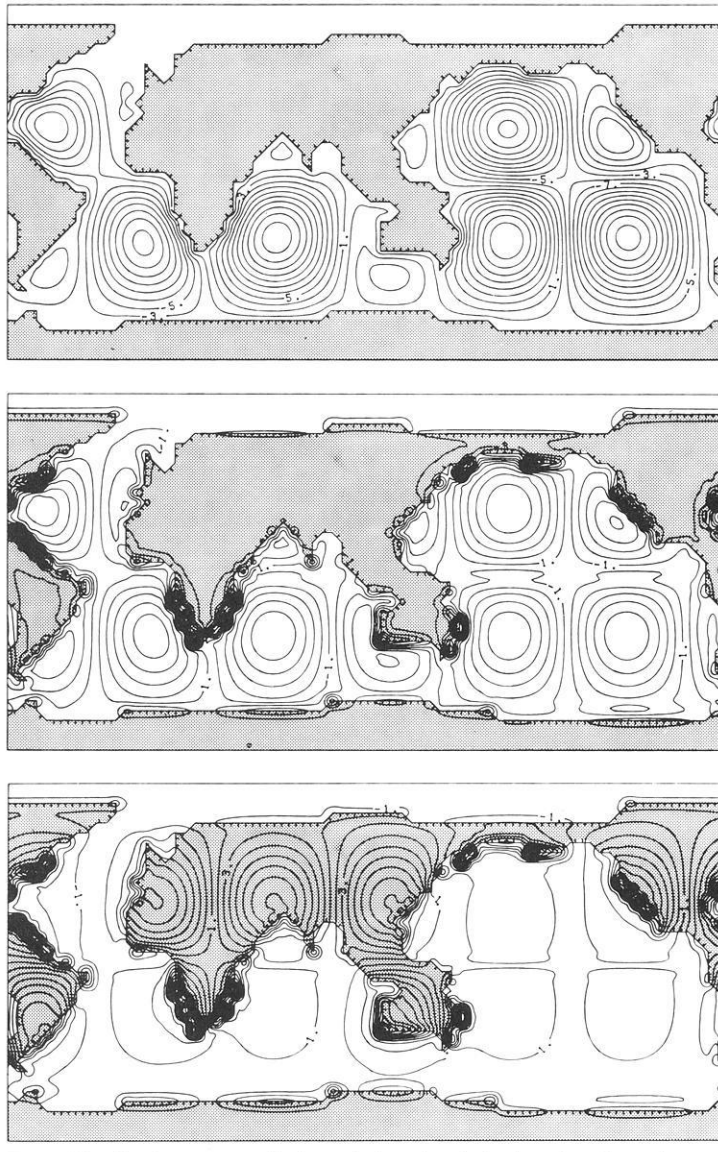


Fig. 5a–c. The in-phase part of the solution for induction by the unit amplitude harmonic $\cos 3\lambda P_4^3(\cos\theta)$ of period 8h. Contour values as in Fig. 2. **a** Streamline function; **b** induced vertical field; **c** total vertical field

Example: $\sin 2\lambda P_3^2(\cos\theta)$; In-Phase; Period 12 h

The relevant diagrams are shown in Fig. 4a–c. This harmonic has 8 vortices, 4 in the northern hemisphere and 4 in the southern hemisphere, with anti-symmetry about $\theta = 90°$. Figure 4a shows how the induced currents attempt to respond. The induced vertical field, Fig. 4b is more intense over oceans and ocean edges

Table 2. Spherical harmonic coefficients used to synthesise the external *Sq* field (extracted from Malin and Gupta, 1977, Table 2). The units are nT

k	n	m	AA^m_{nk}	BA^m_{nk}	AN^m_{nk}	BB^m_{nk}
1	1	1	0.49	−2.96	−4.73	−0.93
1	2	1	11.72	−0.53	−1.29	−9.72
1	4	1	−2.57	−0.51	−0.35	2.33
1	2	0	−0.17	−4.30	0.00	0.00
1	3	0	1.30	2.27	0.00	0.00
2	2	2	1.00	2.07	2.03	−0.69
2	3	2	−5.23	1.74	2.13	5.09
3	3	3	−1.39	−0.67	−0.61	1.04
3	4	3	1.65	−1.52	−1.68	−1.60

than the preceding two examples, and the result of this is seen in Fig. 4c. The total vertical field consists of the 8 vortices, but their strengths are greatly reduced over oceans and there is considerable distortion at ocean edges.

Example: $\cos 3\lambda\, P_4^3(\cos\theta)$; In-Phase; Period 8 h

The highest frequency in this set of examples corresponds to a period of 8 h. The spatial variation of the above harmonic results in 12 vortices, again anti-symmetric about $\theta = 90°$. The induced vortices can be seen in Fig. 5a for ψ, their strengths being dictated by the ocean distribution. The induced vertical field Z^i, Fig. 5b is now quite large over the oceans and their boundaries. Figure 5c shows dramatically the effect on the 12 vortices, their strengths being reduced over the oceans by about 80%. There is a clear progression, seen in Figs. 2c, 4c, and 5c leading towards the high frequency limit in which the in-phase total Z would be reduced to zero over the oceans. The boundary between a continent and an ocean is also that between the unaffected and the reduced vertical field variations. As the frequency increases, this boundary becomes more marked, hence the vertical field values there become more intense.

Application to the Sq Field During the IGY

The scalar magnetic potential for the external part of the field of the daily variations, $\mathrm{Sq}^{(\mathrm{ext})}$, may be written

$$\Omega^{(\mathrm{ext})}_{\mathrm{Sq}} = \mathrm{Re}\{a \sum_{n,m,k} (r/a)^n [(AA^m_{nk} - iBA^m_{nk}) \cos m\lambda + (AB^m_{nk} - iBB^m_{nk}) \sin m\lambda]\, P^m_n(\cos\theta) \exp(ik\alpha t) \tag{26}$$

where t is universal time in seconds and $\alpha = 2\pi/86{,}400$. Many analyses have derived the coefficients AA, AB, BA, and BB for various data sets. A recent analysis for the IGY is given by Malin and Gupta (1977) and we have selected

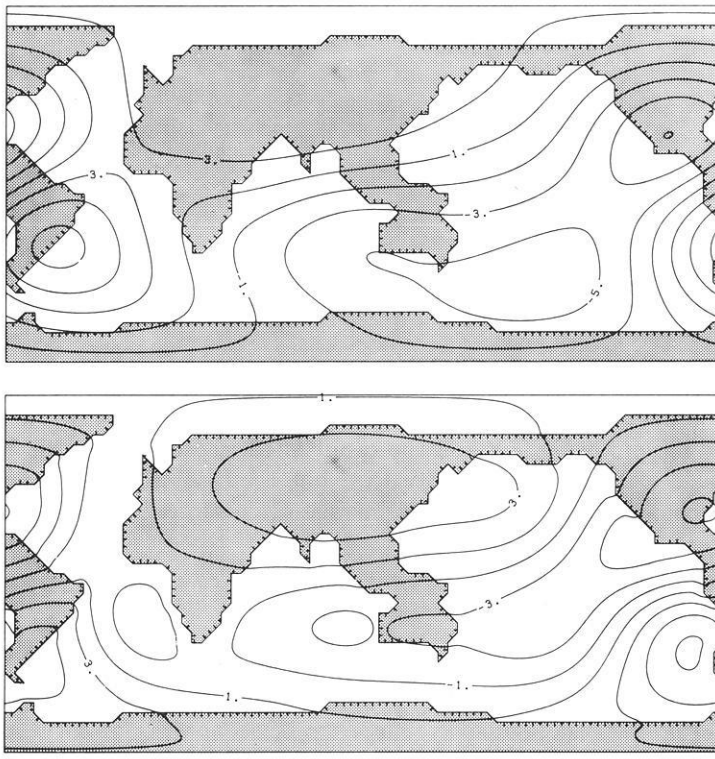


Fig. 6a and b. Equivalent current function for the internal part of *Sq* at universal time $T=16$h. **a** From observations analysed by Malin and Gupta (1977). **b** Calculated, 20 kA flows between adjacent contours

the most important terms to synthesise the $Sq^{(ext)}$ field, i.e., those with the largest coefficients bearing in mind the arguments following Eq. (25) concerning Z^e. The coefficients used, extracted from Malin and Gupta (1977) are given in Table 2.

To determine the response of our Earth/ocean conductivity model, the 16 solutions we have obtained are synthesised, with the $Sq^{(ext)}$ coefficients as weights, in the manner of Eq. (26). In common with previous descriptions of Sq, we present results in terms of an equivalent current system (Malin, 1973) at the surface $r=a$. For the calculated internal part of Sq, this equivalent current system has the streamline function

$$\psi^{(calc)}(t)=\mathrm{Re}(10^3/4\pi)\sum_{n,m,k}\{AA^m_{nk}-iBA^m_{nk})\,\tilde{\psi}^m_{nk,\cos} + (AB^m_{nk}-iBB^m_{nk})\,\tilde{\psi}_{nk,\sin}\}\exp(ik\alpha t) \tag{27}$$

where $\tilde{\psi}^m_{nk,\cos}$ and $\tilde{\psi}^m_{nk,\sin}$ are the equivalent current functions for solutions corresponding to the harmonics in (24). These equivalent current functions are composed of 3 parts, the streamline function for currents flowing in the oceans [e.g., $\psi^m_{nk,\cos}$; a solution to Eq. (1)], the equivalent current function for currents induced directly in the conductosphere by Z^p [e.g., $a((2n+1)/(n+1))\cdot\zeta^{2n+1}\cos m\lambda P^m_n(\cos\theta)$] and the equivalent current function corresponding

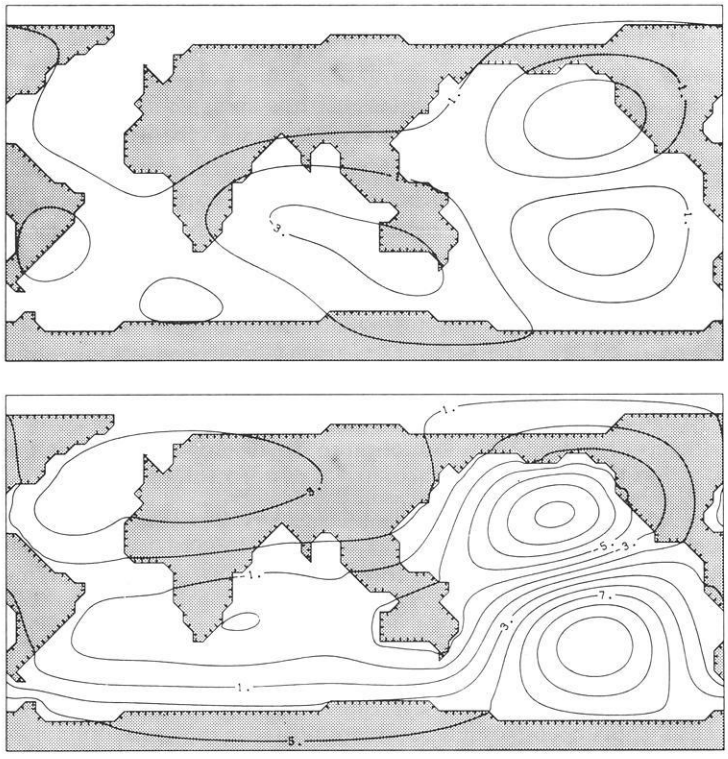


Fig. 7a and b. Equivalent current function for the internal part of *Sq* at universal time $T=20$ h. **a** From observations analysed by Malin and Gupta (1977). **b** Calculated, 20 kA flows between adjacent contours

to mutual induction between the oceans and conductosphere. A surface integral formula for this latter term may be derived using the methods of Hobbs and Price (1970) and is

$$\psi_A(\text{mutual}) = -\frac{1}{4\pi a^2} \iint_S \frac{\zeta(1-\zeta^4)}{(1+\zeta^4-2\zeta^2\cos\Theta)^{\frac{3}{2}}} \psi(\text{oceans})\, dS \tag{28}$$

where Θ is the angle between radius vectors to some point A and the integration point dS. Similarly, for a sine type harmonic the three terms are

$$\tilde{\psi}^m_{nk,\sin} = \psi^m_{nk,\sin} + a[(2n+1)/(n+1)]\, \zeta^{2n+1} \sin m\lambda\, P^m_n(\cos\theta) + \psi^m_{nk,\sin}(\text{mutual}). \tag{29}$$

In (27) we write $t=3600\,T$ so that T is universal time in hours.

The calculated part of Sq may be compared to that derived from the analysis of the observations given by Malin and Gupta (1977). They determine coefficients for the internal part of Sq, from which we may construct the equivalent current system

$$\psi^{(\text{obs})}(t) = \text{Re}(10^3\, a/4\pi) \sum_{n,m,k} [(2n+1)/(n+1)]\, \{(CC^m_{nk} - iDC^m_{nk}) \cos m\lambda + (CD^m_{nk} - iDD^m_{nk}) \sin m\lambda\}\, P^m_n(\cos\theta) \exp(ik\alpha t). \tag{30}$$

The coefficients used are those internal coefficients of Table 2 of Malin and Gupta (1977) labelled as $p=1$, $p=2$ and $p=3$ [in Eq. (30) above, $k \equiv p$].

A comparison between our calculations (27) and the observations (30) for two instants of universal time $T=16$ h and $T=20$ h is shown in Figs. 6 and 7. A numerical way of comparing the solutions is to compute the strengths of the current vortices. In Fig. 6, where the vortices are centred over America, the strengths of the main northern and southern hemisphere vortices are seen to be almost identical, the current flowing betweeen their foci being approximately 140 kA. In Fig. 7, when the vortices are centred over the Pacific Ocean, the strengths are very different and correspond to a current flow betweeen nothern and southern foci of 220 kA for the calculations but only 60 kA for the observations. These represent the best and worse cases, respectively, throughout the period of 24 h. Even in the best example, $T=16$ h Fig. 6, where the current strengths are the same, the current distribution is not. At this instant of universal time, the contribution from induction in the oceans is fairly small, even so, some influence of the oceans in the calculation (Fig. 6b) is evident from the displacement of the current foci towards the Pacific. On the other hand, $T=20$ h (Fig. 7) is a time at which the current vortices are centered over the Pacific Ocean, consequently the oceanic contribution should then be at its maximum. This is clearly so in the calculation, Fig. 7b, but the observations Fig. 7a, suggest the opposite effect, that in the presence of a large ocean, the currents decrease. Such an effect seems to have no physical basis. However the 'observations' are the result of a spherical harmonic analysis of magnetic fields measured predominantly over European and other continental areas. It seems unlikely that they could represent an effect seen only over the oceans and coastal regions. Further work is in progress using also other Sq analyses to see if they can be used in some way as a test of vertical component oceanic induction.

Conclusions

The main point of this paper is to show that it is now possible to solve, numerically, problems of electromagnetic induction in a thin non-uniformly conducting shell surrounding a conductosphere of perfect conductivity. The solutions are determined iteratively in only a few steps and can be obtained for any inducing field and frequency. For a 5° grid they are accurate to about 1 %. It should be noted that these are full solutions to the vertical component oceanic induction problem, as opposed to the simple outer solution approximation of Beamish et al. (1979).

Some of the solutions were used to synthesise the effect of induction by Sq, and a comparison was made between these calculations and observations. The comparison was poor and cannot be said to support the view that vertical component induction in the oceans is the important mode. However the behaviour of the 'observations' in oceanic regions seems to imply that Sq analyses are not useful in testing the appropriateness of thin sheet models.

It is hoped that the solutions will be useful in other applications. For example, a knowledge of the response function of the Earth, that is the ratio of internal to external parts for a given harmonic and frequency, has often been used to infer the Earth's conductivity profile with depth. Such analyses assume the Earth is radially symmetric, so that a given spherical harmonic in the external field induces only that harmonic of the same form within the Earth. This is no longer true when we account for the non-uniform oceans and the solutions we have obtained will be useful in examining the interrelations between the external and the internal hamonics.

Acknowledgements. This work was supported by the Natural Environment Research Council under grant GR3/2705 by means of which one of us (G.J.K.D.) was funded. Edinburgh University provided full computing facilities. The paper was presented at the Fourth I.A.G.A. Workshop on Electromagnetic Induction, Murnau, Germany, September 1978. We would like to thank Dr. W.D. Parkinson for his valuable comments whilst a Senior Visiting Research Fellow in our Department.

References

Beamish, D., Hewson-Browne, R.C., Kendall, P.C., Malin, S.R.C., Quinney, D.A.: Induction in arbitrarily shaped oceans IV: *Sq* for a simple case. Geophys. J.R. Astron Soc. in press, 1979

Bullard, E.C., Parker, R.L.: Electromagnetic induction in the oceans. In: The sea, Vol. IV, Ch. 18, 1970

Cuthill, E.H., Varga, R.S.: A method of normalised block iteration. J. Assoc. Comput. Mach. **6**, 236–244, 1959

Hobbs, B.A.: The calculation of geophysical induction effects using surface integrals. Geophys. J.R. Astron. Soc. **25**, 481–509, 1971

Hobbs, B.A., Brignall, A.M.M.: A method for solving general problems of electromagnetic induction in the oceans. Geophys. J.R. Astron. Soc. **45**, 527–542, 1976

Hobbs, B.A., Price, A.T.: Surface integral formulae for geomagnetic studies. Geophys. J.R. Astron. Soc. **20**, 49–63, 1970

Lahiri, B.N., Price, A.T.: Electromagnetic induction in non-uniform conductors, and the determination of the conductivity of the earth from terrestrial magnetic variations. Philos. Trans. R. Soc. London, Ser. A: **237**, 509–540, 1939

Malin, S.R.C.: Worldwide distribution of geomagnetic tides. Philos. Trans. R. Soc. London, Ser. A: **274**, 551–594, 1973

Malin, S.R.C., Gupta, J.C.: The Sq current system during the International Geophysical Year. Geophys. J.R. Astron. Soc. **49**, 515–529, 1977

Price, A.T.: The induction of electric currents in non-uniform thin sheets and shells. Q.J. Mech. Appl. Math. **2**, 283–310, 1949

Received June 5, 1979; Revised Version August 27, 1979

J. Geophys. 46, 291–299, 1979

Journal of Geophysics

Localized Source Effects on Magnetotelluric Apparent Resistivities

C. Quon [1], K. Vozoff [2, 3], M. Hoversten [3], H.F. Morrison [3], and K-H. Lee [3]

[1] Bedford Institute of Oceanography, Dartmouth, N.S., Canada
[2] Macquarie University, School of Earth Sciences, North Ryde, Sydney, Australia
[3] University of California, Engineering Geoscience, Berkeley, California, USA

Abstract. There is a reawakening of interest in 'source effects' in magnetotellurics, because of the very precise data now being acquired in regions where such effects might be anticipated.

Dipoles at ionospheric heights were taken to be worst-case localized sources. Theoretical apparent resistivities were calculated for horizontal and vertical electric and magnetic dipoles. The earth model was a simple conductive sedimentary basin, where minimal source effects might be anticipated. These results were compared with the usual plane wave apparent resistivities, and with the Price-Wait source models. We found that, with this model, the short period magnetic dipole results were indistinguishable from those due to a planar field, but significant differences occurred at periods greater than 50 s. Horizontal electric dipole sources in particular gave apparent resistivities which differ completely from those due to magnetic dipoles. A more general range of sources and models was also studied.

Key words: Magnetotellurics – Source effects – Three-dimensional modelling.

Historical

For 10 years, from 1955–1965, a great debate was carried on, in and out of the literature, regarding source effects in *MT*. Was Cagniard's assumption, that the source was a plane wave, true or false? Price in 1962 showed that one critical factor was the lateral uniformity of the source field. If it is not uniform, as was assumed by Cagniard, then impedances and apparent resistivities increasingly depart from the plane wave prediction as the period increases. These departures were called 'source effects'.

The geomagnetics discipline viewed the sources of magnetic pulsations in terms of large horizontal loop current sheets, like that responsible for Sq, together with other discrete current systems. These have horizontal scales of 10^4 km or so. They flow at an elevation of 100 km, and Price showed that their

0340-062X/79/0046/0291/$01.80

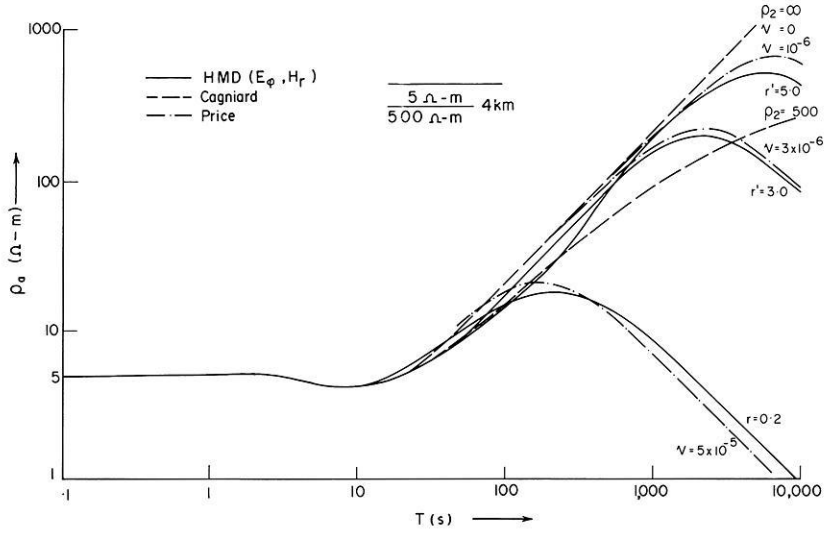


Fig. 1. Summary of results over two-layer models using Cagniard plane wave ($\rho_2=\infty$ and 500), Price's variable horizontal wavenumber ν when $\rho_2=\infty$, and Quon's computations for HMD at various r when $\rho_2=500$. Radial magnetic (H_r) and tangential electric (E_ϕ) components were used to compute dipole ρ_a. The upper layer represents sediments overlying a fixed resistivity basement, r' is r/dipole height, ν is in units of m^{-1}, and ρ is in $\Omega\cdot m$

source effects should be considerable for one particular earth model at least. This, plus the fact that *MT* apparent resistivities obtained at long periods differed from those predicted by purely magnetic observations, cast considerable doubt on Cagniard's approach.

To review this question further, Quon, in 1963, assumed a simplified model of the Alberta sedimentary basin (Fig. 1) and computed the apparent resistivities due to electric and magnetic dipole sources in the ionosphere, at 100 km elevation. These were then compared with the plane wave prediction, and with Price's prediction for a range of horizontal wavelengths. Quon used dipoles at that height since they represented extreme cases of source concentration and hence worst possible concentration of short spatial wavelengths. At periods up to about 100 s, Quon found that with one exception, (which now appears to have been an error) it made little difference whether the source was a plane wave or a dipole. At longer periods the magnetic dipole curves decreased in much the same way as Price's. Also, at any given horizontal distance r from the dipole, ρ_a was nearly the same as if the source field had horizontal wavelength $=r$. That is, horizontal offset and horizontal wavelength play nearly identical roles, as had been anticipated (Fig. 1). Note that r' in all figures is the ratio of radial distance to dipole elevation, ν has the units of $(\text{meters})^{-1}$ and resistivities are given in ohm · m.

When the source was a vertical electric dipole (VED) (Fig. 2), apparent resistivities increased monotonically with increasing period, rather than approaching ρ_2 (as Cagniard predicted) or decreasing (as for the Price and HMD models).

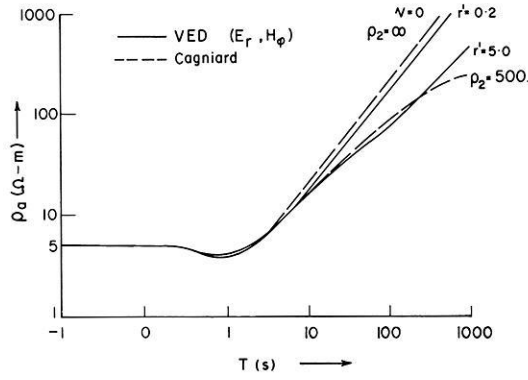


Fig. 2. Comparison of Cagniard $\rho_a(\rho_2=\infty, 500)$ with those of Quon for VED ($\rho_2=500$)

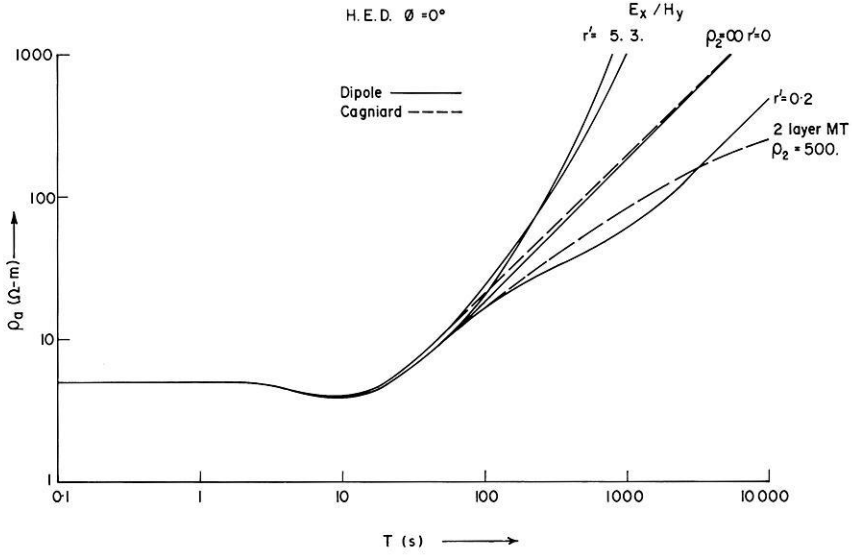


Fig. 3. Corrected two-layer ρ_a for the Horizontal Electric Dipole. The dipole is at 100 km elevation oriented along $\phi=0$. $\phi=0$ corresponds to the x-axis

The horizontal electric dipole (HED) gave ρ_a curves which differed completely from the others. It is these that are erroneous. We recently recomputed those same models, and found that they are very similar to the others at periods less than 10 s. From 10 to 100 seconds they may fall above or below the plane wave response curve, depending on position and component. Beyond 1,000 s ρ_a invariably sloped upward at 45° (Fig. 3). $\phi=0$ corresponds to the x-axis.

Two works appearing in the early 1960's shed more light on the topic. Cantwell's (1960) thesis showed that at least one major addition to Cagniard's model was required. That is, lateral conductivity changes in the earth had to be treated much more seriously. Whereas Cagniard wrote

$$E=zH \tag{1}$$

as a scalar equation relating E and the (perpendicular) H component at each frequency at a site, Cantwell showed that a tensor equation was necessary

$$E_x = Z_{xx}H_x + Z_{xy}H_y$$
$$E_y = Z_{yx}H_x + Z_{yy}H_y. \tag{2}$$

Neglecting this fact, especially near coastlines and other major tectonic features, gave apparent resistivities which were orders of magnitude different from true resistivities, and which varied in time. These were sometimes attributed to source effects.

In 1964 Madden and Nelson issued their classical report 'A defense of Cagniard's magnetotelluric method'. There they showed that, in a laterally uniform earth, the important factor is the ratio of horizontal (source field) wavelength, λ, to skin depth δ

$$\frac{\lambda}{\delta} \approx \frac{(\omega\mu\sigma)^{\frac{1}{2}}}{\nu} \quad \begin{cases} \gg 1 \text{ no source effects} \\ \leqq 1 \text{ large source effects.} \end{cases} \tag{3}$$

Source effects are important only when this is the order of unity or less. In sedimentary basins, where much of the MT work is done, resistivities are low, skin depths are small, and source effects are negligible over the entire range of frequencies used.

These developments, and the usual data scatter resulting from artificial noise and inadequate instrumentation, put an end to the debate at that time.

Very recently, Gamble et al. (1979) showed that, by use of their remote reference technique, apparent resistivity scatter can be reduced to 1% or less. Whereas the method is new, it has since been applied in other surveys in complex areas. It consistently improves the repeatability of results, and makes possible useful measurements in areas so noisy that good data cannot otherwise be obtained.

In another very recent work, Dmitriev and Berdichewsky (1978) showed that the source field over a horizontally layered medium could vary linearly with distance and still give correct values of apparent resistivitiy.

It has therefore become important to reexamine the source effect. The emergence of 3D model programs and of improved, inexpensive 2D model programs, makes this task practicable.

New Computations – Extended Quon Model

The extensive modelling capabilities at Berkeley and Macquarie were used to verify and extend Quon's calculations and to consider other simple models as well.

As mentioned earlier, we verified most of Quon's results except for the HED, which we corrected. Deviation from the plane wave response in the Alberta model was found only at periods >100 s or so. Since this deviation would be expected to diminish for higher conductivities, and since one commonly sees a conductivity increase in the upper mantle, we extended Quon's model by adding

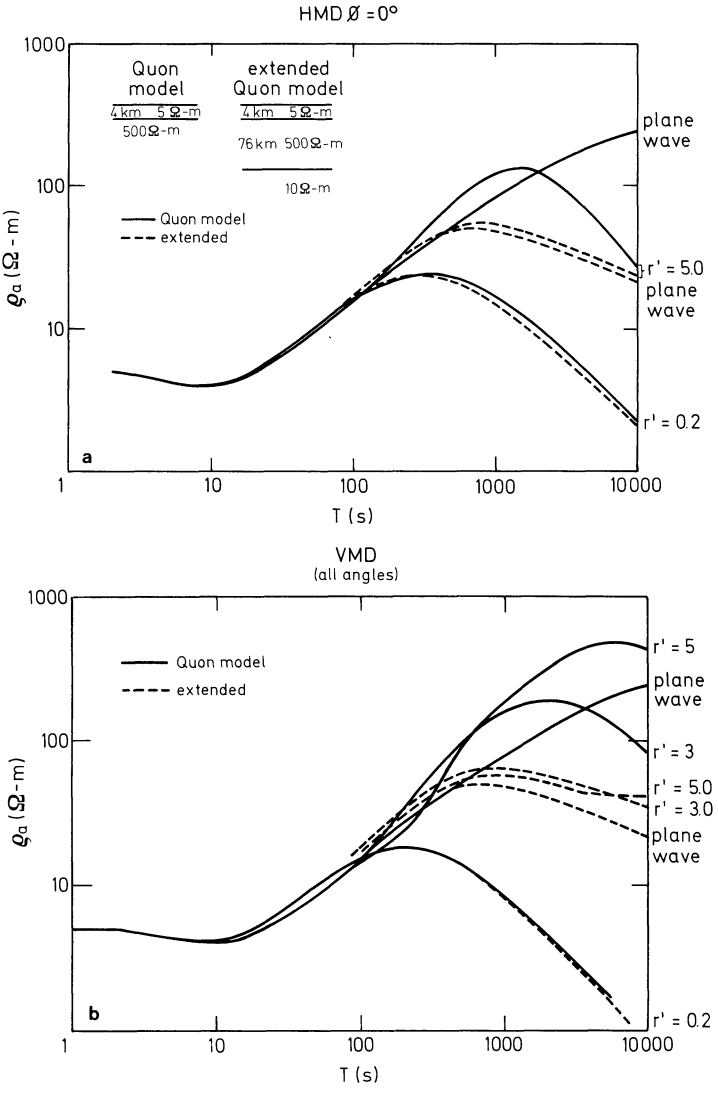


Fig. 4a and b. Plane wave and dipole results over the extended Quon model **a** Horizontal Magnetic Dipole along $\phi = 0°$ using H_x, E_y. **b** Vertical Magnetic Dipole using H_r, E_ϕ

a 10Ω-m mantle at 80km depth. We found that it has considerable influence on VMD and VED results beyond 300s period for large spacing, but makes much less difference to either horizontal dipole (Fig. 4). Therefore we cannot always count on being 'saved' by the mantle from source effects at low frequencies. We note in passing that the field strength due to a dipolar source falls off rapidly with horizontal distance. If source effects are to occur in the presence of plane waves, they must be most evident at small distances from their source.

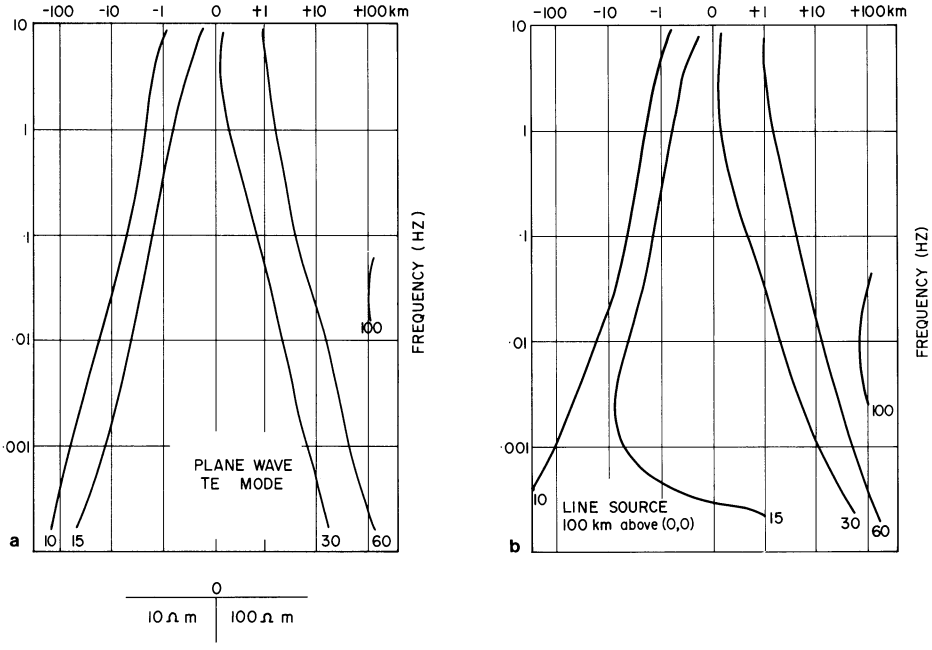


Fig. 5a and b. Apparent resistivity pseudosections for **a** plane wave, *E* parallel to strike, and **b** line current source parallel to strike at 100 km elevation, over the Swift model. Contours are in ohm-meters

New Calculations – Swift Model

While there has been some speculation, there have been no published computations of source effects in more complicated conductivity structures. The simplest of these structures is the vertical exposed fault, and the model shown (bottom, Fig. 5a) is that used by Swift in his early 2D modelling (1967). The question to be examined is whether lateral conductivity variations make much difference in the importance of source effects.

Calculations were carried out for a line source parallel to strike and for horizontal and vertical magnetic dipoles at 100 km above the fault. When the plane wave and the line source apparent resistivities were compared, we found that there were 50% differences at periods of 3000 s, and very little difference at periods less than 100 s (Fig. 5). We also saw that the differences were greatest directly beneath the source. These should diminish as the source moves laterally away from the fault since effective wavelengths at the fault would increase. (The plane wave has the *TE* polarization, and the currents flow parallel to strike in both cases.)

When the line source is replaced by a magnetic dipole, then the result depends on two horizontal coordinates, as did Quon's. It is also evident that, depending on position and dipole orientation, *E* fields can be at any angle to strike. Hence both the smooth transition of the *TE* polarization, and the overshoot-undershoot of the *TM* polarization, should be expected in the various

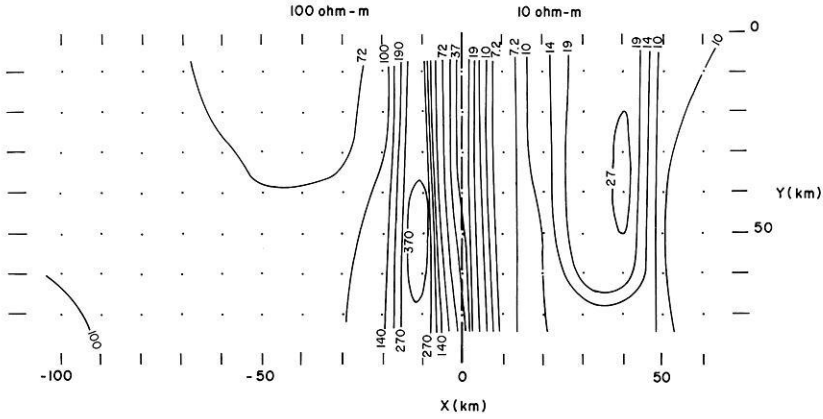


Fig. 6. Plan view of ρ_a due to an x-oriented magnetic dipole above the Swift model. $T = 100$ s, dipole is located at 100 km above (0,0) and ρ_a is calculated from E_x and H_y. Contours are in ohm-meters

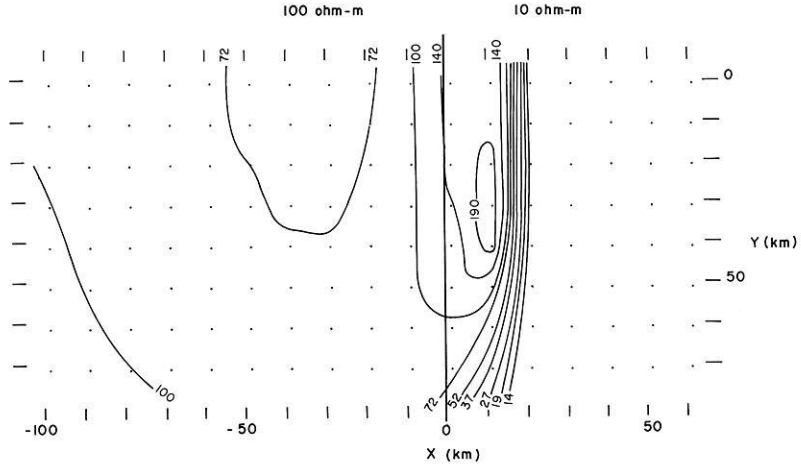


Fig. 7. As Fig. 6, except that the dipole is vertical, and ρ_a is calculated from E_y and H_x. Contours are in ohm-meters

situations. Calculations were done for $T = 100$ s. Larger effects would be expected at longer periods. ρ_{xy} for the x-directed magnetic dipole (Fig. 6; y is strike direction) most clearly shows both overshoot and undershoot. The ratio of extremes (403:6.3) does not quite reach the (100:1) ratio of the TM case, but the values continue to oscillate on the conductive side. The undershoot 10 km on the conductive side is succeeded by a large overshoot at 35 km. The other component, ρ_{yx}, has an overshoot at -80 to -100 km because of the geometric null in the primary magnetic field.

Vertical magnetic dipole results (Fig. 7) at this period are smoother. They show an overshoot (242) but no undershoot. However the overshoot now lies on the conductive side of the contact.

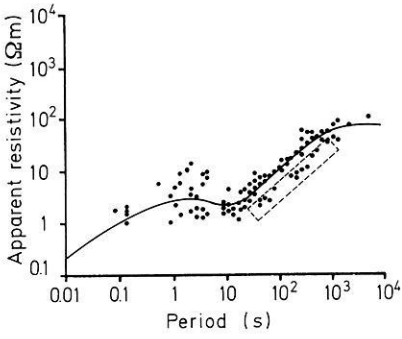


Fig. 8. Rotated tensor ρ_a from a remote site in a simple sedimentary basin. The values in the box appear to be affected by a local source

The Search for Source Effects

Although geomagneticians claim that discrete current or charge systems are unlikely at 100 km elevation, the question of the sources of scatter in good, well smoothed *MT* results remains to be answered. In the result of Fig. 8 the scatter in the 1–10 s period region is due to weak signal. However the lower cluster at 30–700 s was obtained in a single recording at a time of large signal. It differs from the other data only in the time at which it was recorded. The site, in central Australia, is remote from artificial interference, over a sedimentary section of 1.5–4 Ω-m, 2.3 km thick.

Beahn (1976) looking with high precision at small differences between two closely-spaced sites, detected occasional events that had the characteristic of source effects. Large source effects at 500–1,000 s period should be expected at auroral latitudes, where Gokhberg (personal communication, 1976) describes large differences between sites 50 km apart, and several studies of the induction effects associated with the equatorial electrojet can be found (Ducruix et al., 1977). The localized nature of the source currents is central to those results.

Some of the induction effects now ascribed to current channelling, i.e., by 3D conductivity structures – could be caused by combinations of local sources. Consider two discrete, partly coherent sources which are responsible for the *E* and *H* fields at an observing site, and assume that the impedances have different source effects. Then, as the field components add vectorially, the net impedances will vary with the component, and with the relative contribution from each source. If the sources are transitory, then part of the time some components may even cancel, so that null or infinite impedances could be observed for short intervals.

It seems that *MT* could in fact become a valuable technique for studying local heterogeneities in the lower ionosphere, and plans are under way to carry out such an experiment in the auroral zone. An obvious way to approach the subject is with the aid of field intensity and phase distribution measurements at the surface.

As regards the exploration applications of *MT*, if we are fortunate then source effects will remain a curiosity which can be removed, except possibly very near the magnetic equator and the auroral zones. However if we are unfor-

tunate, source effects could be as disruptive to *MT* surveys as magnetic storms are to magnetic surveys.

References

Beahn, T.J.: Geomagnetic field gradient measurements and noise reduction techniques in Colorado. J. Geophys. Res. **81**, 6276–6280, 1976

Cagniard, L.: Basic theory of the magnetotelluric method of geophysical prospecting. Geophysics **18**, 605–635, 1953

Cantwell, T.: Detection and analysis of low frequency magnetotelluric signals. PhD Thesis, MIT, 169 pp. 1960

Dmitriev, W.I., Berdichewsky, M.N.: A new view of Tikhonov-Cagniard magnetotelluric model. Preprint (5 pp). of paper given at the Fourth Workshop on Electromagnetic Induction in the Earth and the Moon, IAGA, Murnau, Federal Republic of Germany, September 1978

Ducruix, J., Courtillot, V., Le Mouël, J.L.: On the induction effects associated with the equatorial electrojet. J. Geophys. Res. **82**, 335–351, 1977

Gamble, T.D., Goubau, W.M., Clarke, J.: Magnetotellurics with a remote reference. Geophysics **44**, 53–68, 1979

Madden, T.R., Nelson, P.: A defense of Cagniard's magnetotelluric method. Tech. Report, ONR Project NR-371-401, MIT, 43 pp, 1964

Price, A.T.: The theory of magnetotelluric methods when the source field is considered. J. Geophys. Res. **67**, 1907–1918, 1962

Quon, C.: Electromagnetic fields of elevated dipoles on a two-layer earth. MSc Thesis, University of Alberta, 250 pp., 1963

Swift, C.M.: A magnetotelluric investigation of an electrical conductivity anomaly in the southwestern United States. PhD Thesis, MIT, 22 pp., 1967

Received March 22, 1979; Revised Version August 28, 1979; Accepted August 31, 1979

J. Geophys. 46, 301–317, 1979

Journal of Geophysics

Long-Wavelength Magnetic Anomalies as a Source of Information About Deep Crustal Structure

Z.A. Krutikhovskaya and I.K. Pashkevich

Institute of Geophysics, Department of the Earth's Magnetic Field, Ukrainian Academy of Sciences, Kiev, USSR

Abstract. The nature of long-wavelength magnetic anomalies ($\lambda = 60$–300 km) computed for the Ukrainian Shield from the original field by a continuation upward to a height of 10 km is studied. The correlation betweeen the regional anomalies, the crustal thickness and the topography of the Curie isotherm of magnetite is examined. The strongest correlation is established between the regional anomalies and the crustal thickness. Similar results have been obtained by us for the Baltic Shield and by D.H. Hall for the Canadian Shield. It is concluded that the entire lower crust is magnetized, the average magnetization being almost the same for all the ancient shields. These results have been used for the construction of a magnetic model of the earth's crust.

With some exceptions, the magnetization of the lower crust is found to be inhomogeneous and 5–10 times higher than that of the upper crust. Theoretical modelling and experimental results show a high magnetization in the entire sequence of blocks for the case of a thickened crust and, vice versa, weakly magnetized rocks correspond to a smaller thickness of the crust. The present approach may be of potential use for distinguishing and studying crust-upper mantle interaction areas as well as for predicting the topography of the Moho discontinuity.

Key words: Magnetic anomalies – Deep crustal structure – Magnetization of earth's crust.

1. Introduction

The problem of the nature of long-wavelength geomagnetic anomalies has been discussed for many years. Numerous publications offer different explanations for the anomalies with wavelengths of 60 to 1,600 km. These anomalies usually have been explained by a superposition of shallow sources, by the existence of local magnetized bodies, by anomalous thermodynamic conditions in the

0340-062X/79/0046/0301/$03.40

lower crust and upper mantle, or by the topography of the Curie isothermal surface of magnetite. A complete list of corresponding references may be found in Green (1976) and Krutikhovskaya (1976).

It is only in recent years that convincing data have been obtained for the Canadian and Ukrainian Shields indicating the relation of the long-wavelength anomalies to deep structure of the crust and upper mantle (Hall, 1974; Krutikhovskaya and Pashkevich, 1974; Krutikhovskaya, 1976). Conclusions have been drawn about a high and inhomogeneous magnetization of the lower crust and about possible applications of magnetic field studies to the solution of deep geology problems.

In the present paper the results of the long-wavelength magnetic component studies for the Ukrainian Shield are discussed. The studies have been carried out at the Institute of Geophysics of the Ukrainian Academy of Sciences.

The Ukrainian Shield was chosen for the investigation as the most favourable geologic region. The shallow Precambrian basement and the availability of data on the distribution of geologic units and on their magnetization make it possible to estimate the high-frequency anomaly effect and provide reliability to separating the magnetic component related to the deep crust.

It is very important that deep seismic sounding (Sollogub and Chekunov, 1975) and heat flow (Kutas, 1976; 1977) investigations have made it possible to study the relationship between the long-wavelength anomalies and both the crustal thickness and the topography of the Curie isotherm of magnetite.

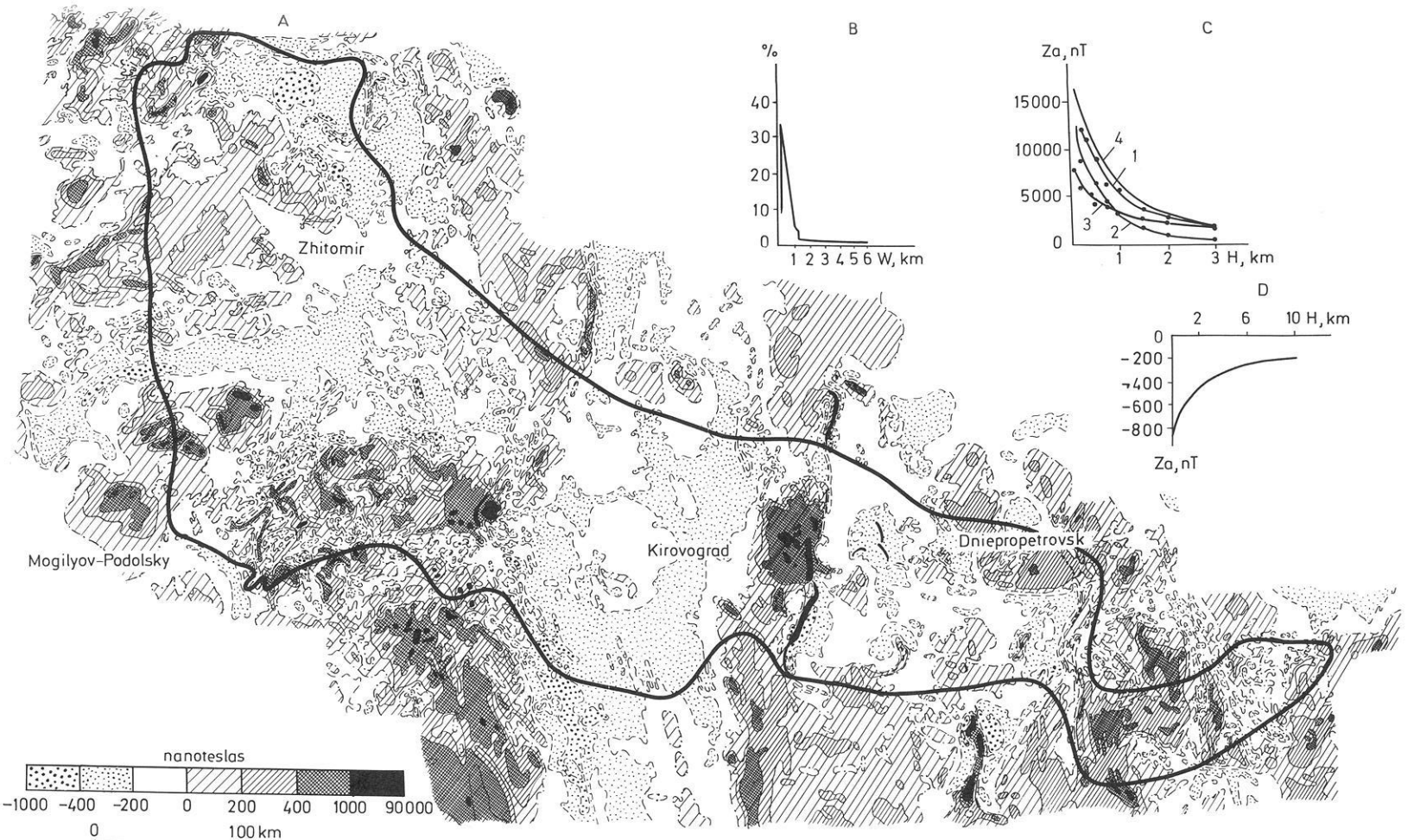
2. Given Data and Data Processing

The anomalous magnetic field of the Ukrainian shield region (Fig. 1a) reflects the block structure of the shield and varies from block to block in number, size, intensity, and dominant strike of the more local anomalies. As shown in Fig. 1b the main contribution to the magnetic field is made by anomalies with widths of no more than 1 km. On the other hand, the magnetic field includes a long-wavelength component with wavelengths of more than 60 km, the existence of which is confirmed by the zero-correlation radius value of the autocorrelation function. These anomalies are called regional ones by us.

The regional anomalies are separated from the local ones by upward continuation of the magnetic field to a certain height. This kind of filtering, compared with others, has some advantages. The main advantage is the possibility to study the attenuation of the amplitudes of the regional anomalies with increasing height of continuation and to compare them with that of the local anomalies.

For the separation of the regional and local anomalies an optimum height should exist where the local anomalies are negligibly small whilst the regional anomalies which are of course also attenuated are still as big as possible.

Fig. 1. A Map of the observed anomalous magnetic field (vertical component Z_a) for the Ukrainian Shield region (border of shield indicated by *thick line*); **B** Frequency distribution of widths of the anomalies shown in part A; **C** Dependence of strongest anomalies on height: 1–3 aeromagnetic survey data, for different parts of the Krivoy-Rog synclinorium; 4 calculated data, for Kremenchug synclinorium; **D** Like C, but for a conjugate negative anomaly from the Verkhovtsevo region


A
B
%
40
30
20
10
0
1 2 3 4 5 6 W, km
C
Za, nT
15000
10000
5000
0
4
1
3
2
1
2
3 H, km
D
2
6
10 H, km
0
-200
-400
-600
-800
Za, nT
Zhitomir
Mogilyov-Podolsky
Kirovograd
Dniepropetrovsk
nanoteslas
-1000 -400 -200 0 200 400 1000 90000
0
100 km

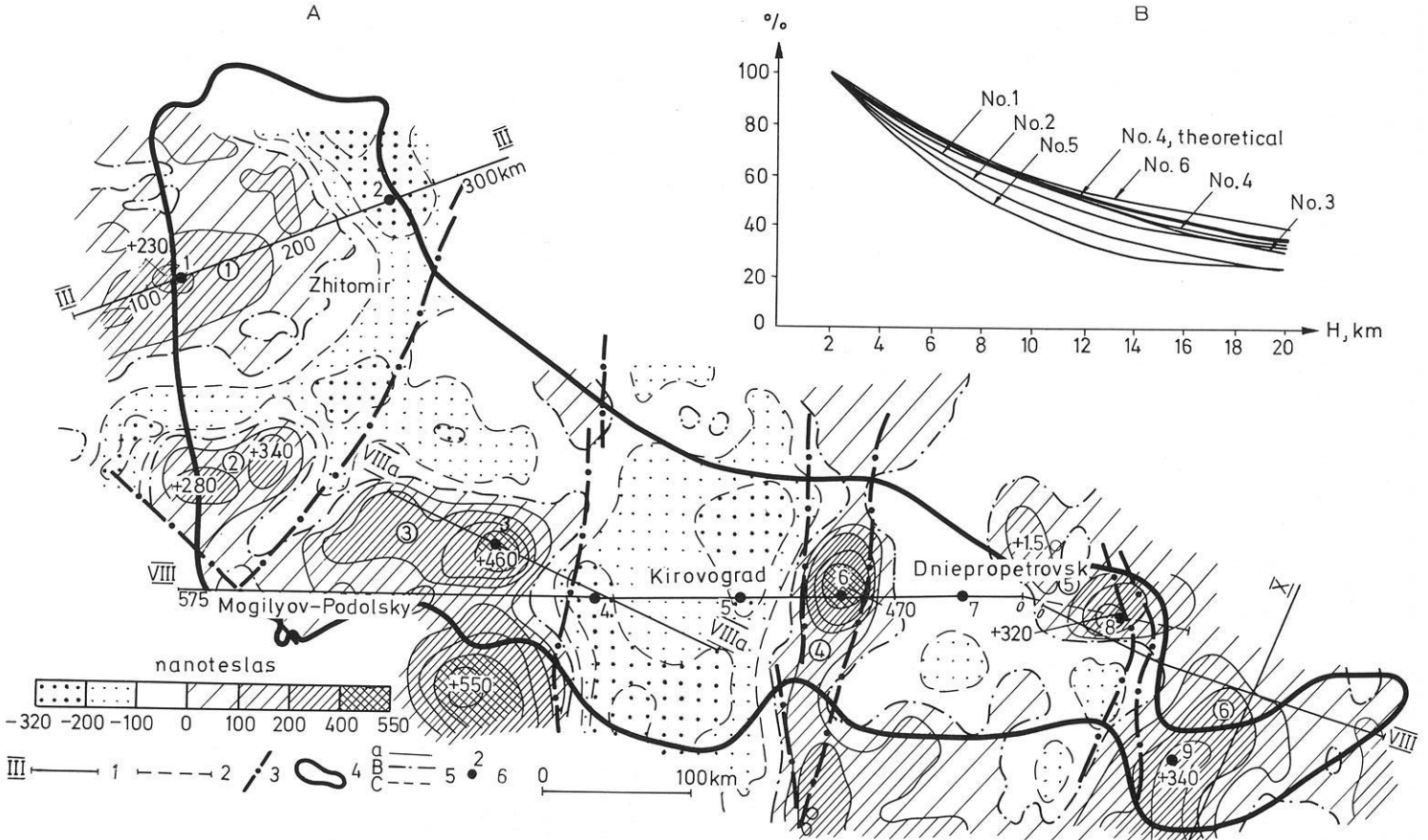
In Fig. 1c the attenuation of the strongest local anomalies within the area of investigation which are situated above the iron-bearing rocks of the Krivoy-Rog-Kremenchug synclinorium is shown. This attenuation is derived from the data of the aeromagnetic survey at different heights and from modelling results. As one can see at a height of 3 km even the biggest anomalies 5 km wide are reduced to 7%–17%. On extrapolating the rate of attenuation and taking into account the modelling data we have found that such anomalies disappear almost completely at a height of 10–15 km. It is reasonable to expect that, at this height, smaller and weaker anomalies should be completely suppressed.

Considering these facts, a height of 10 km was chosen as an optimum level for the field separation. For the purpose of upward continuation the original data were selected over a 2×2 km grid. Such a grid makes it possible to retain the main features of the original field. Anomalies exceeding $\pm 1000\ nT$ – that is a threefold value of the field dispersion – were subtracted from the observed field in advance. However, as studies show, a quite large distortion of the regional anomaly field may be caused by conjugate wide weak lows. They are weaker than $1000\ nT$, but their widths are many times larger than those of the positive part of the anomalies. These lows are attenuated with height much slower than are the conjugate highs (Fig. 1c and d), and the rate of attenuation is commensurable with that of the regional component (Fig. 2b). Therefore when calculating the field at a height of 10–20 km it appears practically impossible to get rid of them (Pashkevich, 1976). The error of field continuation arising from the finiteness of integration varies from block to block and does not exceed $80\ nT$.

3. Results of Data Processing and Comparison to Other Geophysical Results

The derived Ukrainian Shield regional anomalies at a height of 10 km are represented in Fig. 2. Six positive regional anomalies with maxima up to $550\ nT$ and with wavelengths from 60 to 300 km are found above the shield. In the upper right-hand corner of the figure the attenuation of these anomalies is shown. Compared with local anomalies (Fig. 1c) the rate of attenuation is much smaller. These anomalies were assumed to be produced by magnetized bodies at greater depths. The results of our interpretation show that this assumption corresponds well to our knowledge about the amount of magnetization of basic crystalline rocks and about the block structure of the lower crust (see below and Figs. 7 and 8). The negative part of the field includes lows the minimum of which reaches $-340\ nT$. As may be seen, almost all of the

Fig. 2. A Map of the derived regional magnetic anomalies above the Ukrainian Shield. *Legend:* 1: deep seismic sounding lines; 2: eastern section of the line for which the magnetic model has been constructed; 3: deep-seated faults; 4: border of shield (cf. Fig. 1A); 5: magnetic field isolines (*a*: positive; *B*: zero; *C*: negative); 6: points used for study of correlation between regional field values and crustal thickness. **B** Dependence of regional anomalies (cf. *encircled numbers within part A*) on height. The *curve labelled No. 4, theoretical* has been calculated for a three-dimensional model of the source of this anomaly


A
B
%
100
80
60
40
20
0
2 4 6 8 10 12 14 16 18 20
H, km
No.1
No.2
No.5
No. 4, theoretical
No. 6
No. 4
No.3
III
300km
+230
200
100
Zhitomir
+340
+280
VIIIa
+460
VIII
575
Mogilyov-Podolsky
Kirovograd
470
Dniepropetrovsk
+1.5
+320
+550
nanoteslas
-320 -200 -100 0 100 200 400 550
100km
+340

regional highs are located in marginal zones of the shield or above deep-seated faults separating large blocks. An analysis of these anomalies in combination with deep seismic sounding data (Fig. 3) and with results of the upper crust magnetization studies shows that, in the area of the regional highs, the earth's crust is characterized by a higher magnetization of the basement rocks, a greater thickness and the presence of strong reflectors within the mantle ('duplicate'-surfaces according to Sollogub and Chekunov (1975)). According to Chekunov (1976) the layered character of the Mohorovicic discontinuity is caused by crust-mantle merging in the transition zone.

All the areas of an intensive increase of crustal thickness are confined to zones of deep-seated faults. Therefore it is natural to suppose that the magnetic highs are caused by highly magnetized rocks of basic and ultrabasic composition intruding into the fault zones. On the other hand, the correlation of the magnetic anomalies with crustal thickness holds as well for the areas where the M surface is being uplifted. These areas are characterized by minor magnetic anomalies and a predominance of relatively weak magnetization values of the upper crust rocks.

The statistical correlation between the regional anomalies and the crustal thickness as derived from deep seismic sounding is shown in Fig. 4a–d. The correlation coefficients range from 0.70 to 0.84. A higher value from 0.86 to 0.92 is obtained for the relict M surfaces represented by strong mantle reflectors found at depths of 50 to 70 km under the seismic profile VIII.

In Fig. 4e the general correlation between regional anomalies and crustal thickness as established for the whole area of the Ukrainian Shield is shown. The correlation coefficient is computed from the extremum values of the magnetic field and corresponding values of crustal thickness. The corresponding points are shown in the Figs. 2 and 3, respectively. The correlation coefficient amounts to 0.68. This low value may be explained by the following fact: Seven anomalies out of nine correlate with the Moho topography. The value of the correlation coefficient grows up to 0.97 if the points Nos. 6 and 9 are left out of consideration. Any relationship between the regional anomalies and the crustal thickness is not found for these points. There are reasons for excluding these points from the calculation. They belong to anomalies Nos. 4 and 6 (cf. numbers encircled in Fig. 2). A deep origin of the anomaly No. 4 has been proved by Krutikhovskaya et al. (1973). However, the deepest trough in the Moho surface is restricted to the eastern part of the anomaly. There is another version of the deep seismic sounding data interpretation for this region (Pavlenkova, 1973). Therefore, a further detailed study of the position of the crustal bottom is necessary here.

In the case of anomaly No. 6, the position of the Moho discontinuity is not convincingly defined here because the deep seismic sounding profiles cross only the marginal parts of the anomaly (see Fig. 2). Therefore the source of the anomaly is still to be studied.

In Fig. 4f the correlation between the regional magnetic anomalies and the crustal thickness is systematically shown for different ancient shields. For the Soviet part of the Baltic Shield the data were obtained by us along the Sortovala-Belomorsk profile. For the western part of this Shield we have used the data

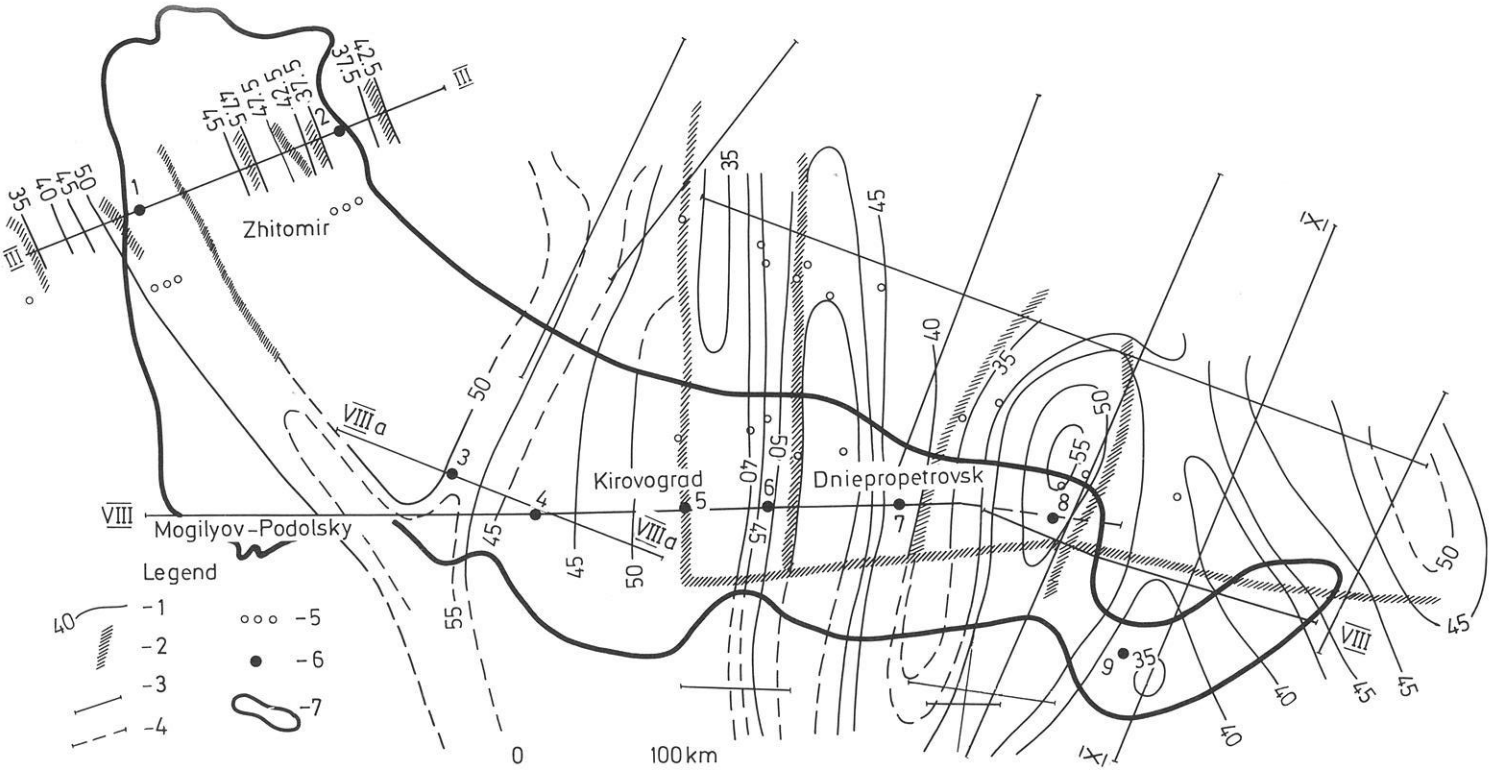


Fig. 3. Map of the Moho topography within the Ukrainian Shield region for the end of the early Proterozoic, after Sollogub (1975). *Legend:* 1: lines of equal depth (as indicated by numbers, in km); 2: faults in the *M* discontinuity; 3: deep seismic sounding lines; 4: eastern section of the line for which the magnetic model has been constructed; 5: points where the depth to the *M* discontinuity has been determined; 6: points used for study of correlation between the regional magnetic field and the depth of the *M* discontinuity; 7: border of shield (cf. Fig. 1A)

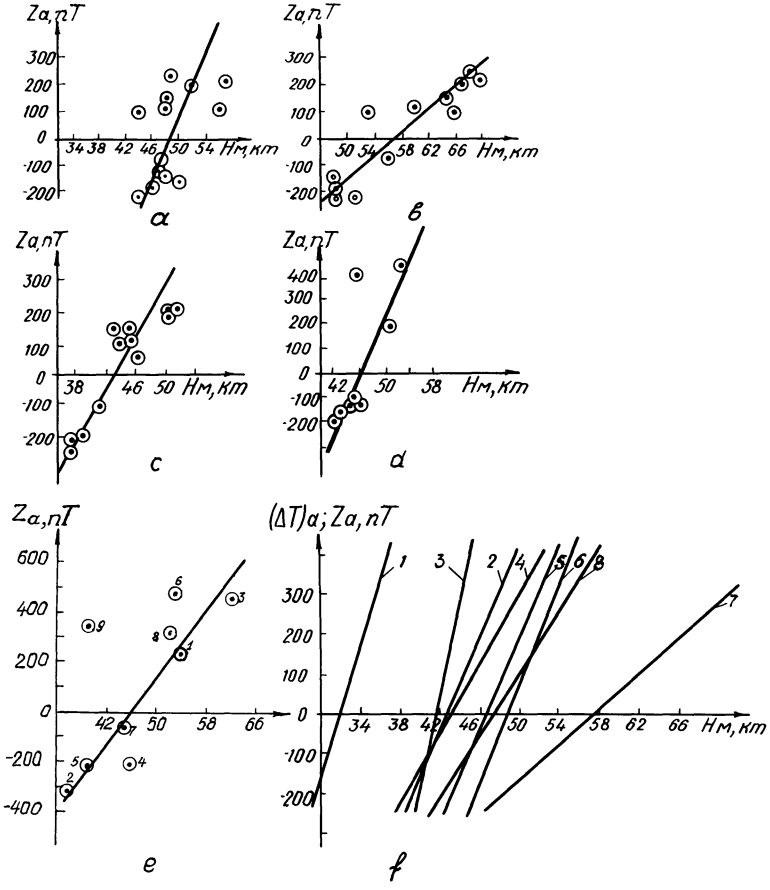


Fig. 4. Graphs showing the correlation between the regional magnetic field (Z_a or ΔT_a) and the crustal thickness (H_M); *a*: deep seismic sounding line VIII, western section; *b*: same line, but for lower positions of *M* discontinuity (horizon B of Fig. 7); *c*: seismic sounding line III; *d*: deep seismic sounding line VIIIa; *e*: whole area of Ukrainian Shield; *f*: ancient shields, with 1: Canadian Shield (Hall, 1974); 2: eastern part of Baltic Shield; 3: western part of Baltic Shield (data from (Vogel and Lund, 1970)); 4–7: parts of Ukrainian Shield, (4 corresponds to part c, 5 to part d, 6 to part a, 7 to part b of this figure); 8: whole Ukrainian Shield (generalized)

reported by Vogel and Lund (1970) for the Trans-Scandinavian deep seismic sounding profiles and the regional magnetic anomalies map published by Riddihough (1972). For the Canadian Shield the correlation has been presented by Hall (1974). A consideration of these data leads to the conclusion that the lower crust of shields is composed of magnetized rocks in such a way that the anomalous magnetic field reflects the Moho topography. However, an analysis of our results for the Ukrainian Shield shows that the regional anomaly No. 4 (Fig. 2) exhibits an exception from this rule. It is located at the border between Precambrian blocks of different age.

The stated statistical correlation between regional anomalies and crustal thickness raises, above all, the question whether this connection is functional

that is, whether the M boundary separates magnetic and nonmagnetic rocks such as, for example, hyperbasite and altered hyperbasite. This suggestion has been checked by various methods, especially by an analysis of the temperature distribution data reported for the Ukrainian Shield crust by Kutas (1976; 1977) and Gordienko (personal communication). The results of the two authors are based on the same original data on heat flow, but when calculating the temperature field different values of the mantle heat flow were adopted. Accordingly, both authors arrived at different results on the topography of isotherms within the crust and, especially, the upper mantle. However, the temperature ranges derived for the M surface were almost the same. They varied from 320°–350° to 540°–575°.

The heat flow change along the deep seismic sounding line VIII, and the corresponding temperature distribution within the crust are shown in Fig. 5. The accuracy of the temperature determination is indicated by hatching. Apparently, a temperature of about 400° C corresponds to the average depth of 40 km of the Moho discontinuity. Both versions of the 575° C isotherm shown suggest that, almost everywhere, the Curie point isotherm of magnetite is estimated to be situated below the Moho. Such thermal conditions justify the assumption that both the lower crust and upper mantle may be magnetized. Therefore the relationship between the regional magnetic anomalies and the depth to the Curie isotherm of magnetite shown in Fig. 5 was studied. The correlation between them seems to be weaker than the correlation of the regional anomalies with crustal thickness.

Previously, the magnetic effect of the earth's crust has been calculated by us assuming the bottom of the magnetized layer at 40 km depth and a magnetization of 4 A m^{-1} below the Conrad discontinuity (extended reflector).

In order to estimate the effect of the bottom of magnetic masses the thus derived effect was subtracted from the regional magnetic anomalies. It has been found that the residual anomalies correlate with the depths of the Curie isotherm of magnetite according to Gordienko (see above, and Fig. 5), the correlation coefficient amounting to 0.80 (Krutikhovskaya and Pashkevich, 1976). However, for the crustal temperature distribution derived by Kutas (1976; 1977; cf. Fig. 5) the correlation coefficient is only 0.45.

We may conclude, that the problem of the upper mantle magnetization cannot be solved unambiguously with the present accuracy of temperature determinations for the crust and upper mantle. At the present stage of studies the correlation between the regional magnetic field and the crustal thickness established for the Ukrainian and other shield regions does not contradict to the assumption that the bottom of the magnetized masses coincides with the Moho surface.

4. Construction of a Model and Geological Implications

The former conclusion has been accepted by us as a basis for constructing a magnetic model of the earth's crust of the Ukrainian Shield. For this purpose the magnetization of different parts of the crustal section along the deep seismic

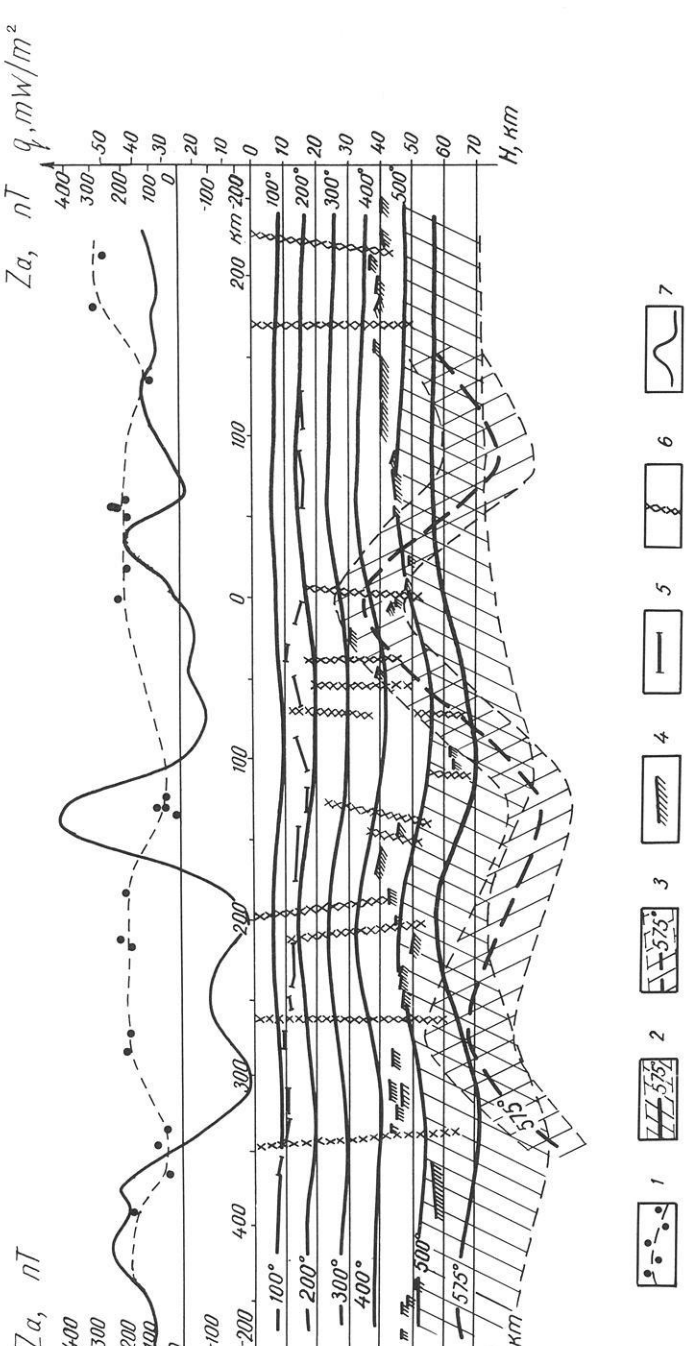


Fig. 5. Regional magnetic field (Z_a), heat flow (data from geothermal line Taganrog-Golovanevsk (Kutas, 1977), crustal seismic reflectors (Sollogub and Chekunov, 1975), and calculated temperature distribution within earth's crust, along the deep seismic sounding line VIII. *Legend*: 1: heat flow; 2: isotherms, partly with errors (*shaded*), according to Kutas (1977); 3: 575° C isotherm (with errors indicated) according to Gordienko (personal communication); 4: *M* discontinuity; 5: seismic boundaries within crust; 6: deep-seated faults; 7: regional magnetic field

sounding lines III and VIII was estimated. The magnetization of the granite-gneissic layer was determined from magnetic susceptibility measurements of the rocks outcropping on the crystalline basement surface and from results of the magnetic anomalies interpretation. The ratio of the remanent magnetization to the induced one as derived from the experimental data is 1. The average magnetization values in different blocks of the Ukrainian Shield range from 0.1 to 0.5 A m^{-1}. The average value for the granite-gneissic layer is extended to depth down to a long reflecting horizon which, probably, marks the lower boundary of the most intense folding of this layer and, correspondingly, the limit depth of the lower edges of the sources of local magnetic anomalies.

In estimating the lower crust magnetization the general results on composition and ferromagnetic content of deep-seated rocks have primarily been used. Most of the investigators in the Soviet Union believe that the lower crust is composed of various sedimentary-volcanic complexes with predominance of basic and intermediate rocks both metamorphosed to the granulite facies, and essentially nongranitized. These rocks probably consist of charnokite, amphibolite, and eclogite-like formations. The general tendency of basicity to grow with depth is confirmed by the corresponding seismic velocity increase.

According to Kalyaev (1976) the ophiolitic association of the Ukrainian Shield representing the ancient oceanic crust is composed of amphibolite, gabbro-amphibolite, gabbro-peridotite, pyroxene gneiss and associated orthoschists alternating with highly metamorphosed pelitic rocks. The high total content of iron in sedimentary, effusive, and intrusive primary rocks of the ophiolitic association has resulted in its separation in the form of magnetite as a result of the subsequent metamorphism. Nalivkina (1976) recognizes several stages in the process of magnetite formation. In the Archean, it had a regional character with transformation of basic rocks from the ophiolitic association into the charnokite association. Several stages are stated by the author for the early Proterozoic as well.

As follows from the data reported by Christencen and Fountain (1975), the lower crust is composed mainly of granulite facies rocks and has lateral inhomogenities in chemical and mineralogical composition.

The observation of the high magnetization of the lower crust corresponds to the results of experiments on the thermodynamic conditions of the occurrence of magnetite. Our conclusion on the magnetization of the lower crust is also supported by the existence of a correlation between density and magnetization of widely developed rocks of various metamorphic facies and by the existence of magnetic inhomogeneities within the lower crust as derived from the interpretation of regional anomalies. By accepting a direct relationship between the seismic velocity and the density of rocks and by basing on the relation between magnetic susceptibility and density, a magnetization increase with depth may be deduced. Numerous experimental and published data have been used in studying that relation. Sedimentary-volcanogenetic rocks of granulite facies and intrusive rocks of normal line have been investigated (Fig. 6). For intrusive rocks, the correlation between susceptibility and density was studied by the SiO_2 content.

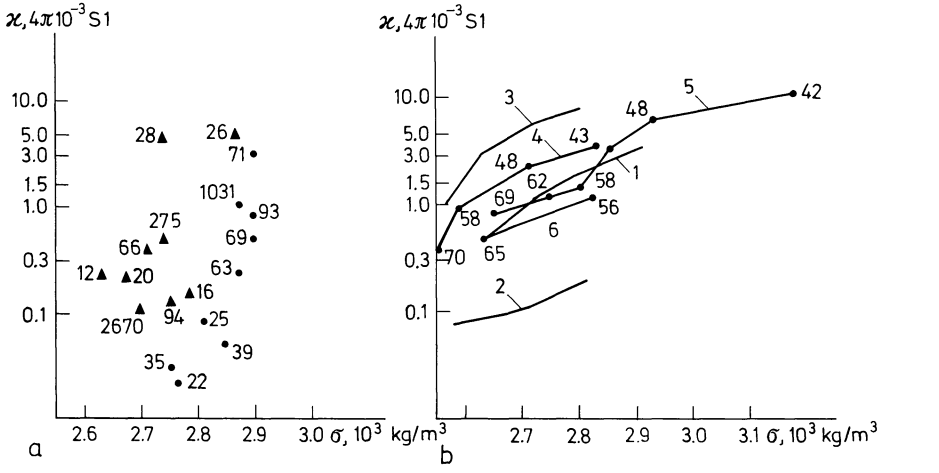


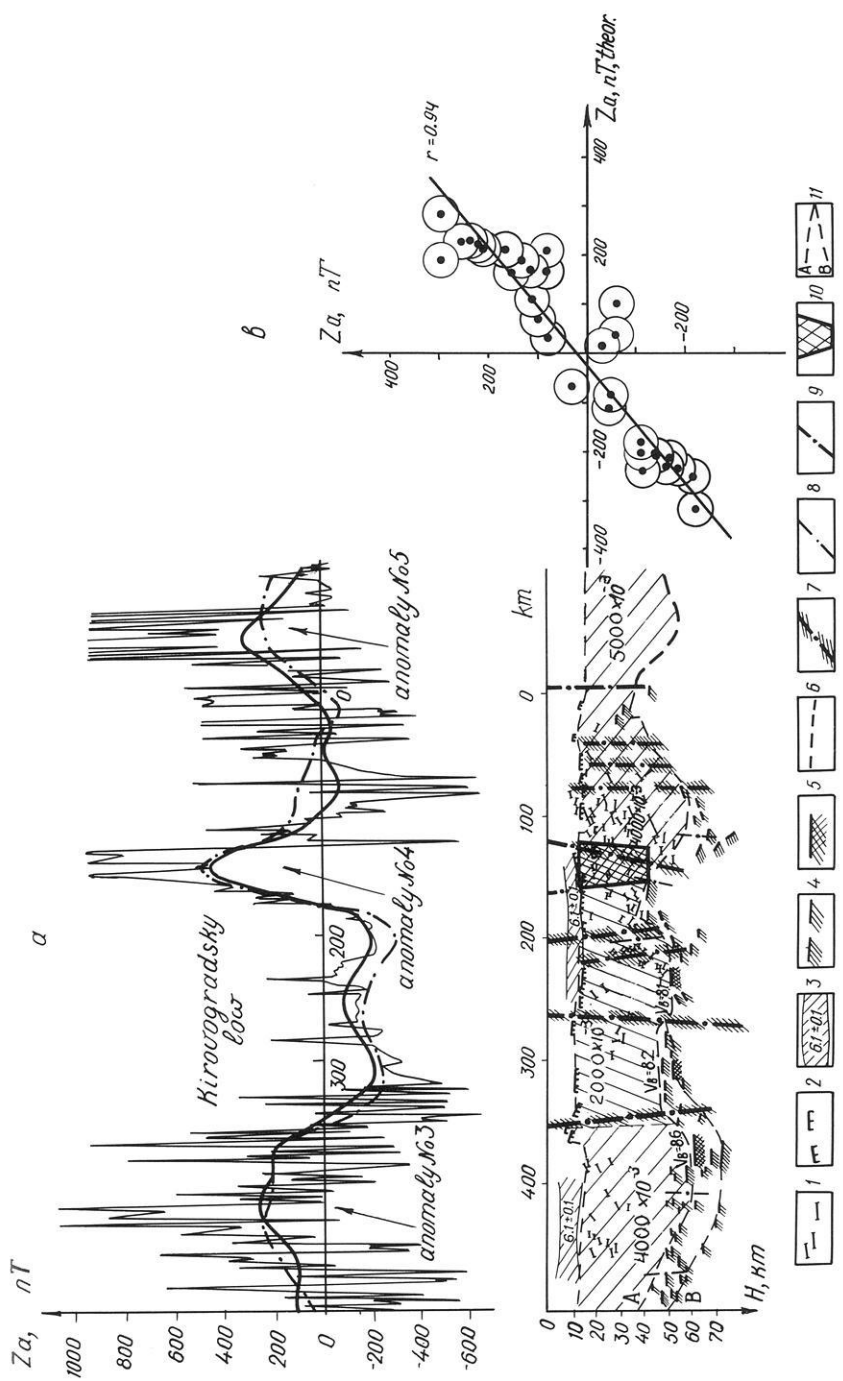
Fig. 6a and b. Graphs demonstrating the correlation between the magnetic susceptibility and the density. **a** Rocks of granulite facies from Ukrainian Shield: ● pyroxene plagioclase gneiss, garnet pyroxene plagioclase gneiss, hypersthene gneiss (1949 specimens); ▲ charnokite (3207 specimens. Average κ- and σ-values for uniform data sets have been used; numbers indicate numbers of specimens which have been averaged. **b** Intrusive rocks: 1: gabbro-diorite-granodiorite formation; 2: granite formation; 3: granite-granosyenite formation (1–3 according to Dortman, 1976) magmatic rocks of normal line according to Semyonova, 1973); 4: abyssal rocks; 5: hypabyssal rocks; 6: effusive rocks. Numbers indicate SiO_2 content (%)

The use of the aforementioned data and the following modelling of the magnetic effect enable us to determine the possible limits of the magnetization values for the lower crust. The average value ranges from 0.3 to 4 A m^{-1}, reaching 6–7 A m^{-1} in certain blocks. The magnetic data thus independently lead to a conclusion about the chemical and mineralogical inhomogeneity of the lower crust.

According to the experimental data the maximum depth at which the formation of the ferromagnetic minerals took place is believed to be 50–60 km (under a pressure of 15–18 kbar). Hence, the magnetization values obtained for the rocks in the lower crust may be considered valid for the entire of its thickness.

Direct modelling included the determination of the magnetic effect of the whole crust along the deep seismic sounding lines III and VIII (Figs. 7 and 8).

Fig. 7. a Magnetic model of the crust along the deep seismic sounding line VIII. *Legend:* 1: crustal reflectors; 2: extended refracting horizon; 3: low velocity layer; 4: *M* discontinuity reflectors; 5: *M* discontinuity refractors (V_B indicates sub-Moho seismic velocity, in km s^{-1}), 7: deep-seated large faults according to deep seismic sounding results; 8: same as 7 but minor faults; 9: faults derived from other geological and geophysical data; 10: deep reaching zone with magnetization 6 A m^{-1}; 11: averaged *M* discontinuity position used for studying correlation with regional magnetic field; *A*: position like in Fig. 3; 12: observed magnetic field (Z_a); 13: magnetic field calculated from the observed field for a height of 10 km; 14: magnetic field calculated for the same height from the magnetic model of the crust; 15: rocks with magnetization 4–5 A m^{-1}; 16: rocks with magnetization 2 A m^{-1}. **b** Graph showing correlation between the magnetic field (Z_a) at 10 km height as calculated from observed field and magnetic field ($Z_{a,\ theor.}$) at same height corresponding to the magnetic model of the crust


a
Za, nT
Kirovogradsky low
anomaly №3
anomaly №4
anomaly №5
km
H, km
5000×10⁻³
4000×10⁻³
2000×10⁻³
Vs=82
Vs=86
61±01
61±19
в
Za, nT
Za, nT, theor.
r=0.94

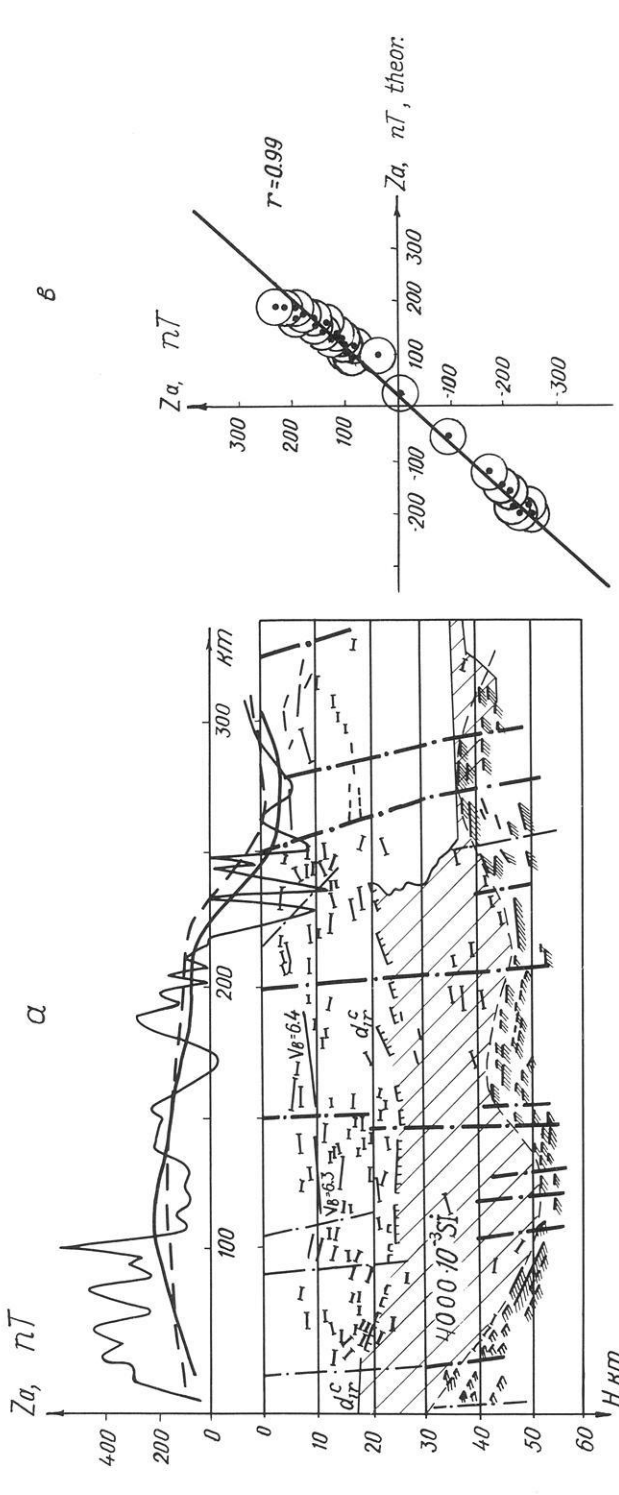


Fig. 8. Same as Fig. 7, but for deep seismic sounding line III. *Legend:* 1: crustal reflectors; 2: reflectors and boundary velocity values (km s^{-1}); 3: extended refracting horizon; 4: Moho; 5: deep-seated large faults from deep seismic sounding results; 6: same as 5 but minor faults; 7–9: same as legend numbers 12–14, respectively, within Fig. 7; 10: rocks with magnetization 4 A m^{-1}; 11: rocks with average magnetization 0.3–0.5 A m^{-1}

Average magnetization values for the upper crustal section of large blocks were estimated, as mentioned, from the experimental data and from results of the magnetic field interpretation. However, as calculations show, the effect of the upper crust at a height of 10 km falls within the error of the regional anomalies separation. Hence, the field produced by the entire crust corresponds actually only to the effect of the lower crust.

The ambiguity in estimating the magnetization has necessitated a discussion of two versions of the magnetic model, with the effective magnetization of large blocks remaining at a constant value.

A magnetization of the lower crust with values relatively smaller than in the adjacent blocks is found under the central part of the deep seismic sounding profile VIII in the area of the Kirovogradsky regional low (Fig. 7). In this block the upper crust is composed of weakly magnetized rocks of various origin including intrusive rocks. A pluton of intrusive rocks of acid and basic composition was formed during the final platform stage of the shield evolution.

The closest correlation between theoretical and original profiles of the regional magnetic anomalies has been obtained with the magnetization distribution within the lower crust as shown in Fig. 7a. The correlation coefficient is 0.94 in this case (Fig. 7b). The correlation is interrupted in the eastern section of the deep seismic sounding line at the location of anomaly No. 5. When modelling we did not intend to reach a complete coincidence of theoretical and original data since too many aspects of structure, petrology and magnetization of rocks as well as thermodynamic conditions within the earth's crust have not been thoroughly studied up to now.

In constructing the magnetic model along the deep seismic sounding line III (Fig. 8a) it is necessary to assume that the magnetization of the lower crust under the eastern section of the line is commensurable with the upper crust magnetization and amounts to $0.5\ A\ m^{-1}$. According to deep seismic sounding data, in the area of the Moho discontinuity rise reflectors are absent in the lower crust and the extended reflecting horizon previously regarded as a boundary between the upper and lower crust is not found. Here, on the Precambrian basement, a large pluton of acid and basic weakly magnetized rocks has been mapped. This pluton, alike to one in the Kirovogradsky block, was formed also during the platform stage of the shield evolution. This suggests that weakly magnetized rocks build up the earth's crust down to a great depth, and that the lower crust is of intrusive nature like the upper crust. The correlation coefficient of the original profile with the theoretical curve for the given distribution of the crustal magnetization is 0.99 (Fig. 8b).

5. Conclusions

The results of the regional anomalies studies for the magnetic field of the Ukrainian Shield lead to the following main conclusions:

Regional magnetic anomalies with a wavelength between 60 and 300 km reflect, primarily, crustal thickness, allowing thus to predict the Moho topography.

The magnetization of the lower crust is not homogeneous and, generally, is considerably larger than the magnetization of the upper crust. An abnormally high magnetization of the entire crustal sequence is observed in zones with a thicker crust. On the basement surface, these zones are represented by the products of metamorphism and granitization of the most ancient predominantly magmatic and sedimentary-volcanic rocks of ophiolitic association.

Blocks with a thin crust are characterized by the lowest magnetization. The upper parts of these blocks are mostly represented by the products of metamorphism and granitization of the pelitic rocks and by plutons of gabbro-anorthosite and rapakivi granite. The lower crust in these blocks is likely to be composed of andesite and anorthosite.

The close correlation between the magnetization of the upper and lower crust reflects the processes of the crust-upper mantle interaction. Taking into account the density-magnetization correlation on the one hand, and the interrelation between gravity and magnetic anomalies on the other hand, the conclusion may be drawn that the earth's crust of the Ukrainian Shield is isostatically balanced. The 'heavy' crust corresponds to the Moho surface depressions and the 'light' crust to the Moho uplifts. An analysis of the results obtained for the Ukrainian and Baltic Shields and the published data on the Canadian shield leads to the conclusion that, in general the types and major features established for the earth's crust of the ancient shields are similar.

The problem of the nature of the regional anomalies is of utmost interest and requires further investigations. This is especially important in connection with the application of the plate tectonics concept to studying the continental crust evolution.

Acknowledgements. We wish to thank T.V. Babenko, A.A. Garbuza, A.Ph. Demut, and L.D. Klivadenko from the staff of the Geomagnetic Field Department, Institute of Geophysics, Ukrainian Academy of Sciences, for the help in preparation of the paper.

References

Chekunov, A.V.: Earth's crust of the Ukrainian Shield and some important points in formation of the ancient platform basement (in Russian). Dokl. Akad. Nauk SSSR, Ser. B. No. **10**, 893–896, 1976

Christencen, N.I., Fountain, D.M.: Constitution of the lower continental crust based on experimental studies of seismic velocities in granulite. Geol. Soc. Am. Bull. **86**, 227–234, 1975

Dortman, N.B.: Effect of thermodynamical conditions of formation on magnetic characteristics of the intrusive rocks (in Russian). In: Magnetic anomalies of the Earth's interior. Kiev: Naukova Dumka 1976

Green, A.G.: Interpretation of project Magnet aeromagnetic profiles across Africa. Geophys. J. **44**, 203–228, 1976

Hall, D.H.: Long-wavelength aeromagnetic anomalies and deep crustal magnetization in Manitoba and Northwestern Ontario, Canada. J. Geophys. **40**, 403–430, 1974

Kalyaev, G.I.: The most ancient stage of continental crust formation (in Russian). Geofiz. Sbornik, No. **73**, 56–60, 1976

Krutikhovskaya, Z.A.: Problems of creating the earth's crust magnetic model for ancient shields (in Russian). Geofiz. Sb. No. **73**, 3–29, 1976

Krutikhovskaya, Z.A., Pashkevich, I.K., and Simonenko, T.N.: Magnetic anomalies of Precambrian shields and some problems of their geological interpretation. Can. J. Earth Sci. **10**, 629–636, 1973

Krutikhovskaya, Z.A., Pashkevich, I.K.: Magnetic field and structure of the earth's crust for ancient shields (in Russian). Geofiz. Sb. No. **62**, 3–15, 1974

Krutikhovskaya, Z.A., Pashkevich, I.K.: Is the crust of the ancient shields and platforms magnetized? (in Russian). Geofiz. Sb. No. **71**, 89–91, 1976

Kutas, R.I.: Temperature distribution in the earth's crust for the territory of the Ukrainian SSR (in Russian). Geofiz. Sb. No. **70**, 86–94, 1976

Kutas, R.I.: A thermal model for the earth's crust of the Ukrainian Shield along DSS line VIII (in Russian). Geofiz. Sb. No. **80**, in press, 1977

Nalivkina, E.B.: Petrological model of the Baltic Shield eastern part (in Russian). Geofiz. Sb. No. **73**, 46–55, 1976

Pashkevich, I.K.: Procedure of distinguishing and interpretation of regional magnetic anomalies (in Russian). Geofiz. Sb. No. **73**, 30–36, 1976

Pavlenkova, N.I.: Velocity fields and earth's crust model (in Russian). Kiev: Naukova Dumka 1973

Riddihough, R.P.: Regional magnetic anomalies and geology in Fennoscandia. A discussion. Can. J. Earth Sci. **9**, 219–232, 1972

Semyonova, S.G.: Relationship between physical properties, mineralogical composition and structural factors (in Russian). Geofiz. Razved., No. **57**, 128–136, 1973

Sollogub, V.B., Chekunov, A.V.: Deep structure and earth's crust evolution (in Russian). In: The earth's physics problems in the Ukraine. Kiev: Naukova Dumka, 118–141, 1975

Vogel, A., Lund, C.E.: Combined interpretation of the Trans-Scandinavian seismic profile, section 2–3, University of Uppsala, Report 4, Uppsala, 1970

Received August 28, 1977; Revised February 27, 1979; Accepted August 14, 1979

J. Geophys. 46, 319–334, 1979

Journal of Geophysics

Palaeomagnetism and the Early Magmatic History of Fuerteventura (Canary Islands)

K. M. Storetvedt, H. Mongstad Våge, S. Aase, and R. Løvlie

Department of Geophysics, University of Bergen,
Allégt. 70, N-5014 Bergen-Universitetet, Norway

Abstract. Thermal and alternating field demagnetization combined with studies of the convergence points of remagnetization circles have been carried out on a variety of rocks from Fuerteventura. The older (post-Albian) rocks, including the basement plutonics, the sheeted dike complex and the earliest subaerial lava sequence (lava Series I), have a multicomponent remanence while the younger lava Series II of Miocene/Pliocene age has dominantly a one-component magnetization. Comparison of the results with recent palaeomagnetic data from Gran Canaria/Tenerife and with results from continental Africa and Europe suggest a late Cretaceous origin of the basal intrusive rocks while the subaerial volcanism (lava Series I) most likely initiated at around the Cretaceous/Tertiary boundary. This implies that there is a nearly 50-m.y.-long period of volcanic quiescence and erosion between lava Series I and lava Series II. The apparent contradiction between these conclusions and the few K/Ar dates available is discussed.

Key words: Palaeomagnetism – Age of early Fuerteventura magmatism.

1. Introduction

The oldest exposed rocks in Fuerteventura are found in the western part of the island, the so-called Betancuria massif (Fuster et al., 1968; Rothe, 1968) (Fig. 1). This basement complex, which Gastesi (1973) believes to be a piece of uplifted oceanic crust, has a core of layered basic and ultra-basic plutonics which in part occurs in faulted contact with Lower-Middle Cretaceous terrigenous and calcareous clastic, deep-water sediments of basically turbiditic origin (Robertson and Stillman, 1979). Upwards, the sedimentary sequence, after a hiatus, passes into submarine volcanics and volcaniclastic sediments. This rock association is in turn cut by a very dense dyke system followed by plutonic intrusions with a second generation of dykes, a carbonatite sequence and finally by a syenitic and trachytic ring dyke system. After an unconformity (Stillman

0340-062X/79/0046/0319/$03.20

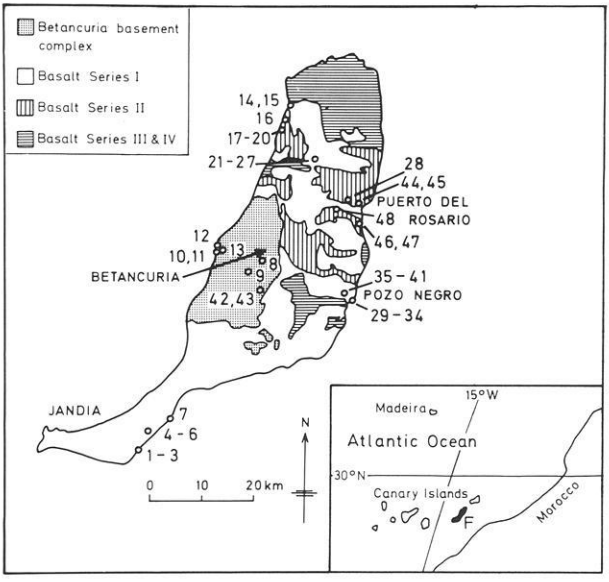


Fig. 1. Simplified geological sketch map of Fuerteventura showing distribution of palaeomagnetic sampling sites.

et al., 1975), four series of subaerial basaltic lavas (Series I–IV; Fuster et al., 1968) were laid down.

A fairly strong secondary alteration has affected all rocks older than lava Series I. Fuster et al. (1968) and Stillman et al. (1975) ascribe these mineral changes to a regional metamorphism equivalent to the greenschist facies. In the Betancuria massif the tectonic situation is complex but at least the Cretaceous sedimentary succession and in part the overlying submarine volcanics have been involved in a relatively strong deformation. The timing of this phase of margin unrest, pre-dating the subaerial lava sequences (Series I–IV), has been a matter of speculation, not least because the existing radiomatric age determinations from the older Fuerteventura volcanics in part are inconsistent and difficult to interpret (cf. Chap. 5). Against these uncertainties an age of Middle Miocene – Pliocene for basalt Series II appears fairly well established (Fuster et al., 1968; Rothe, 1966).

Palaeomagnetic results may provide alternative information on the age problem, but unfortunately the data published so far (Watkins et al., 1966; Watkins, 1973) are not sufficiently detailed to be of much help in this context: only a modest laboratory treatment was performed and discordant directions as well as a marked decrease in the precision parameter (k) towards the base of the subaerial lava succession (Series I) were left unexplained.

In the present study the encountered palaeomagnetic problems are ascribed to unresolved multicomponent magnetizations. After having considered these problems and hopefully estimated the directions of single remanence components the results are compared with recent palaeomagnetic data from Gran Canaria/Tenerife (Storetvedt et al., 1978), and conclusions regarding the early tectono-magmatic evolution of Fuerteventura are drawn.

Table 1. Rock collection details

Site	Sample nos.	Rock type/ series	Locality
1–3	14	Series I	Beach at Cases del Butihondo
4–6	15	Series I	Between Montana Blanca and the sea, ca. 100 m above sea level
7	5	Series I	Beach south of Montana Bay
8	5	Trachyte	Vega de Rio Palmas
9a–e	15	Dyke swarm	Morro Fenduca
10	4	Sediment	Puerto de la Pena
11	4	Dyke	Puerto de la Pena
12	5	Series II	Puerto de la Pena
13a–e	14	Dyke swarm	Puerto de la Pena
14	5	Series I	Toston Cotillo
15	4	Dyke	Toston Cotillo
16–20	19	Series II	Playa del Aguila
21–27	27	Series I	La Matilla
28	4	Series II	4 km west of Puerto del Rosario
29–34	26	Series I	Pozo Negro
35–41	35	Series I	2 km west of Pozo Negro
42–43	8	Trachyte	Tuineje
44–45	10	Series II	Just west of Puerto del Rosario
46–47	10	Series II	Between Puerto del Rosario and the airport
48	5	Series II	Between Casillas del Angel and Puerto del Rosario

2. Palaeomagnetic Collection

The present palaeomagnetic study is based on a total of 234 samples from 56 sites, comprising trachyte (3 sites), dykes (12 sites), sediment (1 site) lava Series I (28 sites) and lava Series II (12 sites). Only one site has been sampled from each rock body (lava, dyke etc.). About half of the sites were collected by portable drill, the remaining material being collected by standard hand sampling. On the average four samples (drill cores or hand samples) were collected from each site. Sun compass readings are available for about 25% of the sites. The mean declination estimated from the difference between magnetic and sun compass readings is 11.9 W which is in fairly good agreement with the inferred present regional declination for Fuerteventura (ca. 10° W). Total variation in declination estimates is between 6.6 W and 22.3 W, but the majority of the results are close to the mean value. For further collection details refer to Table 1 and Fig. 1.

3. Palaeomagnetic Results

In the laboratory an average of six specimens from each site has been subjected to progressive demagnetization in alternating field (AF) or temperature, ranging upwards to 1,500 Oe and 580° C respectively. Palaeomagnetic properties of individual specimens after each demagnetization step were determined by using

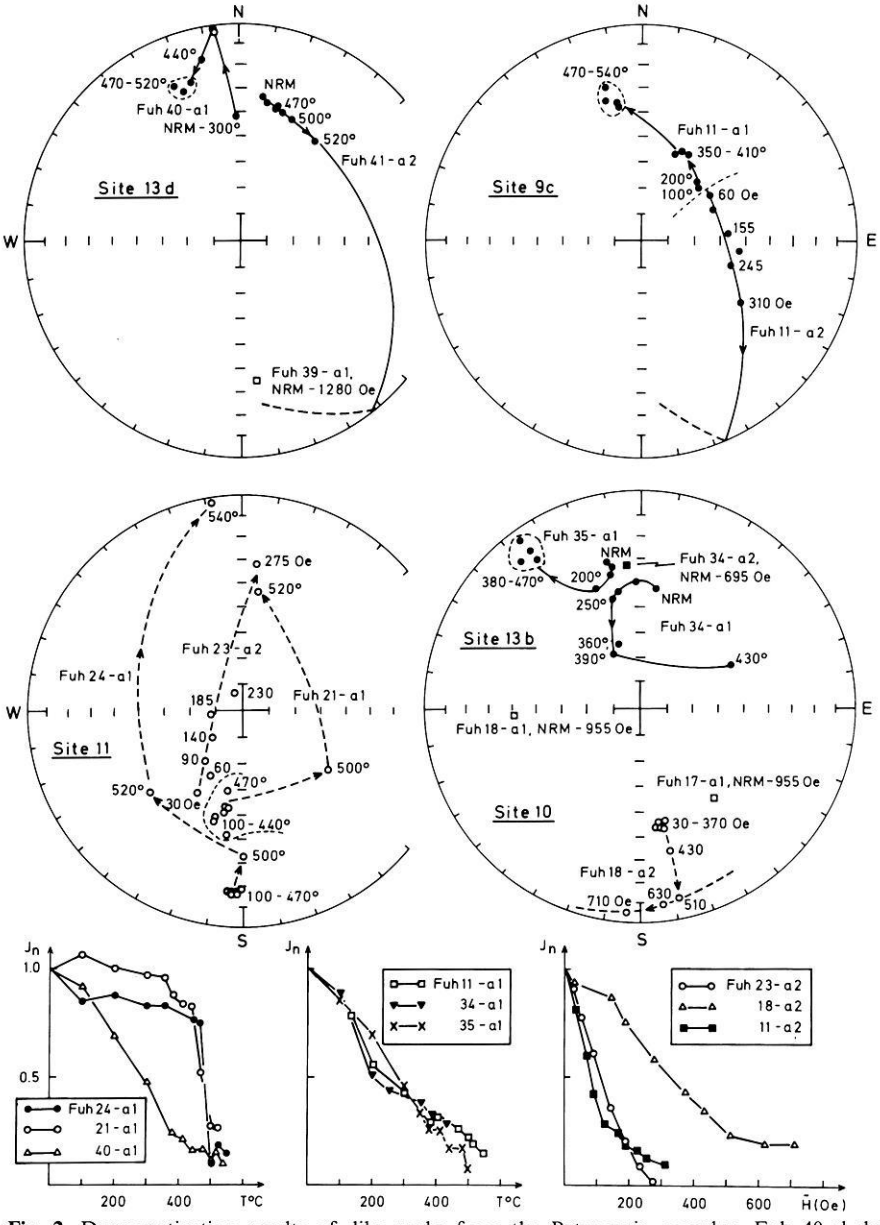


Fig. 2. Demagnetization results of dike rocks from the Betancuria complex. Fuh 40-a1 denotes specimen a1 cut from sample Fuh 40 etc. Projection is equal area and open (*closed*) symbols represent the direction of upward (downward) pointing magnetization vectors. Square symbols are mean magnetization directions of closely spaced results obtained within demagnetization ranges (field of temperature) as indicated on the diagrams. Note that neighbouring specimens (from same drill core or site) may show opposite polarity of their bulk remanences (cf. site 13 d) or may be heading towards opposite directions as demagnetization progresses (cf. sites 9c and 13d). Also, the initial (and sometimes fairly stable) remanence directions have often a smeared distribution that lines up with the directional movements that take place at a more advanced stage of demagnetization (cf. site 11). In specimen Fuh 18-a2 (site 10) the predominating remanence appears to be the C (or possibly A) magnetization (cf. Fig. 5) but at fields $\geqq 510$ Oe the directional pattern tends to define a practically horizontal great circle path probably associated with an interplay between the normal and reverse B components. Note also the discordant (but stable) direction of magnetization of the adjacent specimen Fuh 18-a1

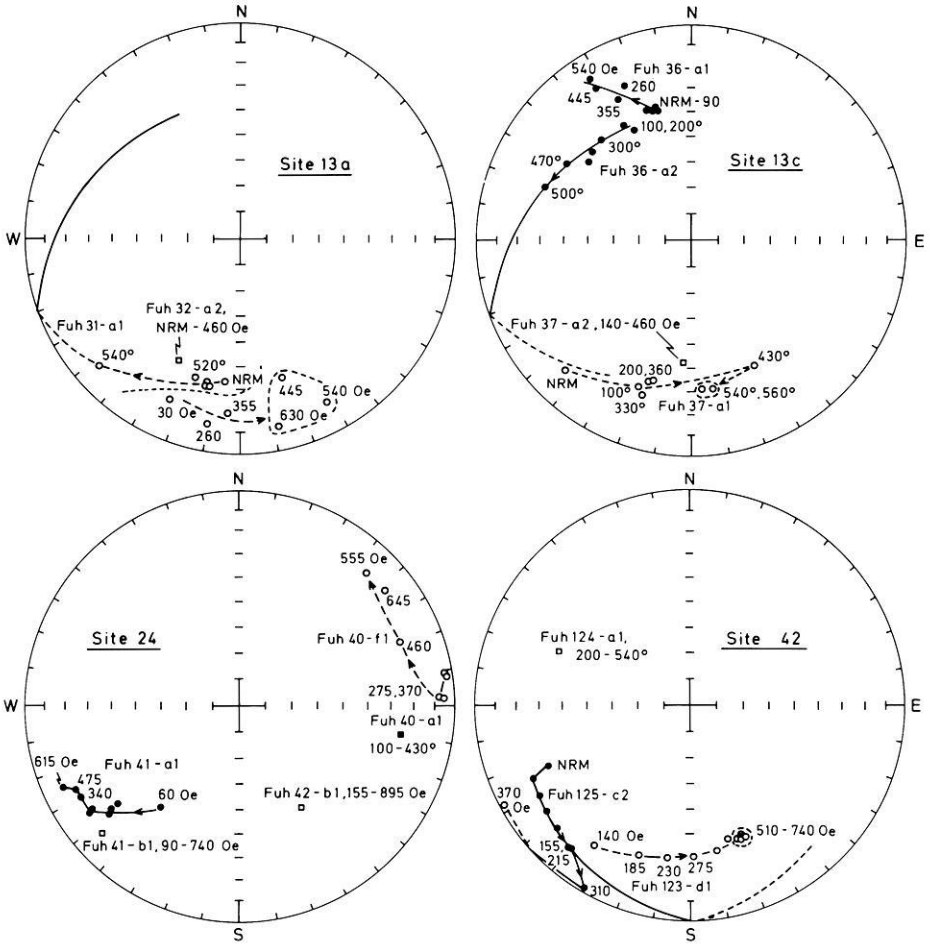


Fig. 3. Further examples illustrating magnetization complexities in the older Fuerteventura rocks: sites 13a and c from the sheeted dike complex, site 24 from lava Series I and site 42 from a trachyte intrusion. Diagram comventions as for Fig. 2

commercial spinner magnetometers. For specimens showing systematic directional trends (AF or thermal) the adopted acceptance criteria for 'stable end points' is that such terminal directions must be experimentally confirmed by at least three successive demagnetization steps (for each specimen). However, at some stages of demagnetization a great number of specimens came into an erratic stage (anomalously varying results both in direction and intensity) before stable end points could be defined. Specimens of the latter category and which, over a certain range of demagnetization, were shown to move along great circle paths (see below) have been employed in an analysis to estimate palaeomagnetic components through estimation of best intersection points of remagnetization circles (Halls, 1976; 1978).

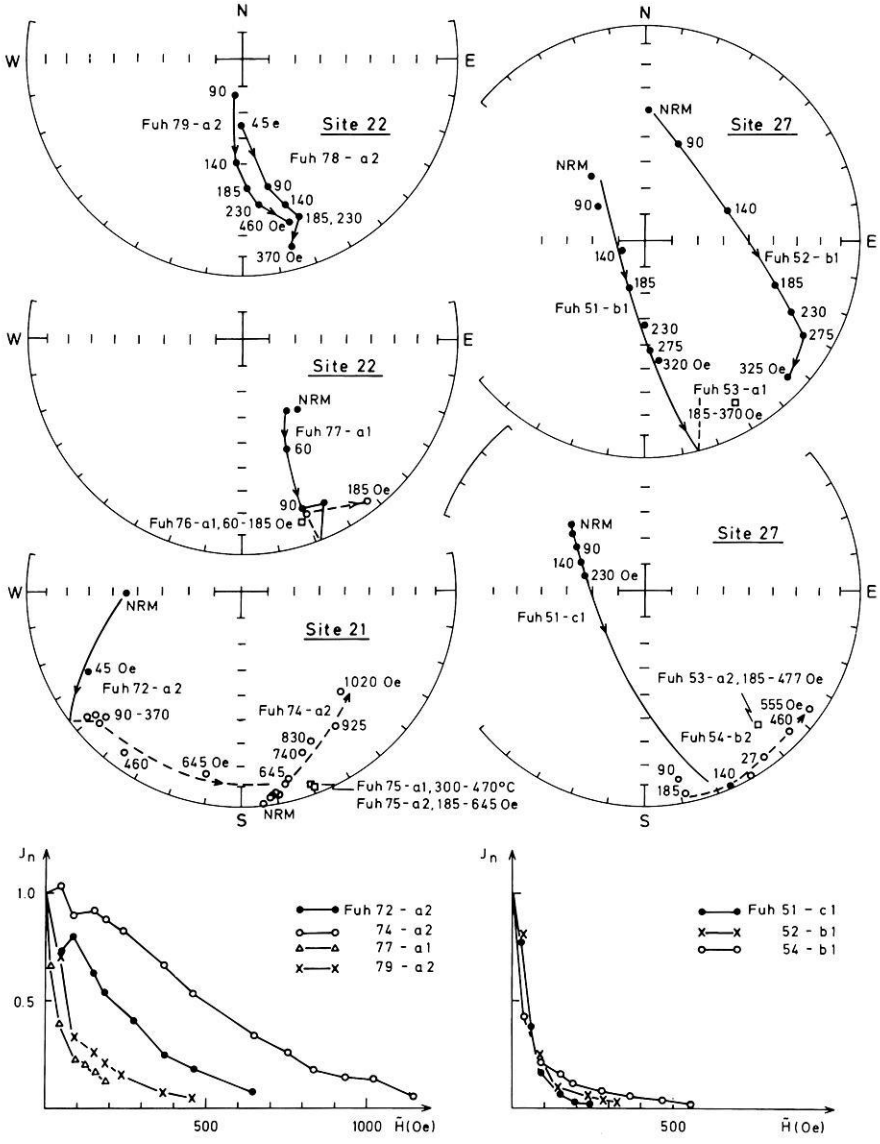


Fig. 4. Demagnetization results from lava Series I sites in which the B magnetization appears to be an important palaeomagnetic component. Diagram conventions as for Fig. 2

Lava series I (the oldest sequence of subaerial volcanics) and the Betancuria rocks have primarily a multivectorial NRM. Directional changes versus increasing demagnetization characterize these rocks but very often the remanence decays into the 'noise level' before a terminal direction is reached. An important observation is the occurrence of superimposed normal and reverse magnetizations, demonstrated by the directional behaviour versus demagnetization of individual specimens or by 'anti parallel' directions of stable bulk magnetization in neighbouring specimens (cf. Figs. 2–4). For many sites the total palaeomagnetic

Table 2. Palaeomagnetic results ('stable end point' data) from Fuerteventura. The D and I values represent averaged figures over the ranges quoted. See text for other details

Site		Lava series I (palaeomagnetic groups A and B)					
		Specimen	D	I	κ	α_{95}	Range (th. or AF)
1	Fuh	1–a3	161	−14			230–650 Oe
		3–b2	182	−29			190–370 Oe
		3–a3	189	−28			250–520° C
		5–b1	172	−28			30–220 Oe
2		6–b2	170	−29			NRM–920 Oe
		7–a2	160	−30			45–830 Oe
		8–b3	160	−40			830–920 Oe
		9–a	179	−27			NRM–920 Oe
Site mean			168	−32	69	11°	
3		12–a2	174	−25			NRM–620 Oe
		12–b2	170	−24			100–475° C
		13–b2	185	−33			NRM–830 Oe
		14–b3	160	−27			70–920 Oe
Site mean			172	−28	67	11°	
4		15–b3	159	− 2			140–920 Oe
		16–b2	169	− 9			90–830 Oe
		17–b2	150	−11			280–510 Oe
		17–a2	156	−17			450–520° C
		18–a3	166	− 3			140–9200e
		19–a	155	−17			320–1200 Oe
		19–b	156	−37			375–470° C
5		20–a1	155	+ 2			NRM–550 Oe
		20–b2	153	+ 2			NRM–550° C
		21–a2	161	+ 2			NRM–830 Oe
		21–a1	161	+ 1			NRM–525° C
		21–b1	168	−10			540–580° C
		22–b2	169	+ 4			NRM–650 Oe
		22–a2	167	+ 4			NRM–560° C
		22–b1	165	+ 6			NRM–510° C
		23–a2	150	+ 3			NRM–740 Oe
		23–b1	153	+ 6			NRM–525° C
		23–b2	153	+ 3			NRM–540° C
		24–b1	171	+ 4			NRM–510° C
		24–b2	169	+ 3			NRM–550 Oe
Site mean			161	+ 2	91	4°	
32		72–a2	318	+24			NRM–740 Oe
		72–b2	329	+24			NRM–400° C
		74–a	352	+29			NRM–465° C
		75–a	339	+40			NRM–540° C
		76–c	343	+42			NRM–740 Oe
Site mean			336	+32	32	14°	
33		77–b2	344	+27			NRM–920 Oe
		77–a3	339	+28			NRM–570° C
		78–a	353	+19			NRM–920 Oe
		79–a2	334	+22			NRM–920 Oe
		79–b2	334	+20			NRM–540° C
Site mean			341	+23	92	8°	

Table 2 (continued)

Site		Lava series I (palaeomagnetic groups A and B)					
		Specimen	D	I	κ	α_{95}	Range (th. or AF)
34	Fuh	82–a	354	+36			45–920 Oe
		84–b	317	+20			45–920 Oe
		85–b2	342	+22			500–540° C
		86–b	338	+19			45–415 Oe
Site mean			337	+25	26	18°	
36		92–a2	342	+20			45–450 Oe
		93–b	345	+33			NRM–510 Oe
		94–c2	354	+19			830–1480 Oe
		95–a	345	+31			NRM–560° C
Site mean			347	+26	87	10°	
37		97–a3	340	+15			480–550° C
		97–b2	336	+29			45–460 Oe
		99–a3	326	+29			90–1480 Oe
		99–b2	316	+34			400–555° C
		99–b3	311	+30			140–1525 Oe
		100–b2	338	+26			480–540° C
		100–b3	327	+25			45–920 Oe
		101–b3	335	+27			90–1480 Oe
Site mean			329	+27	55	7°	
38		102–a3	343	+29			NRM–460 Oe
		102–b2	340	+12			495–550° C
		104–b2	339	+28			NRM–1480 Oe
		105–a2	351	+18			100–400° C
		106–b2	343	+37			NRM–500° C
Site mean			343	+25	57	10°	
39		107–a3	344	+26			45–450 Oe
		107–b3	346	+28			200–540° C
		108–a3	340	+30			NRM–1480 Oe
		109–a	336	+21			400–540° C
		110–a2	325	+35			400–560° C
		110–b	352	+34			45–1480 Oe
		111–a2	341	+26			NRM–475° C
		111–b	341	+34			NRM–550° C
Site mean			341	+30	87	6°	
40		112–a3	344	+41			NRM–460 Oe
		112–b3	341	+33			NRM–450° C
		113–b2	337	+25			NRM–1480 Oe
		114–b	336	+26			90–920 Oe
		115–a3	340	+34			NRM–520° C
		116–b2	327	+15			180–920 Oe
Site mean			337	+29	61	9°	
41		117–a2	335	+34			NRM–1380 Oe
		117–b2	335	+32			NRM–500° C
		118–a3	335	+34			NRM–475° C
		118–b2	333	+36			NRM–740 Oe
		119–a2	339	+34			NRM–555° C
		121–a2	332	+20			NRM–320 Oe
Site mean			335	+32	171	5°	

Site		Lava series II (palaeomagnetic group C)					
		Specimen	D	I	κ	α_{95}	Range (th. or AF)
12		26–a2	183	−29			355–585 Oe
		27–a1	189	−31			NRM–460 Oe
		27–a2	184	−38			355–490 Oe
		28–a2	182	−22			30–370 Oe
		29–a2	196	−16			30–370 Oe
Site mean			187	−27	66	9,5°	
16		53–a2	184	−25			350–860 Oe
		54–a1	179	−24			NRM–390 Oe
		55–a2	182	−22			NRM–860 Oe
		56–a1	168	−22			NRM–470° C
Site mean			178	−23	145	7,5°	
17		57–a1	171	−48			NRM–560° C
		58–a1	187	−37			30–830 Oe
		58–a2	189	−40			30–755 Oe
		59–a1	187	−28			NRM–520° C
		59–a2	183	−30			NRM–755 Oe
		60–a1	187	−22			NRM–560° C
		61–a2	193	−26			NRM–560° C
Site mean			186	−33	59	8°	
18		62–a1	181	−28			NRM–830 Oe
		62–a2	184	−32			470–550° C
		63–a1	197	−25			NRM–390° C
		64–a1	190	−27			NRM–830 Oe
		64–a2	191	−28			NRM–755 Oe
		65–a1	187	−32			NRM–700 Oe
		65–a2	192	−35			90–660 Oe
Site mean			189	−30	193	4,5°	
19		67–a1	183	−29			NRM–830 Oe
		67–a2	189	−29			NRM–755 Oe
		68–a1	190	−29			NRM–540° C
		68–a2	189	−23			NRM–830 Oe
Site mean			188	−28	387	4,5°	
20	Fuh	70–a1	178	−25			90–740 Oe
		70–a2	175	−26			60–370 Oe
		71–a1	180	−30			NRM–740 Oe
		71–a2	176	−26			100–470° C
Site mean			177	−27	747	3°	
44		131–a2	203	−37			45–925 Oe
		132–b	191	−34			NRM–520° C
		133–a2	190	−18			NRM–500° C
		134–a	204	−21			NRM–475° C
		134–b2	203	−21			NRM–1480 Oe
Site mean			198	−26	57	10°	
45		136–a	188	−39			NRM–925 Oe
		138–a2	186	−28			45–740 Oe
		139–a2	171	−38			450–540° C
		139–b2	191	−34			45–740 Oe
Site mean			184	−35	86	10°	

Table 2 (continued)

Site	Lava series II (palaeomagnetic group C)					
	Specimen	D	I	κ	α_{95}	Range (th. or AF)
46	140–a3	165	−39			45–740 Oe
	141–a3	169	−47			45–740 Oe
	142–a	167	−33			NRM–540° C
	143–c2	184	−29			90–510 Oe
	144–b2	175	−42			NRM–500° C
Site mean		172	−38	73	9°	
47	145–b	184	−40			NRM–740 Oe
	146–a2	186	−39			NRM–740 Oe
	147–a2	192	−36			NRM–525° C
	148–b2	179	−33			NRM–520° C
	149–a2	180	−48			NRM–520° C
	149–b2	178	−53			NRM–1480 Oe
Site mean		183	−41	90	7°	

information consists of a number of directional paths (without stable end points) in addition to one or more stable (apparently one-component) directions. There may however, be large disagreements between such stable specimen directions, indicating that at least in part they may represent unresolved multicomponent remanences. This assumption is strenghtened by the unsuccessful attempt of estimating the direction(s) of erased single-component vectors for specimens following great circle paths on demagnetization: there appears to be no range of demagnetization for which vector subtracted directions cluster, indicating that the stability spectra are closely overlapping. Therefore, in order to eliminate as many as possible of unresolved multicomponent (but stable) directions of magnetization from the distribution of true palaeomagnetic observations a strict acceptance criteria has been adopted: only sites which have a minimum of four relatively well-grouped specimens (i.e., 60−100% of analysed specimens per site) have been considered for estimation of palaeomagnetic parameters. In none of these sites does the total directional spread of stable directions exceed 35° of arc. In addition to the accepted directions of magnetization these sites may exhibit a number of specimens that perform systematic directional variation without achieving stable end points, but they have no stable directions in divergent positions. Of the older Fuerteventura rocks only lava Series I sites have given acceptable palaeomagnetic results (i.e., stable end points corresponding to criteria outlined above). Lava series II have a simple remanence: the results from a total of 10 out of 12 sampled flows comply with the acceptance criteria. The bulk magnetization of the two remaining Series II lavas (site nos. 28 and 48) are also in accord with the Series II results but directional trends on demagnetization are relatively frequent and the number of stable specimens are fewer than here required.

The individual specimen results from all acceptable sites are listed in Table 2 and the total directional variation is shown in Fig. 5. Many of these specimens

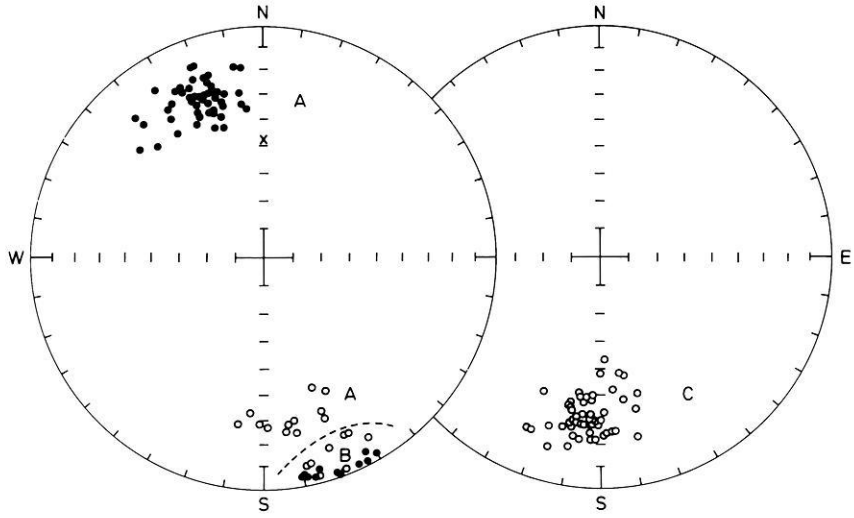


Fig. 5. Palaeomagnetic directions (specimens) from Fuerteventura. Groups *A* and *B* represent lava Series I and Group C Series II. The *cross* is the present axial dipole field direction relative to the Canary Islands. Further diagram conventions as for Fig. 2

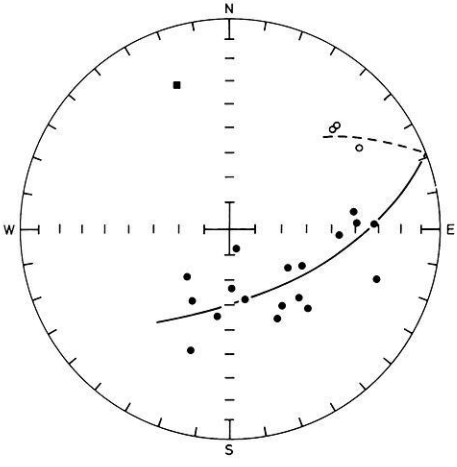


Fig. 6. Specimen remagnetization circle data (21 specimens). See text for details on data base. *Closed (open) circles* are downward (upward) pointing normals to specimen remagnetization circles. *The closed square* is the downward pointing normal to the great circle (drawn on diagram) fitted by least squares to the specimen normals

exhibit directional stability over the entire or major range of demagnetization, involving a minimum of 90% of the NRM moment. The Series II data (magnetization C) are exclusively reversed while Series I shows both polarities. From the reversed Series I data there is evidence that this lava sequence actually contains two palaeomagnetic axes: a major two polarity magnetization with an inclination of about 30°, referred to as magnetization A, and a subordinate reversed component that is practically horizontal (magnetization B). Following the acceptance criteria for stable end point results here adopted, only one site

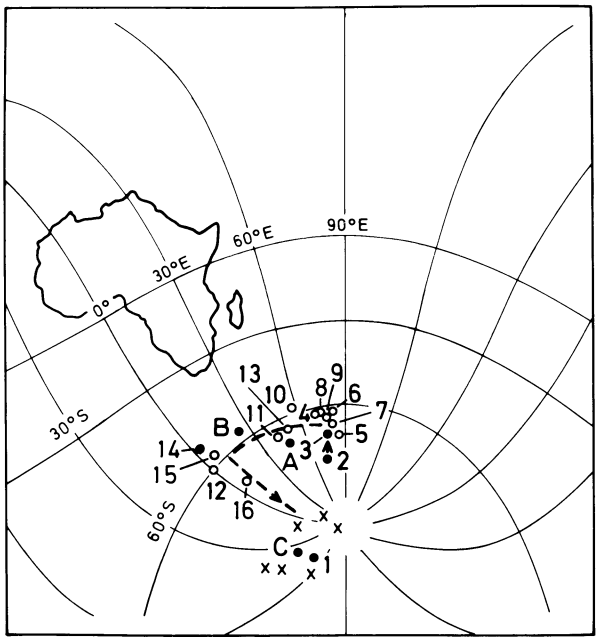


Fig. 7. Estimated pole positions for Fuerteventura basalts, poles A – C, in comparison with other relevant poles. Closed symbols are data from the Canary Islands. Poles A and B represent lava Series I and pole C lava Series II. Pole 1 is the Upper Tertiary/Quaternary pole for Gran Canaria/Tenerife and poles 2 and 3 represent the older volcanic rocks in these islands, before and after a 5° anticlockwise adjustment of the mean declination respectively (Storetvedt et al., 1978). Pole 14 is from lava Series I of N.Lanzarote (Johansen, 1976). *Crosses* are African Tertiary poles. Other poles are as follows: 4: Hoachanas (Gidskehaug et al., 1975); 5: Mlanje (Gough and Opdyke, 1963; Briden 1967); 6: Lupata, 106 m.y. (Gough and Opdyke, 1963; Gough et al., 1964); 7: mean Mesozoic SE. Africa (Hailwood and Mitchell, 1971); 8: mean Mesozoic NW. Africa (Hailwood and Mitchell, 1971); 9: Shava ijolite (Gough and Brock, 1964); 10: Kimberlite pipes, 83–89 m.y. (McFadden and Jones, 1977); 11: Wadi Natash volcanics, 81–90 m.y. (El Shazly and Krs, 1973); 12: mean Lower Tertiary of Europe (Storetvedt, 1978); 13: volcanics SE. Sicily, 70–80 m.y. (Schult, 1973); 15 and 16: L.-(M) Tertiary volcanics Egypt (Gouda Hussain et al., 1979)

(no. 5) shows an unambiguous B magnetization. In the other sites this component either interferes with the A magnetization, causing for example an inclination spread of the resultant remanence (see for example site 4, and probably also site 1, Table 2) or it may form part of a highly variable and complex magnetization. Examples of within-site magnetization and detailed demagnetization behaviour for sites that are not included in the final palaeomagnetic results are given in Figs. 2 – 4. A number of specimens have ranges of demagnetization in which the resultant vectors move along practically horizontal great circle paths, suggesting that two antiparallel *B* axis components are present. In Fig. 5 the distinction between magnetizations A and B is fairly tentative but considering also all the directional information from sites not included in Table 2 (see for example Fig. 4) there appears to be little doubt that Series I contains two axes of magnetization. Additional evidence in support of the B magnetization comes from the nearby Island of Lanzarote (cf. Fig. 7) and where the A component has not been found (Johansen, 1976; Skårnes, 1977).

Table 3. Mean palaeomagnetic data from Fuerteventura

Formation		N	R	K	α_{95}	$\bar{D}$	$\bar{I}$	Pole	
Group C; lava series II	a	51	50.1	53.5	2:8	184.8,	−31.4	324.5E,	77.9S
	b	10	9.9	89.1	5.2	184.3,	−31.0	326.2E,	77.8S
Group B; lava series I	a	20	19.7	59.6	4.3	160.6,	−02.1	023.6E,	57.0S
Group A; lava series I	a	62	60.7	46.6	2.7	340.5,	+28.4	043.9E,	67.6S
	b	12	11.9	98.7	4.4	342.2,	+28.2	040.8E,	68.8S
	c	21	–	–	–	340.5,	+28.2	043.9E,	67.6S

N: Number of unit vectors (a: specimens, b: sites, c: specimens in great circle analysis)
R: Length of resultant vector
K: Precision parameter
α_{95}: Radius of circle of confidence at 95% significance level
$\bar{D}, \bar{I}$: Declination and inclination of mean vector

In an attempt to retrieve palaeomagnetic information from specimens that do not achieve stable end points on demagnetization the great circle technique (Halls, 1976; 1978) was used for all specimen trends that had a quality factor $Q < 1 \cdot 10^{-5}$ (Halls, 1978). Fifty specimens satisfied this requirement. The average number of demagnetization steps for each great circle segment is 6. Figure 6 shows the distribution of poles for specimen great circles used for independent estimation of the direction of normal A magnetization. Of the 21 specimens concerned 12 are Series I rocks, 8 are from dykes and 1 from a trachyte intrusion. Eleven of the specimens have northward trending great circle paths for which the individual great circle segments, traced out by the successive total vectors, are mostly several tens of degrees away from the convergence area, while the remaining 10 specimens move into and basically within the area of stable normal A magnetization. As seen from Table 3 the best normal intersection point of this group of specimens is in perfect agreement with the mean A axis magnetization based on stable end points. The excellent agreement of the two methods and the fact that they are based on totally different specimen populations strengthen the palaeomagnetic reliability of the A magnetization.

While the northward trending remagnetization circles have a reasonably well defined intersection point this is not the case for those moving towards the reversed component(s). For this group (29 specimens) the distribution of poles for individual specimen trends are too complex to allow estimation of palaeomagnetic directions. Since the majority of these specimens have resultant magnetizations that are reversed, it seems likely that an interplay between the three reversed components (A, B, C) causes the complexity in the distribution of poles to specimen remagnetization circles.

Account of investigated specimens:

(a) 332 specimens have been studied, (b) 134 have given palaeomagnetic results according to acceptance criteria (Table 2), (c) of the remaining 198 specimens ca. 50% gave stable end points, in frequent agreement with inferred magnetization components (A, B, or C), but the associated site magnetizations still remained multicomponent (and scattered) after demagnetization. Finally,

ca. 100 specimens did not define stable end points – 50 of these formed great circle segments on progressive demagnetization and were used in the remagnetization circle analysis, and the rest had a more complex behaviour.

4. Palaeomagnetic Interpretations

Summary of palaeomagnetic results including mean directions of magnetization, statistical parameters and pole locations are given in Table 3.

As seen from Fig. 7 the Fuerteventura C pole (lava Series II) plot in close agreement with a late Tertiary-Quaternary pole recently obtained from the Canary Islands (Storetvedt et al., 1978) and corresponds well with data of similar age for continental Africa. Like the pole for the older Gran Canaria/Tenerife volcanic strata (Storetvedt et al., 1978) the Fuerteventura A pole, corresponding to the principal palaeomagnetic direction in baselt Series I, agrees fairly well with those for the Mesozoic of Africa. Pole B has a rather limited data base but the scatter of individual specimen directions is similar to those for groups A and C. In view of the ample evidence for partial remagnetization in basalt Series I (as well as in the older rocks) it is likely that the magnetization at least in part is of CRM or TCRM origin (note that from Table 2 and Figs. 2–4 some sites are likely to contain both the A and the B magnetization). Considering all the available evidence it appears reasonable to conclude that the B pole reflects an average palaeomagnetic field axis rather than being an 'artifact' of incomplete averaging-out of geomagnetic secular variation.

Figure 7 shows the estimated Fuerteventura pole positions in comparison with other relevant data. The eastern pole cluster (poles 5–9) defines the representative pole lacation relative to Africa for the major part of Mesozoic time, the youngest representative being that for the Lupata volcanics that date back to ca. 106 m.y. Poles 10, 11, and 13, situated west of the previous group, represent volcanic rocks from Africa and SE. Sicily in the approximate age range 70–90 m.y. A further westward extension of the polar pattern encounters the Lower Tertiary pole position for Europe at around 0° E, 60° S (pole 12) which is in very good agreement with recent Lower-(Middle) Tertiary palaeomagnetic results from Egypt, poles 15 and 16 (Gouda Hussain et al., 1979). It has been concluded that the European Tertiary polar shift towards the present pole location occurred at around Middle Oligocene (Storetvedt, 1973) and from the available data there is good evidence to infer a similar axis shift relative to Africa at the same time. Therefore, it appears that Europe and Africa define 'identical' polar patterns for the last 90 m.y. or so. It is interesting to note that based on palaeoclimatological evidence Köppen and Wegner (1924) suggest a Mesozoic-Lower Tertiary-Upper Tertiary polar track relative to Africa that is closely similar to the one outlined here (cf. their Fig. 22). The Fuerteventura A and B poles fit well into this framework. With reference to Fig. 7 and the geological evidence summarized above (Chap. 1) it is appropriate to conclude that the basal complex originated in the Upper Cretaceous (post-Albian) while basalt Series I was laid down at around the Cretaceous-Tertiary boundary.

5. Conclusion and Discussion

From geological evidence (Robertson and Stillman, 1979) passive margin conditions in the area of Fuerteventura ended in Albian time and was followed by uplift and tectonism. According to the palaeomagnetic results outlined above it appears reasonable to assume that emplacement of the basal igneous core of Fuerteventura (the Betancuria complex) accompanied these tectonic processes, forming sometime in the Upper Cretaceous. The plutonic activity was succeeded by an erosional unconformity after which the Series I plateau basalts were laid down probably at around the Cretaceous/Tertiary boundary.

After Series I had been formed there followed a period of ca. 50 m.y. of magmatic quiescence and erosion. In Middle Miocene-Pliocene the eroded surface was overstepped by a second major sequence of plateau basalt (Series II). We believe that the long period of erosion and weathering (in a tropical – subtropical environment) along with the penetrative thermochemical effects accompanying the younger lava series led to partial remagnetization, manifested by the present multicomponent magnetization, of the older formations.

These conclusions are at variance with the radiometric age determinations (conventional whole rock K/Ar results) available from the older igneous rocks of Fuerteventura (Rona and Nalwalk, 1970; Abdel-Monem et al., 1971; Grunau et al., 1975; Stillman et al., 1975). However, the radiometric data pose interpretional problems. For example, Grunau et al. (1975) have noted that, following a normal differentiation trend for the plutonics, the oldest age of these rocks (38 m.y.) in fact comes from one of the supposed youngest intrusive members (an alkali syenite). The same authors also report on extremely poor reproducibility of the Argon analyses from different splits of two plutonic samples (diorite and syenite). For the sheeted dyke complex the total apparent age ranges from 12.1 m.y. to 46 m.y. At least the oldest of these ages is in contradiction to the 35 m.y. age estimate that has been obtained from the submarine volcanics (at the base of the igneous series). Some of these problems are certainly related to the general metamorphism of the basement complex. However, we believe in particular that the extensive magmatism that swept the island in Miocene/Pliocene time lead to a partial degassing of the older igneous rocks (including basalt Series I), giving rise to a complex and geologically too young K/Ar age distribution.

Acknowledgements. We wish to thank T. Storetvedt, O. Johansen, and A. Skårnes for their assistance during the field work. We are greatly indebted to Drs. C.J. Stillman and M. Le Bas for providing unpublished geological information. This work was funded by the Norwegian Research Council for Science and the Humanities.

References

Abdel-Monem, A., Watkins, N.D., Gast, P.W.: Potassium – argon ages, volcanic stratigraphy and geomagnetic polarity history of the Canary Islands; Lanzarote, Fuerteventura, Gran Canaria, and La Gomera. Am. J. Sci. **271**, 490–521, 1971

Briden, J.C.: A new palaeomagnetic result from the Lower Cretaceous of East Central Africa. Geophys. J. R. Astron. Soc. **12**, 375–380, 1967

El Shazly, E.M., Krs, M.: Palaeogeography and palaeomagnetism of the Nubian Sandstone, Eastern Desert of Egypt. Geol. Rundschau **62**, 212–225, 1973

Fuster, J.A., Cendrero, A., Gastesi, E., Lopez Ruiz, J.: Geologia y volcanologia de las Islas Canarias – Fuerteventura. 239 pp. Madrid: Consejo Sup. Investig. Cient. 1968

Gastesi, P.: Is the Betancuria Massif, Fuerteventura, Canary Islands, an uplifted piece of oceanic crust? Nature Phys. Sci. **246**, 102–104, 1973

Gidskehaug, A., Creer, K.M., Mitchell, J.G.: Palaeomagnetism and K-Ar ages of the South-West African basalts and their bearing on the time of initial rifting of the South Atlantic Ocean. Geophys. J.R. Astron. Soc. **42**, 9–20, 1975

Gouda Hussain, A., Schult, A., Soffel, H.: Palaeomagnetism of the basalts of Wadi Abu Tereifiya, Mandisha and diorite dykes of Wadi Abu Shihat, Egypt. Geophys. J.R. Astron. Soc. **56**, 55–61, 1979

Gough, D.I., Brock, A.: The palaeomagnetism of the Shawa ijolite. J. Geophys. Res. **69**, 2489–2493, 1964

Gough, D.I., Brock, A., Jones, D.L., Opdyke, N.D.: The palaeomagnetism of the ring complexes at Marangudzi and the Mateke Hills. J. Geophys. Res. **69**, 2499–2507, 1964

Gough, D.I., Opdyke, N.D.: The palaeomagnetism of the Lupata Alkaline volcanics. Geophys. J.R. Astron. Soc. **7**, 457–468, 1963

Grunau, H.R., Lehner, P., Cleintuar, M.R., Allenbach, P., Bakker, G.: New radiometric ages and seismic data from Fuerteventura (Canary Islands) and Sao Tomé (Gulf of Guinea). In: Progress in Geodynamics. Amsterdam: North-Holland Publ. Comp. 1975

Hailwood, E.A., Mitchell, J.G.: Palaeomagnetic and radiometric dating results from Jurassic intrusions in South Morocco. Geophys. J.R. Astron. Soc. **24**, 351–364, 1971

Halls, H.C.: A least-squares method to find a remanence direction from converging remagnetization circles. Geophys. J.R. Astron. Soc. **45**, 297–304, 1976

Halls, H.C.: The use of converging remagnetization circles in palaeomagnetism. Phys. Earth Planet. Inter. **16**, 1–11, 1978

Johansen, O.: Thesis. Bergen: University of Bergen 1976

Köppen, W., Wegener, A.: Die Klimate der geologischen Vorzeit. Borntraeger, 256 pp. Berlin 1924

McFadden, P.L., Jones, D.L.: The palaeomagnetism of some Upper Cretaceous Kimberlite occurrences. Earth Planet. Sci. Lett. **34**, 125–135, 1977

Robertson, A.H.E., Stillman, C.J.: Late Mesozoic sedimentary rocks of Fuerteventura, Canary Islands: implications for West African continental margin evolution. J. Geol. Soc. London **136**, 47–60, 1979

Rona, P., Nalwalk, A.J.: Post-early Pliocene unconformity of Fuerteventura, Canary Islands. Geol. Soc. Am. Bull. **81**, 2117–2122, 1970

Rothe, P.: Zum Alter der Vulkanismus auf den östlichen Kanaren. Soc. Sci. Finnica. Commentar. Phys. Math. Soc. Sci. Fenn. **31**, (13), 1–80, 1966

Rothe, P.: Mesozoische Flysch – Ablagerungen auf der Kanarieninsel Fuerteventura. Geol. Rundschau **58**, 314–322, 1968

Schult, A.: Palaeomagnetism of Upper Cretaceous volcanic rocks in Sicily. Earth Plant. Sci. Lett. **19**, 97–100, 1973

Skårnes, A.: Thesis. Bergen: University of Bergen 1977

Stillman, C.J., Fuster, J.M., Bennell-Baker, M.J., Munoz, M., Smewing, J.D., Sagredo, J.: Basal complex of Fuerteventura (Canary Islands) is an oceanic intrusive complex with rift-system affinities. Nature **257**, 469–471, 1975

Storetvedt, K.M.: The rotation of Iberia; Caenozoic palaeomagnetism from Portugal. Tectonophysics **17**, 23–39, 1973

Storetvedt, K.M.: Structure of remanent magnetization in some Skye lavas, N.W. Scotland. Phys. Earth Planet. Inter. **16**, 45–58, 1978

Storetvedt, K.M., Svalestad, S., Thomassen, K., Langlie, Aa., Nergård, A., Gidskehaug, A.: Magnetic discordance in Gran Canaria/Tenerife and its possible relevance to the formation of the NW. African continental margin. J. Geophys. **44**, 317–332, 1978

Watkins, N.D.: Palaeomagnetism of the Canary Islands and Madeira. Geophys. J.R. Astron. Soc. **32**, 249–267, 1973

Watkins, N.D., Richardson, A., Mason, R.G.: Palaeomagnetism of the Macaronesian insular region: The Canary Islands. Earth Plant. Sci. Lett. **1**, 225–231, 1966

Received February 5, 1979; Revised Version July 16, 1979; Accepted September 22, 1979

J. Geophys. 46, 335–342, 1979

Journal of Geophysics

Ray Theoretical Seismograms for Laterally Inhomogeneous Structures*

V. Červený

Institute of Geophysics, Charles University, Ke Karlovu 3, 121 16 Praha 2, Czechoslovakia

Abstract. Applications of the ray method to the construction of theoretical seismograms for laterally varying layered structures are discussed. Numerical examples are presented. It is shown that the refracted waves are very sensitive to the curvature of interfaces. Certain modifications and improvements of ray theoretical seismograms to increase their accuracy are suggested.

Key words: Ray theoretical seismograms – Laterally inhomogeneous structures – Modifications of ray method.

1. Introduction

Various methods can be used to compute theoretical seismograms for vertically (or radially) inhomogeneous media. It is not possible to give here a full survey of these methods, it can be found in Červený et al. (1978). We shall mention here only some of them. The most popular are probably the reflectivity method (Fuchs, 1968; Fuchs and Müller, 1971), the exact ray method (also called the generalized ray method) and its various modifications (Cagniard, 1962; de Hoop, 1960; Müller, 1968; 1969; Helmberger, 1968; Chapman, 1976; Gilbert and Helmberger, 1972; Wiggins, 1974; etc.), and methods based on modal summation (Buland, 1977). Recently some new important methods have been developed, such as the method of partial separation of variables in the combination with finite differences (Alekseev and Mikhailenko, 1976; 1977; 1978).

Most of these methods, however, can be hardly used to compute theoretical seismograms for general laterally inhomogeneous media with curved interfaces. It is the regions with strong lateral inhomogeneities, however, that play a very important role in the present geodynamic studies. Ray theoretical seismograms can give valuable results even for these types of media, especially when we

* Presented at the Workshop Meeting on Seismic Waves in Laterally Inhomogeneous Media, Liblice, ČSSR, February 27–March 3, 1978

0340-062X/79/0046/0335/$01.60

use some modifications to increase their accuracy in singular regions and in some other situations.

Let us note that under 'ray theoretical seismograms' we understand here the theoretical seismograms computed by the standard ray method or by some of its modifications.

The application of ray methods to the construction of theoretical seismograms is most natural in the case of small epicentral distances (reflection methods in seismic prospecting, etc.). In this case, the construction of theoretical seismograms does not cause great difficulties and the accuracy of computations is usually quite satisfactory, even in the case of rather complicated structures. At larger epicentral distances, where we deal mainly with refracted waves and supercritical reflections (e.g., 50–300 km for the models of the Earth's crust), the situation is more complicated in several respects. However, it is shown in Červený (1979) that even at these epicentral distances the ray theoretical seismograms can give satisfactory results from an interpretational point of view. In this paper, we shall shortly describe the procedures to construct theoretical seismograms for such epicentral distances. We shall also present some numerical examples. For very large epicentral distances (say, 500 km or more for the models of the Earth's crust), the rays of waves propagating within the Earth's crust are mostly horizontal, and the waves are formed by a superposition of a large number of elementary waves. The accuracy of ray theoretical seismograms in this region is expected to be lower. These large epicentral distances are not considered here, they would need special investigation.

Many other details on various aspects of the construction of ray theoretical seismograms for laterally inhomogeneous media can be found in Červený et al. (1977), Červený and Pšenčík (1978), Smith (1977), Jedrzejowska-Zwinczak (1978).

2. Construction of Ray Theoretical Seismograms for Laterally Varying Layered Structures

The basic principles of the construction of ray theoretical seismograms are described in Červený et al. (1977), and in Červený (1979), for vertically inhomogeneous media. In principle, the same procedures can be used in the case of laterally inhomogeneous media. In some details, however, these procedures differ.

The first important difference consists in the computation of travel-times and complex amplitudes. At the present time, however, the methods of computation of travel times and complex amplitudes are well known, even for complicated laterally inhomogeneous media with curved interfaces, see Červený et al. (1977). The first and the most important step in these computations is the computation of rays. Rays in such media are described by a system of ordinary differential equation of the first order. The system can be solved by standard numerical procedures, such as the Runge-Kutta method. Some complications in these computations are connected with the fact that the computation of the ray trajectory is not a Cauchy initial value problem, but a two-point bound-

ary value problem (since we are looking for the ray connecting the source with the receiver). The solution of a two-point boundary value problem is usually easier at smaller epicentral distances; with increasing epicentral distance it becomes more complicated and cumbersome. Moreover, the travel-time curves of certain important waves may have many branches at larger epicentral distances, and we must determine all of them. Methods of a two-point ray tracing are discussed elsewhere, see Julian and Gubbins (1977).

As soon as the ray connecting the source and the receiver is determined, the evaluation of amplitudes is easier. Some problems are connected with the computation of geometrical spreading. Several methods for computing geometrical spreading are described in Červený et al. (1977).

The second difference consists in the impossibility of grouping the elementary waves in laterally inhomogeneous media into families of kinematically analogous waves, as in the case of vertically inhomogeneous media. It is necessary to compute elementary waves independently, one after another. This makes the computations more cumbersome and lengthy. The number of elementary waves arriving at the receiver within a time window of a given length increases considerably with the increasing epicentral distance.

The third difference consists in the algorithms for the generation of numerical codes of individual elementary waves. In comparison with the same algorithms for vertically inhomogeneous media, the generation algorithms are more complicated. The problem of the generation of numerical codes of elementary waves is closely connected with the problem of parameterization of the medium, and with the system of numerical coding of elementary waves. It is not simple to find an automatic or semi-automatic generation system for quite general laterally inhomogeneous media with block structures. It is often more suitable to read the numerical codes of elementary waves used to construct the theoretical seismogram for a specific model of medium in cards, as input data. The preparation of these input data is rather cumbersome but the experience of the interpreter may be also very important. However, when we use proper parameterization of the medium and proper numerical codes of waves, it is not difficult to develop fully or partially automatic generation algorithms. For example, it is very convenient to use such a parameterization of a medium in which all the interfaces are continuous from the left border to the right border of the medium under investigation. Then the interfaces and the corresponding individual layers can be numbered from the top to the bottom. Some interfaces might in part coincide with other interfaces or they might be in part only fictitious (with the same velocities and densities on both sides). Then we can use quite similar numerical coding of elementary waves as in vertically inhomogeneous media, and very similar generation algorithms. Some elementary waves arriving at the receiver will be, of course, of zero amplitudes. The generation algorithms, however, become much more simple then in the general case.

Similar parametrization of the medium, coding of elementary waves and generation algorithms are used in our programs. They permit fully automatic generation of all primary P, S, PS, SP, PP, and SS reflected and refracted waves for an arbitrary surface or burried source and a receiver. The generation of other multiply reflected/refracted waves we wish to consider can be read in as input data.

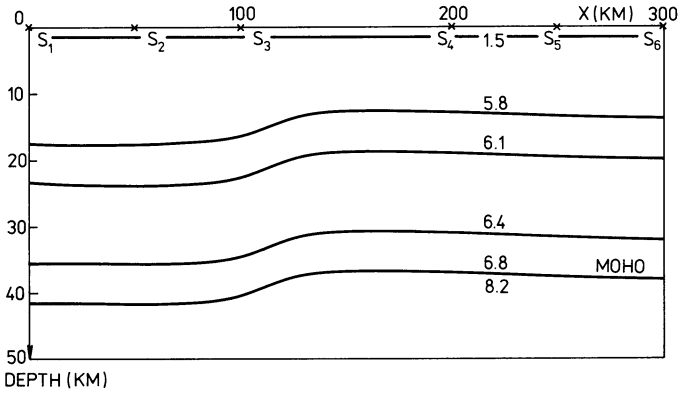


Fig. 1. Model of a laterally varying layered Earth's crust, used for a computation of theoretical seismograms. Thick lines denote the geometrical shape of interfaces, the numbers the velocity in individual layers. The points *S*1–*S*6 show the receiver positions

3. Examples

We shall now give examples of ray theoretical seismograms for laterally varying layered structures. Assume that the medium is composed of homogeneous layers separated by curved interfaces. (Let us note that a smooth laterally inhomogeneous medium can be simulated by a system of thin homogeneous layers with curved interfaces.) The interfaces are specified by a set of points and approximated by splines. A program allowing for lateral variations of velocity within individual layers is under preparation. We consider an explosive point source of *P* waves, with the symmetrical directional characteristics, situated near to the Earth's surface. The source-time function is given by the formula

$$f(t)=\exp(-4\pi^2 t^2 f_M^2/\gamma^2)\cos(2\pi f_M t+v),$$

with $f_M=4$Hz, $\gamma=4$, $v=0$. Only ideal registration is considered, and a possible distortion by the recording equipment is not taken into account. The receiver is also situated near to the Earth's surface, and the vertical displacement component is presented. For simplicity, only primary reflected and refracted *P* waves are considered, no multiply reflected/refracted *P* waves, converted waves, *S* waves, etc. (Let us note that these waves can be optionally taken into account in the used programs.) No scaling of amplitudes with respect to the epicentral distance is applied.

The basic characteristics of the model of the Earth's crust, used for computations, see Fig. 1, is the slope of interfaces at distances $x=100$–130 km. At other distances, the interfaces are roughly horizontal. In Fig. 2, three systems of theoretical seismograms corresponding to three various positions of source are shown (see *S*1, *S*2, and *S*3 in Fig. 1). It can be clearly seen from Fig. 2 that the theoretical seismograms for the model of medium shown in Fig. 1 depend considerably on the position of the source, especially at large epicentral distances. Mainly the refracted waves are very sensitive to the slope of interfaces. For example, in the case of the source *S*2 ($x=50$ km), the refracted waves are

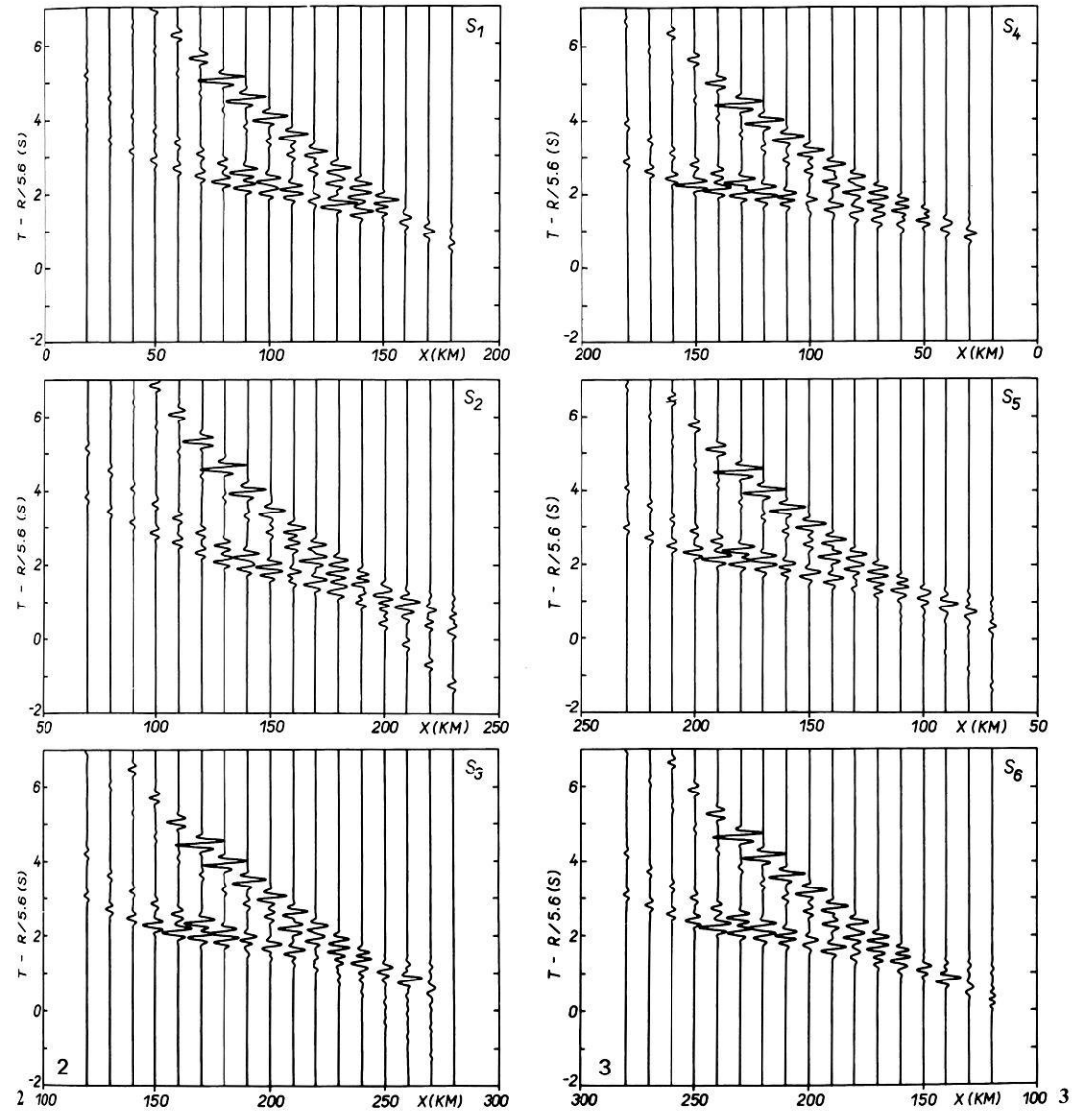


Fig. 2. Ray theoretical seismograms for the model of laterally varying Earth's crust shown in Fig. 1, for sources *S*1 (*top figure*), *S*2 (*middle figure*), *S*3 (*lower figure*). The epicentral distance is denoted by R, the distance along the profile by x. Reduction velocity equals 5.6 km/s

Fig. 3. Ray theoretical seismograms for the model of laterally varying Earth's crust shown in Fig. 1, for sources *S*4 (*upper figure*), *S*5 (*middle figure*) and *S*6 (*lower figure*). The epicentral distance is denoted by R, the distance along the profile by x. Reduction velocity equals 5.6 km/s

very strong. In fact, at the epicentral distance $R \sim 200$ km ($x \sim 250$ km), they are of the same order of magnitude or even higher than the primary reflections. (It should be noted that at these large epicentral distances the refracted wave propagating below the Moho comes at the receiver at the first arrivals.) In the case of the source $S3$ ($x = 100$ km), the refracted waves are rather weaker, as the curvatures of the interfaces are smaller in the relevant regions. In the case of the source at $S1$ ($x = 0$ km), the refracted waves vanish altogether; the reason can be simply understood from the geometry of interfaces.

The slope of interfaces in the region $x \sim 100$–130 km has also some influence on the amplitudes of the waves reflected from the corresponding interfaces at certain epicentral distances. This increase of amplitudes is connected in some cases with the loop on the travel-time curve, in other cases simply with the strong curvature of the travel-time curve.

Figure 3 shows three systems of theoretical seismograms computed along reversed profiles, corresponding to the sources situated at $S4$ ($x = 200$ km), $S5$ ($x = 250$ km) and $S6$ ($x = 300$ km). As in the preceding case, the refracted waves are not observed in the case of sources $S4$ and $S6$. They are, however, clearly visible in the case of the source $S5$. They again indicate the curvature of the interface. In this case, however, the refracted waves are weaker than in the case of the source at $S2$.

4. Possible Modifications of Ray Theoretical Seismograms

The main disadvantage of ray theoretical seismograms consists in their limited accuracy in some singular regions. Many of these limitations can be removed by means of various modifications of the ray method. An application of these modifications to the program for constructing ray theoretical seismograms needs primarily further development of the ray theory and of its various modifications, as well as progress in the numerical realization of these new theoretical approaches. The same can be said about a possibility of taking into account some of the non-ray and inhomogeneous waves, diffracted waves, etc. Much progress has been reached recently in the approximate computation of diffracted waves connected with wedges at interfaces (block structures, etc.), see Klem-Musatov et al. (1975). It would not be difficult to supplement the above described programs by the routines for an approximate computation of diffracted waves. Moreover, recent development in the theory of diffraction at curved interfaces suggests certain advantages of simulating smooth interfaces by a piece-wise linear approximation. This would remove some problems connected with caustics, triplications of travel-time curves, etc., of reflected waves from curved interfaces (Klem-Musatov, personal communication, 1977).

Some difficulties in constructing ray theoretical seismograms which are connected with a great number of elementary and with some non-ray waves (such as tunnel waves) can be simply removed by a combination of ray and matrix methods. This approach is effective mainly in studying the wave field in a medium composed of thick layers separated by thin transition layers. The mentioned modification was used successfully in seismic prospecting (Ratnikova,

1973). Up to this time, it has been used for vertically inhomogeneous media only. Nevertheless, it can be applied to laterally inhomogeneous media too.

Another promising approach is the computation of the wave field along the rays by more precise methods, see Kirpichnikova (1971). To perform these computations, it might be very convenient to use finite difference procedures. Generally, the combination of the ray method with the method of finite differences can be useful in various situations.

The ray method can be also generalized for some essentially different types of media, e.g., for anisotropic media, prestressed media, absorbing media, etc. Numerical computations of ray theoretical seismograms for such media should not cause too great difficulties, even in the case of horizontal inhomogeneities.

Acknowledgements. The author appreciates discussions with many colleagues during the course of this investigations. They include Drs. I. Pšenčík, J. Zahradník, G. Müller, J. Fertig, A.S. Alekseev, K.D. Klem-Musatov, A.B. Belonosova, and others.

References

Alekseev, A.S., Mikhailenko, B.G.: Solution of Lamb's problem for a vertically inhomogeneous elastic halfspace. Izv. Akad. Nauk SSSR, Fiz. Zemli, No. **12**, 11–25, 1976

Alekseev, A.S., Mikhailenko, B.G.: Numerical modelling of seismic wave propagation processes in a radially inhomogeneous model of the Earth. Dokl. Akad. Nauk SSSR **235**, 46–49, 1977

Alekseev, A.S., Mikhailenko, B.G.: A solution of dynamic problems of elastic wave propagation in inhomogeneous media by a combination of partial separation of variables and finite-difference methods. Presented at the Workshop on 'Seismic Waves in Laterally Inhomogeneous Media', Liblice, February 27–March 3, 1978

Buland, R.: Synthetic seismograms by modal summation (Abstract). Eos. **58**, 1190, 1977

Cagniard, L.: Reflection and refraction of progressive seismic waves. New York: McGraw Hill 1962

Červený, V.: Accuracy of ray theoretical seismograms. J. Geophys. **46**, 135–149, 1979

Červený, V., Fuchs, K., Müller, G., Zahradník, J.: Theoretical seismograms for inhomogeneous elastic media. Submitted to: Vopr. dinamicheskoy teorii rasprostraneniya seismicheskikch voln, Vol. 20, G.I. Petrashen, ed. Leningrad: Nauka 1978

Červený, V., Molotkov, I.A., Pšenčík, I.: Ray method in seismology. Praha: Univerzita Karlova 1977

Červený, V., Pšenčík, I.: Ray theoretical seismograms for laterally varying layered structures. In: Proceedings XV. General Assembly European Seismological Commission, Kraków 1976, Part I. Publ. Inst. Geophys. Pol. Acad. Sci. A-4 (115), R. Teisseyre, ed., pp. 173–185. Warszawa-Lódź: PWN 1978

Chapman, C.H.: Exact and approximate generalized ray theory in vertically inhomogeneous media. Geophys. J.R. Astron Soc. **46**, 201–233, 1976

Fuchs, K.: The reflection of spherical waves from transition zones with arbitrary depth-dependent elastic moduli and density. J. Phys. Earth **16** (Special Issue), 27–41, 1968

Fuchs, K., Müller, G.: Computation of synthetic seismograms with the reflectivity method and comparison with observations. Geophys. J.R. Astron. Soc. **23**, 417–433, 1971

Gilbert, F., Helmberger, D.V.: Generalized ray theory for a layered sphere. Geophys. J.R. Astron. Soc. **27**, 58–80, 1972

Helmberger, D.V.: The crust-mantle transition in the Bering Sea. Bull. Seism. Soc. Am. **58**, 179–214, 1968

Hoop, A.T. de: A modification of Cagniard's method for solving seismic pulse problems. Appl. Sci. Res. **B8**, 349–356, 1960

Jedrzejowska-Zwinczak, H.: On the application of the simulation method to exploration seismic data interpretation. Presented at the Workshop on 'Seismic Waves in Laterally Inhomogeneous Media', Liblice, February 27–March 3, 1978

Julian, B.R., Gubbins, D.: Three-dimensional seismic ray tracing. J. Geophys. **43**, 95–113, 1977
Kirpichnikova, N.J.: Construction of solutions concentrated close to rays for the equations of elasticity theory in an inhomogeneous isotropic space (in Russian, English translation by Am. Math. Soc., 1974). In: Matematicheskiye voprosy teorii diffrakcii i rasprostraneniya voln, Vol. 1, V.M. Babich, ed., pp. 103–113. Leningrad: Nauka 1971
Klem-Musatov, K.D., Kovalevskiy, G.L., Chernyakov, B.G., Maksimov, L.G.: Mathematical modelling of diffraction of seismic waves in angle-shaped regions. Geol. Geofiz. **11**, 116–124, 1975
Müller, G.: Theoretical seismograms for some types of point sources in layered media. Part 1: Theory. Z. Geophys. **34**, 15–35, 1968. Part 2: Numerical calculations. Z. Geophys. **34**, 147–162, 1968. Part 3: Single force and dipole sources of arbitrary orientation. Z. Geophys. **35**, 347–371, 1969
Ratnikova, L.: Methods of computation of seismic waves in thin-layered media. Moscow: Nauka 1973
Smith, S.G.: A reflection profile modelling system. Geophys. J.R. Astron. Soc. **49**, 723–737, 1977
Wiggins, R.A., Madrid, J.A.: Body wave amplitude calculations. Geophys. J.R. Astron. Soc. **37**, 423–433, 1974

Received August 1, 1978; Revised Version July 16, 1979

J. Geophys. 46, 343–347, 1979

Journal of Geophysics

Short Communication

Macroseismic Field of the Earthquake of September 3, 1978, in the Swabian Jura

D. Procházková[1], G. Schneider[2], E. Schmedes[3], J. Drimmel, E. Fiegweil, G. Lukeschitz[4], J. Vogt, P. Courtot, P. Godefroy[5], G. Grünthal[6], D. Mayer-Rosa, R. Berger[7]

[1] Geophysical Institute of Czechosl. Acad. Sci., 141 31 Praha 4 – Spořilov, Czechoslovakia
[2] Institut für Geophysik der Universität Stuttgart, Richard-Wagner-Str. 44, D-7000 Stuttgart, Federal Republic of Germany
[3] Inst. für Allgemeine und Angewandte Geophysik der Universität, Theresienstr. 41, D-8000 München, Federal Republic of Germany
[4] Zentralanstalt für Meteorologie und Geodynamik, Hohe Warte 38, A-1190 Wien, Austria
[5] Bureau de Recherches Géologiques et Minières, F-450 18 Orléans, Cédex, France
[6] Zentralinstitut für Physik der Erde, Telegrafenberg, Potsdam, German Democratic Republic
[7] Institute of Geophysics, E.T.H. Hönggerberg, Zürich, Switzerland

Key words: Seismology – Swabian Jura – Macroseismic observations.

1. Introduction

The earthquake of September 3, 1978, 05 h 08 m 32 s UT in the region of the Western Swabian Jura was felt in the Federal Republic of Germany, Austria, Czechoslovakia, France, German Democratic Republic, Switzerland and only weakly in Northern Italy. The parameters of the main shock are as follows:

Geographic latitude:	48, 29° N
Geographic longitude:	8, 94° E
Origin time:	05 h 08 m 31.8 s UT
Focal depth:	9 km
Surface wave magnitude:	5,4

(All data after CSEM = Centre Séismologique Europeo-Meditteranéen, Strasbourg)

Epicentral macroseismic intensity: 7–8

(On the macroseismic scale after MEDVEDEV, SPONHEUER, KÁRNÍK = MSK-64: All 'intensity' data in this paper are related to this scale).
Seismic focal moment: $3.35.\ 10^{16}$ Nm
(This value is based on the broad-band record of the Central Seismological

0340-062X/79/0046/0343/$01.00

Observatory, Erlangen-Gräfenberg, assuming a P-velocity of $7.35 \cdot 10^3$ m/s along the propagation path).

At the seismic station Mössingen (MSG) about 340 aftershocks with a local magnitude larger than 1.0 have been recorded until the end of 1978. The strongest aftershock took place on September 3, 1978, 10 h 02 m 43,3 s. Its surface magnitude reached 4.4 (after CSEM).

2. Description of the Macroseismic Field

Austria. In order to increase the number of reports 649 additional inquiries were sent to police stations, post offices and schools. From 179 districts or points in Austria (besides 8 reports from abroad) 377 reports and 539 negative answers could be used in working out the macroseismic map. As Fig. 1 shows, a maximum intensity of 5 was experienced in the northwestern part of Vorarlberg. The isoseismal intensity 3, being practically the limit of perceptibility, extends to epicentral distances of 350 km near Kötschach-Mauthen in Carinthia and 550 km near Vienna. In Austria 48,000 km^2, about 58% of the whole country felt a minimum intensity of 3, 9,600 km^2 an intensity 4 and 450 km^2 an intensity 5. No damage was reported in Austria.

Czechoslovakia. Czechoslovakia is situated at the border of the macroseismic field, and, therefore, only singular observations have been collected. The intensities determined on the basis of these observations show values of 3 to 4.

The intensity 4 was observed in the south-western part of Bohemia at isolated places. In view of the small number of observations, (15 reports from 10 localities) the isoseismal was marked by a dashed line (Fig. 1). An intensity 3 was not felt by higher percentage of the population because the shock occurred early in the morning when people slept still.

Federal Republic of Germany. The epicentral area of the shock is situated in Baden-Württemberg. The pleistoseismal area is characterized by an epicentral intensity of 7–8, restricted to two suburbs Onstmettingen and Tailfingen of Albstadt (Fig. 2). The most damage in these areas was situated above the focal line of a unilateral fracture running from the hypocenter towards the South. The macroseismic 'islands', North and Northeast of Stuttgart correspond to communities with many tall buildings in the industrialized environs of the capital of Baden-Württemberg (Fig. 1). The greater part of Baden-Württemberg, outside of the epicentral area, felt an intensity of 4–5. In Bavaria the shock was felt with intensities between 3 and 5. Reports came from 70 localities. No observations were available from the central and the northern parts of the Federal Republic of Germany and, therefore, the isoseismal for intensity 4 was extrapolated by a dotted line.

France. The intensities observed in the French regions of Alsace, Vosges, Lorraine and N.E. of Franche-Comté ranged from 2 to 6; an intensity 6 was

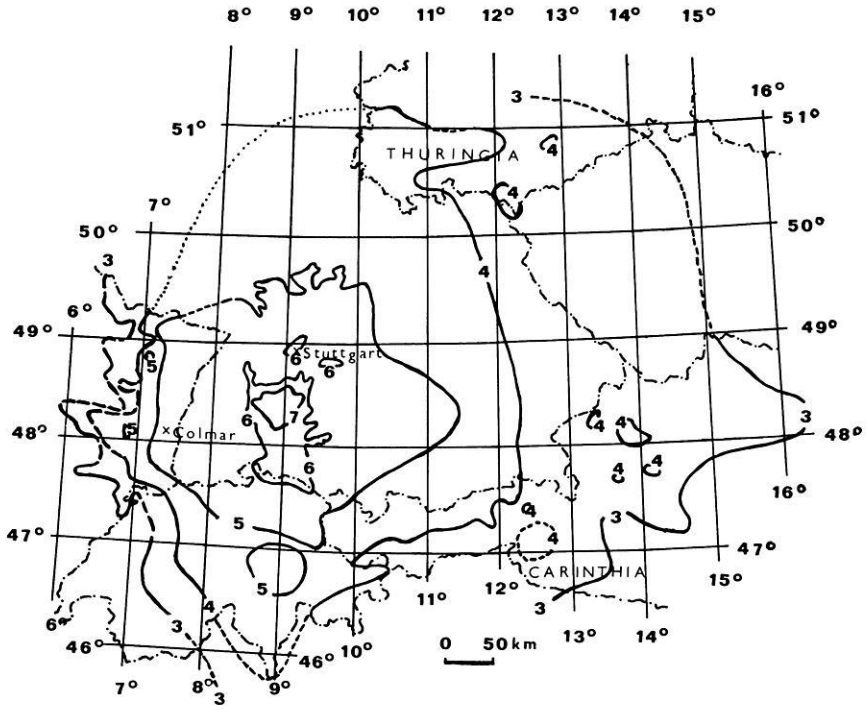


Fig. 1. Isoseismal map

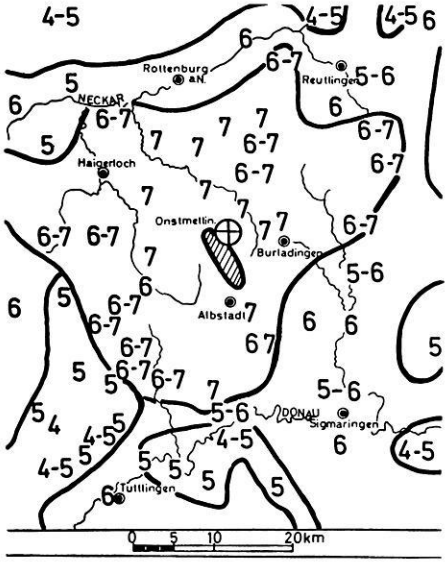


Fig. 2. Macroseismic intensity in the epicentral area

observed in a small area south of the city of Colmar (with regard to the small size of this area, this isoseismal is not marked in Fig. 1). Altogether 400 reports were obtained. The macroseismic field in France extended to some 30,000 km^2. The shock was also felt by most participants of the European Seismological Commission in Strasbourg.

German Democratic Republic. For collection of the macroseismic observations in the German Democratic Republic a questionnaire-inquiry was carried out. From 54 towns or villages detailed reports were obtained while from 93 localities no replies were recieved. Most localities observed an intensity 4, smaller intensities were observed rarely, because the shock occurred early in the morning and small intensities as a rule do not wake people. Intensities of 4 were observed in the southwestern part of the German Democratic Republic, in the western part of Thuringia and to the north as far as to the line Mühlhausen-Erfurt-Weimar-Jena-Gera. South of the line Gera-Suhl and in Southeastern Thuringia the shock was not felt. This lack of perceptions is well verified. Some further reports of intensity 4 were obtained from the Vogtland.

Switzerland. 600 reports from 320 villages all over the country were evaluated. According to these data, the earthquake was strongly felt in the northern, central and southern part of Switzerland. On the other hand, it was not felt in the west, such as in the Jura Mountains and in the Lake Geneva area. The strongest effects were observed in north-eastern Switzerland along the borderline to the Federal Republic of Germany particularly in the Lake of Constance area and Kanton Schaffhausen.

In this region maximum intensity was 6 and many reports contained descriptions of moving objects and displaced furniture. In the Kanton of Zürich and Schaffhausen more than 100 cases of slight damage (mainly cracked walls), were reported to the insurance companies. But not more than 20% of these cracks could be proved of recent nature. In many descriptions, the duration of the strong shaking was estimated between 5 and 30 s. A significant number of people indenpendently reported earthquake-generated noise, comparable to 'trucks passing by'. In general the earthquake was felt more as a 'rocking or rolling' movement with relatively low frequency, and was especially marked in high-rise buildings. This caused some problems in using the MSK-64 scale, since the observations had to be corrected subjectively in intensity by a certain factor.

3. Remarks

On the basis of the observations described in paragraph 2 the isoseismal map in Fig. 1 was compiled. From this map follows:

1. For the macroseismic determination of the focal depth, the mean radii of isoseismals were calculated using the surfaces encircled by the isoseismals. The results are as follows: $r_7 = 20$ km, $r_6 = 41$ km, $r_5 = 135$ km, $r_4 = 230$ km and $r_3 =$ ca. 330 km. Using the formulae of Kövesligethy and Blake with the

parameters $k=3{,}4$ and $\alpha=0{,}001$ (Procházková, 1979) a focal depth of about 10 km, was determined. This value is of the same order of magnitude as the value given by CSEM.

2. A comparison with the effects caused by the famous 1911 earthquake (Hiller, 1960, Procházková and Kárník, 1978) in the same area shows, that the intensities outside of the epicentral area for the 1978 earthquake were half to one degree lower than in 1911.

3. On the isoseismal map (Fig. 1) the islands of greater intensity may be caused by differences in the local geological foundations. For example the old shields, to which the Bohemian Massif belongs, are characterized by small absorption of seismic energy (e.g., a great distance between isoseismals is observed, Procházková, 1979). In Thuringia the course of the isoseismals shows a conspicuous form. The zone without macroseismic observations in southeastern Thuringia corresponds well with an area of Paleozoic basement, while north of the line Gera-Suhl on a Mesozoic foundation the shock was felt with an intensity 4. This zone in southeastern Thuringia where no effects were observed, was reported previously in macroseismic investigations of other Swabian Jura earthquakes and so seems to be a typical effect in the case of these earthquakes.

Acknowledgement. The authors are indebted to Dr. V. Kárník for discussion and critical comments.

References

Hiller, W.: Die Erdbeben der Zollernalb. Der Landkreis Balingen, **1**, 27–37, 1960

Procházková, D., Kárník, V. (editors): Atlas of Isoseismal Maps for Central and Eastern Europe. Geoph. Inst. of the Czechosl. Acad. Sci., Prague, 1978

Procházková, D.: Parameters of the Macroseismic Fields in Central and Eastern Europe. (in Czech.). Geophysical Conference in Liblice in 1979, in press 1979

Received June 18, 1979; Revised Version September 7, 1979; Accepted September 13, 1979

J. Geophys. 46, 349–367, 1979

Journal of Geophysics

Approximate Diffraction Theory for Transparent Half-Planes With Application to Seismic-Wave Diffraction at Coal Seams

J. Fertig* and G. Müller**

Geophysical Institute, University of Karlsruhe, Hertzstr. 16, D-7500 Karlsruhe, Federal Republic of Germany

Abstract. Starting with the exact theory of diffraction of plane SH or plane acoustic P waves at an opaque half-plane, i.e., a rigid screen or a crack, an approximate theory is given for diffraction at a transparent half-plane, realized, e.g., by a thin layer in a homogeneous medium. The results are extended for linesource excitation. The diffraction formula is similar to formulas based on Kirchhoff diffraction theory, but it includes both a term related to the reflected wavefield and a term related to the direct plus transmitted field. Moreover, use is made of the reciprocity principle. A comparison with finite-difference calculations for SH waves shows that the approximate theory has a rather broad range of applicability. Results of calculations are presented which are related to wave-propagation problems encountered in seismic prospecting for coal: the reflections and diffractions generated by a sequence of coal seams with an offset along a fault are calculated, and the diffractions produced by a realistic vertical offset of a horizontal seam are studied in some detail.

Key words: Diffraction theory – Theoretical seismograms – Seismic prospecting for coal.

Introduction

The seismic reflection response of coal-seam sequences normally has a complicated pattern. The three main reasons for this are: (1) the seams are often closely grouped such that individual seams cannot be resolved even with high frequencies, (2) wave conversion upon oblique incidence of the wave from the source can produce strong additional reflections, (3) seam offsets due to faults

* *Present address:* Preussag AG, Erdöl und Erdgas, Arndtstr. 1, D-3000 Hannover, Federal Republic of Germany

** *Present address:* Institute of Meteorology and Geophysics, University of Frankfurt, Feldbergstr. 47, D-6000 Frankfurt, Federal Republic of Germany

0340-062X/79/0046/0349/$03.80

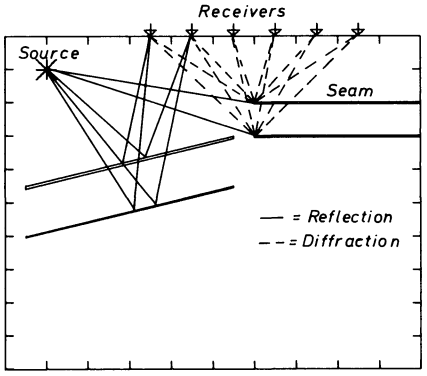


Fig. 1. Two-dimensional model with an arbitrary number of seams of finite width, and reflected and diffracted rays from the line source to the receivers

cause diffractions and offsets in the reflections. We have investigated the first two effects in an earlier paper (Fertig and Müller, 1978) with the aid of the reflectivity method and assumed for these purposes horizontally layered seam sequences without offsets. In the present paper we are mainly interested in the effects produced by the interruption and termination of seams due to faulting, and we present an approximate method for the calculation of theoretical seismograms in such cases. Such theoretical seismograms can help to clarify the circumstances under which coal-seam offsets can be detected by seismic measurements.

Our model of the subsurface is two-dimensional, the source of waves is a line source, and we consider only the SH-wave and the acoustic P-wave case. This is a restriction to the simplest conditions of wave propagation which, nevertheless, are of practical importance; for instance, SH-waves may well become a more routinely used tool in seismic exploration because of their greater simplicity and resolving power, compared with P-waves. The seams are assumed to be thin plane layers of finite width, located horizontally or non-horizontally in an otherwise homogeneous medium (Fig. 1). The plane-wave reflection and transmission response of an individual seam is calculated with the reflectivity method and corrected by the geometrical-spreading factor of cylindrical waves. Multiples between the seams are disregarded; calculations for horizontal seam sequences with the reflectivity method, which yield also these multiples, show that this neglection is often permitted, inspite of the large reflection coefficients of coal seams. The diffractions from the seam ends are calculated with an approximate theory which is an extension of the exact theory of diffraction of plane SH-waves at an opaque half-plane. This approximate diffraction theory is described in the first part of the paper, and its results for a special case are compared with the results of finite-difference calculations.

Later in this paper we present a few examples of theoretical seismogram sections. The computations illustrate the reflection and diffraction of P waves at a single seam, and at a sequence of dipping seams which are separated into two blocks by a fault. Finally, we investigate the reflection, transmission and diffraction of P and SH waves at a realistic vertical offset of a horizontal seam.

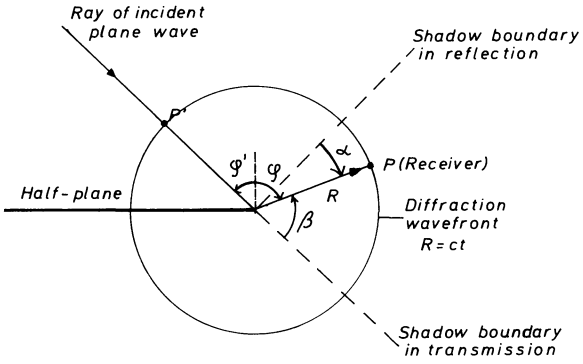


Fig. 2. Definition of diffraction angles α and β, relative to the shadow boundaries, and of the angles φ' and φ. α and β as indicated are positive

Approximate Diffraction Theory for a Transparent Half-Plane

Plane-Wave Excitation. Our starting point is the exact theory of diffraction of a plane SH wave with unit-step function time behavior at an opaque half-plane (either a rigid screen or a crack), as given, e.g., by Pao and Mow (1973, pp. 572–586). In a first step, we introduce into their formulas (5.15) for a rigid screen and (5.22) for a crack the diffraction angles α and β with respect to the shadow boundaries in reflection and transmission (see Fig. 2). Then, we observe that Pao and Mow's formulas give the *total* displacement field for distances R from the edge of the half-plane less than or equal to ct, where c is the S-wave velocity of the medium and t the time relative to the arrival of the incident plane wave at the edge. In order to obtain the *diffracted* displacement field alone we have to subtract the direct wave for receivers with $\alpha>0$ and $\beta>0$, and the direct plus the reflected wave for receivers with $\alpha<0$ and $\beta>0$. The result is:

$$\Psi^{(0)}_{\text{diff}}=\frac{1}{\pi}\left\{r\arctan\frac{\left[\frac{c}{2R}\left(t-\frac{R}{c}\right)\right]^{\frac{1}{2}}}{\sin\frac{\alpha}{2}}-\arctan\frac{\left[\frac{c}{2R}\left(t-\frac{R}{c}\right)\right]^{\frac{1}{2}}}{\sin\frac{\beta}{2}}\right\}H\left(t-\frac{R}{c}\right) \tag{1}$$

The coefficient r is -1 for a rigid screen and $+1$ for a crack, and $H(t)$ is the unit-step function.

The first term in (1) dominates close to the shadow boundary in reflection, where it is discontinuous and jumps from $rH(t-R/c)/2$ for $\alpha>0$ to $-rH(t-R/c)/2$ for $\alpha<0$. If for $\alpha<0$ the reflection $rH(t-R/c)$ is superposed, the total wave field is continuous at the shadow boundary in reflection, as it should be. Similarly, the second term in (1) dominates close to the shadow boundary in transmission ($\beta=0$). Its discontinuity is removed by adding, for $\beta>0$, the direct wave.

The generalization of the first term of (1) for a transparent half-plane is straightforward, if one considers the fact that r in (1) is just the plane-wave reflection coefficient of the half-plane. As in usual Kirchhoff diffraction theory,

one obtains a convolution of the plane-wave reflection response $r(t)$ of the half-plane [or more precisely of its derivative $r'(t)$] with a diffraction operator which here is the first arc tan function in (1). The second term in (1) has to be interpreted as the contribution to diffraction from the direct wave. In order to generalize this term for a transparent half-plane we have to include additionally a contribution from the transmitted wave on the underside of the half-plane. If the direct wave is $d(t)$ [not necessarily equal to $H(t)$] and the plane-wave transmission response of the half-plane $b(t)$, we arrive at the convolution of the difference $b'(t)-d'(t)$ and the second arc tan function in (1). The total diffraction is

$$\Psi_{\text{diff}}^{(1)}=r'(t)*\frac{1}{\pi}H\left(t-\frac{R}{c}\right)\arctan\frac{\left[\frac{c}{2R}\left(t-\frac{R}{c}\right)\right]^{\frac{1}{2}}}{\sin\frac{\alpha}{2}}$$

$$+[b'(t)-d'(t)]*\frac{1}{\pi}H\left(t-\frac{R}{c}\right)\arctan\frac{\left[\frac{c}{2R}\left(t-\frac{R}{c}\right)\right]^{\frac{1}{2}}}{\sin\frac{\beta}{2}}.$$

This formula has the disadvantage that it does not satisfy the reciprocity principle which in the present case requires unchanged displacements, when (Fig. 2) the receiver is shifted to P' and the ray of the incident plane wave passes through P. In an ad-hoc procedure we enforce reciprocity by averaging the two diffractions; this step, of course, needs justification which will be given later. Averaging the diffractions is equivalent to averaging the plane-wave reflection and transmission responses, since the diffraction operators are the same. If we define

$\bar{r}(t)[\bar{b}(t)]=$ arithmetic mean of the half-plane reflection (transmission) responses to a plane wave $d(t)$ with angle of incidence φ' and φ, respectively (see Fig. 2),

we obtain the final diffraction formula for plane-wave excitation:

$$\Psi_{\text{diff}}^{(2)}=\bar{r}'(t)*\frac{1}{\pi}H\left(t-\frac{R}{c}\right)\arctan\frac{\left[\frac{c}{2R}\left(t-\frac{R}{c}\right)\right]^{\frac{1}{2}}}{\sin\frac{\alpha}{2}}$$

$$+[\bar{b}'(t)-d'(t)]*\frac{1}{\pi}H\left(t-\frac{R}{c}\right)\arctan\frac{\left[\frac{c}{2R}\left(t-\frac{R}{c}\right)\right]^{\frac{1}{2}}}{\sin\frac{\beta}{2}} \tag{2}$$

Discussion of Formula (2). For an opaque half-plane and the direct wave $d(t)=H(t)$, we have $\bar{r}(t)=\pm H(t)$ and $\bar{b}(t)=0$, and (2) reduces to (1). Moreover, $\Psi_{\text{diff}}^{(2)}$ satisfies all continuity requirements of the total field at the wavefront of the

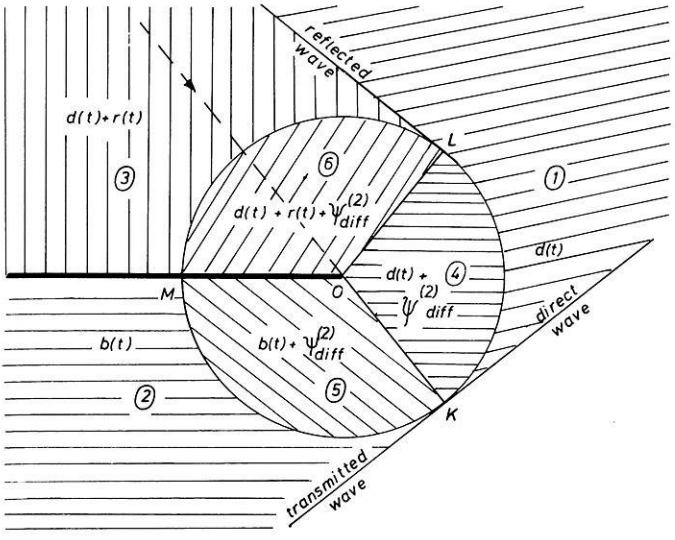


Fig. 3. Wave fronts for reflection, transmission and diffraction of a plane SH wave at a transparent half-plane. The total displacement in six different domains is indicated. For details see text

diffraction and at the shadow boundaries. In order to show this, we consider the wavefront picture for a fixed time in Fig. 3; it is subdivided into six different domains in each of which the *total* displacement field is indicated. The displacements at the wavefront of the diffraction are continuous, since $\Psi^{(2)}_{\text{diff}}$ starts from zero at $t=R/c$. The continuity of the displacements at the shadow boundaries OK and OL will now be demonstrated in the case of OK. The first term of (2) is continuous across OK. The second term in domain ⑤ close to OK (i.e., for small negative β) is

$$[\bar{b}'(t)-d'(t)]*\frac{1}{\pi}H\left(t-\frac{R}{c}\right)\frac{-\pi}{2}\approx-\tfrac{1}{2}b\left(t-\frac{R}{c}\right)+\tfrac{1}{2}d\left(t-\frac{R}{c}\right),$$

and in domain ④, also close to OK, it is

$$[\bar{b}'(t)-d'(t)]*\frac{1}{\pi}H\left(t-\frac{R}{c}\right)\frac{\pi}{2}\approx\tfrac{1}{2}b\left(t-\frac{R}{c}\right)-\tfrac{1}{2}d\left(t-\frac{R}{c}\right).$$

The total field in ⑤ close to OK is (addition of the transmitted wave)

$$\tfrac{1}{2}b\left(t-\frac{R}{c}\right)+\tfrac{1}{2}d\left(t-\frac{R}{c}\right)+\text{first term of (2)},$$

and in ④, also close to OK, it is (addition of the direct wave)

$$\tfrac{1}{2}b\left(t-\frac{R}{c}\right)+\tfrac{1}{2}d\left(t-\frac{R}{c}\right)+\text{first term of (2)},$$

i.e., the total field is continuous across OK. Similar results hold for the shadow boundary OL.

A limitation of the simple diffraction formula (2) is that $\Psi_{\text{diff}}^{(2)}$ for transparent half-planes does not satisfy the boundary conditions at points on the half-plane between O and M (Fig. 3). There is no direct way to determine the degree of disagreement, since in the framework of the simple theory presented only the plane-wave reflection and transmission responses of the half-plane are needed; the field inside the half-plane does not enter. All that can be done is to test (2) against exact results, calculated, e.g., with finite-difference methods. This will be discussed later.

In geophysical applications of (2) the transparent half-plane will be a homogeneous layer (e.g., a coal seam) or even a layer with internal layering. It is clear that (2) can only be used, if the dominant wavelength is at least several times the layer thickness. Furthermore, the S velocity *in* the layer should be less than in the surrounding medium, since otherwise refraction along the layer would occur and diffraction at the edge would start before the direct wave arrives there. Probably this contribution is small, but a more detailed investigation is necessary before one can conclude that (2) is a useful approximation also in the case of a high-velocity layer.

Line-Source Excitation. In a final step we derive from (2) the diffracted field due to excitation by a line source whose distance from the diffracting edge is R'. The incident wave at the edge is assumed to be

$$\Psi_{\text{inc}} = \frac{1}{R'^{\frac{1}{2}}} d\left(t - \frac{R'}{c}\right), \tag{3}$$

i.e., the line source radiates isotropically, beginning at $t=0$, and geometrical spreading is that for cylindrical waves. The corresponding formula for Ψ_{diff} should satisfy the following plausible requirements [see (2) for reference]:

(a) The first term should be, for sufficiently small α,

$$\frac{\operatorname{sign}\alpha}{2(R'+R)^{\frac{1}{2}}} r(t-t_d),$$

i.e., half the reflection response at the shadow boundary (apart from the sign). Likewise for small β the second term should be

$$\frac{\operatorname{sign}\beta}{2(R'+R)^{\frac{1}{2}}} [b(t-t_d) - d(t-t_d)].$$

Here and in the following, $t_d = (R'+R)/c$ is the arrival time of the diffraction.

(b) For larger values of α the first term of Ψ_{diff} should be $1/R'^{\frac{1}{2}}$ times the first term of (2), i.e.,

$$\frac{\bar{r}'(t)}{R'^{\frac{1}{2}}} * \frac{1}{\pi} H(t-t_d) \arctan \frac{\left[\frac{c}{2R}(t-t_d)\right]^{\frac{1}{2}}}{\sin\frac{\alpha}{2}}, \tag{4}$$

and similarly the second term of Ψ_{diff} for larger β should be

$$\frac{\bar{b}'(t)-d'(t)}{R'^{\frac{1}{2}}} * \frac{1}{\pi} H(t-t_d) \arctan \frac{\left[\frac{c}{2R}(t-t_d)\right]^{\frac{1}{2}}}{\sin\frac{\beta}{2}}. \tag{5}$$

The following form of Ψ_{diff} satisfies these requirements:

$$\Psi_{\text{diff}} = \frac{\bar{r}'(t)}{(R'+R)^{\frac{1}{2}}} * \frac{1}{\pi} H(t-t_d) \arctan \frac{\left[\frac{c}{2R}(t-t_d)\right]^{\frac{1}{2}}}{\left(\frac{R'}{R'+R}\right)^{\frac{1}{2}} \sin\frac{\alpha}{2}}$$

$$+ \frac{\bar{b}'(t)-d'(t)}{(R'+R)^{\frac{1}{2}}} * \frac{1}{\pi} H(t-t_d) \arctan \frac{\left[\frac{c}{2R}(t-t_d)\right]^{\frac{1}{2}}}{\left(\frac{R'}{R'+R}\right)^{1/2} \sin\frac{\beta}{2}} \tag{6}$$

The first of the above conditions is fulfilled, since for small α or β the additional term in the argument of the arc tan functions, $[R'/(R'+R)]^{\frac{1}{2}}$, does not change the step-function behavior of the diffraction operators. The second condition is approximately satisfied, since for larger α and β the arc tan functions in (4), (5) and (6) can be expanded into Taylor series with restriction to the first term. In this approximation, which actually is a high-frequency or wavefront approximation, (4) and the first term of (6) agree, and likewise (5) and the second term.

Formula (6) was derived for SH waves, and Ψ_{diff} is the out-of-plane displacement of the diffracted wave. This formula can also be used for acoustic P waves. In this case Ψ_{inc} in (3) or Ψ_{diff} in (6) can be either the displacement potential, the pressure or the displacement along the ray, and $\bar{r}(t)$ and $\bar{b}(t)$ are the averaged compressional reflection and transmission responses. In the applications, given later in this paper, mainly the acoustic case is treated. Formula (6) can, in principle, also be applied in the case of P-SV waves in solid media; a few corresponding remarks are made at the end of this paper.

The treatment of diffraction at a half-plane with Kirchhoff diffraction theory (Berryhill, 1977; Trorey, 1977) so far has given results similar to the first term in (6). No contribution similar to the second term has been found. The reason is that only backward diffraction was of interest and hence only the reflected field at the topside of the half-plane was continued upwards by Kirchhoff's formula. However, depending on the parameter contrasts between the half-plane and the surrounding medium and hence on the difference between the direct wave $d(t)$ and the averaged transmission response $\bar{b}(t)$, the second term in (6) which dominates the forward diffraction can also be of importance for backward diffraction. Another difference to Kirchhoff diffraction theory is the incorporation of reciprocity in (6) which gives a considerably broader range of applicability, as will be seen later.

Comparison of the Approximate Diffraction Theory With Finite-Difference Calculations

To test formula (6) we computed synthetic seismograms, both with (6) and with a finite-difference method, for vertical incidence of an SH wave on a transparent half-plane which is represented by a thin homogeneous layer.

Plane-Wave Response of a Homogeneous Layer. In the following we summarize the formulas which describe the reflection and transmission of a plane wave $d(t)$ by a homogeneous layer of infinite extent in all directions, embedded in a full-space. Besides the SH-wave case we consider also the acoustic P-wave case. The full-space is characterized by the P velocity α_1, the S velocity β_1 and the density ρ_1. The corresponding parameters of the layer are α_2, β_2 and ρ_2; its thickness is h. The plane-wave reflection and transmission coefficients, r_{ss} or r_{pp} and b_{ss} or b_{pp}, for monochromatic waves of circular frequency ω are (j = imaginary unit):

SH waves	Acoustic P waves	
$r_{ss}=\dfrac{r_0(1-E)}{1-r_0^2E}$	$r_{pp}=\dfrac{r_0(1-E)}{1-r_0^2E}$	(7)
$b_{ss}=\dfrac{1-r_0^2}{1-r_0^2E}$	$b_{pp}=\dfrac{1-r_0^2}{1-r_0^2E}$	(8)

$$E=\exp\left(-2j\omega\frac{h}{v}\cos\varphi_2\right)$$

$v=\beta_2$	$v=\alpha_2$	
$r_0=\dfrac{\rho_1\beta_1\cos\varphi_1-\rho_2\beta_2\cos\varphi_2}{\rho_1\beta_1\cos\varphi_1+\rho_2\beta_2\cos\varphi_2}$	$r_0=\dfrac{\rho_2\alpha_2\cos\varphi_1-\rho_1\alpha_1\cos\varphi_2}{\rho_2\alpha_2\cos\varphi_1+\rho_1\alpha_1\cos\varphi_2}$	(9)

φ_1 is the angle of incidence, and φ_2 the angle of refraction in the layer, related to φ_1 by Snell's law. r_0 is the plane-wave reflection coefficient of the upper boundary of the layer.

The reflection and transmission responses of the layer, $r(t)$ and $b(t)$, which are needed in the diffraction formula (6), follow from (7) and (8) by multiplication with the spectrum of the incident wave $d(t)$ and by an inverse Fourier transform; both the angle φ' and φ, defined in Fig. 2, enters as the angle of incidence.

For wavelengths much longer than the layer thickness h the following low-frequency approximations are useful [they are obtained from (7) and (8) by power-series expansions in ω]:

$$\left.\begin{aligned} r(t)&=\frac{r_0\tau}{1-r_0^2}\,d'(t)\\ b(t)-d(t)&=-\frac{r_0^2\tau}{1-r_0^2}\,d'(t)=-r_0r(t)\\ \tau&=\frac{2h}{v}\cos\varphi_2. \end{aligned}\right\} \tag{10}$$

The error of these approximations is less than 10%, if the wavelength in the layer is greater than 15 to 20 times h.

With these results the diffraction for line-source excitation is readily obtained from (6). To obtain the complete wavefield we have to add the direct, reflected or transmitted wave, depending on where the receiver is located. These contributions are $L^{-\frac{1}{2}}$ times $d(t-L/c)$, $r(t-L/c)$ or $b(t-L/c)$, where L is the length of the wavepath from the line source to the receiver and c is either α_1 or β_1. $r(t)$ and $b(t)$ are determined for the corresponding angle of incidence which normally is different from the angles φ' and φ.

Finite-Difference Method. When finite-difference methods are applied to the wave propagation in heterogeneous media, one has the choice to use either the equation of motion for homogeneous media, combined with a formulation of the boundary conditions at the boundaries between different homogeneous parts of the medium, or to use the equation of motion for heterogeneous media. We chose the second method (Kelly et al., 1976), mainly because it allows also the modelling of complicated structures. The two-dimensional equation of motion for SH waves in this case and for spatially constant density is

$$\frac{\partial^2 v}{\partial t^2}=\frac{\partial}{\partial x}\left(\beta^2\frac{\partial v}{\partial x}\right)+\frac{\partial}{\partial z}\left(\beta^2\frac{\partial v}{\partial z}\right). \tag{11}$$

Here, x and z are the two spatial coordinates, $v(x,z,t)$ is the out-of-plane displacement, and $\beta(x,z)$ the S velocity. The density was kept constant in order to save computer storage and time; including an inhomogeneous density distribution into (11) would pose no problems.

In the finite-difference approximation of (11) we use standard central differences. If the time step is Δt and the grid spacings are Δx and Δz, and if $v^M_{K,L}$ and $\beta_{K,L}$ are the discrete approximations of $v(K\Delta x, L\Delta z, M\Delta t)$ and $\beta(K\Delta x, L\Delta z)$, we obtain:

$$\frac{\partial^2 v}{\partial t^2}=\frac{1}{\Delta t^2}(v^{M+1}_{K,L}-2v^M_{K,L}+v^{M-1}_{K,L})$$

$$\begin{aligned}\frac{\partial}{\partial x}\left(\beta^2\frac{\partial v}{\partial x}\right)=\frac{1}{2\Delta x^2}&[(\beta^2_{K+1,L}+\beta^2_{K,L})(v^M_{K+1,L}-v^M_{K,L})\\&-(\beta^2_{K,L}+\beta^2_{K-1,L})(v^M_{K,L}-v^M_{K-1,L})]\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial z}\left(\beta^2\frac{\partial v}{\partial z}\right)=\frac{1}{2\Delta z^2}&[(\beta^2_{K,L+1}+\beta^2_{K,L})(v^M_{K,L+1}-v^M_{K,L})\\&-(\beta^2{}_{K,L}+\beta^2_{K,L-1})(v^M_{K,L}-v^M_{K,L-1})]\end{aligned}$$

Then (11) yields for $\Delta z=\Delta x$:

$$\begin{aligned}v^{M+1}_{K,L}=&2v^M_{K,L}-v^{M-1}_{K,L}\\&+\left(\frac{\Delta t}{\Delta x}\right)^2[\beta^2_E(v^M_{K+1,L}-v^M_{K,L})-\beta^2_W(v^M_{K,L}-v^M_{K-1,L})\\&+\beta^2_S(v^M_{K,L+1}-v^M_{K,L})-\beta^2_N(v^M_{K,L}-v^M_{K,L-1})]\end{aligned} \tag{12}$$

$$\beta^2_E=\tfrac{1}{2}(\beta^2_{K+1,L}+\beta^2_{K,L})\qquad \beta^2_W=\tfrac{1}{2}(\beta^2_{K,L}+\beta^2_{K-1,L})$$

$$\beta^2_S=\tfrac{1}{2}(\beta^2_{K,L+1}+\beta^2_{K,L})\qquad \beta^2_N=\tfrac{1}{2}(\beta^2_{K,L}+\beta^2_{K,L-1})$$

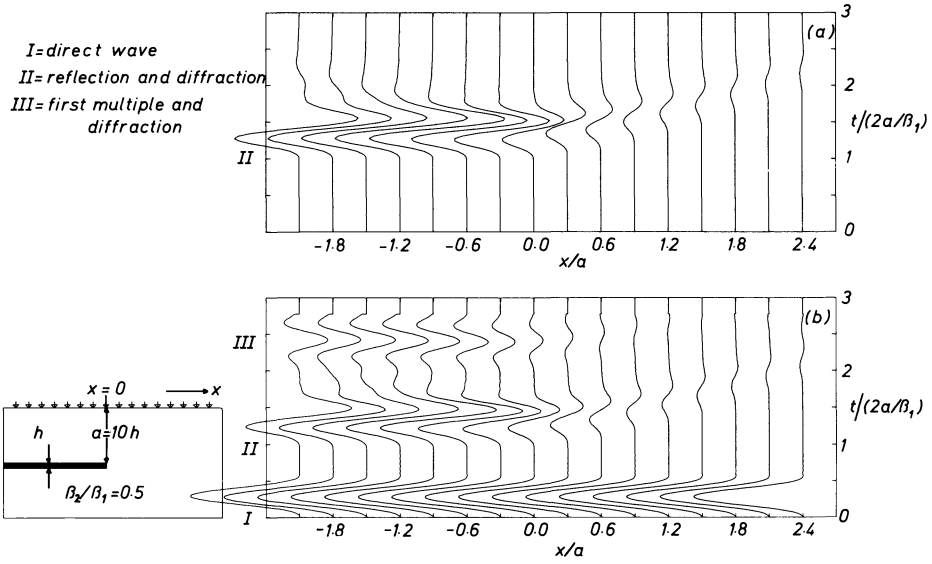


Fig. 4a and b. Reflection and diffraction of plane, vertically travelling SH waves at a transparent half-plane: **a** analytical calculation; **b** finite-difference calculation

To keep the explicit scheme (12) stable we use the relation $\Delta t = \Delta x/(\beta_{\max}\sqrt{2})$ between the time step and the grid spacing, where $\beta_{\max}$ is the maximum S velocity of the model (Boore, 1972). Grid dispersion is reduced by choosing Δx equal to about 1/12 of the dominant wavelength in those parts of the model with the minimum S velocity, $\beta_{\min}$.

The scheme (12) is supplemented by prescribed displacements at the source, by the boundary condition of vanishing stress at the boundaries of the rectangular model, and by initial conditions corresponding to vanishing displacements and particle velocities for $t=0$.

First-order discontinuities parallel to the x or z axis are modelled by a jump in β over one grid spacing and are thus located approximately in the middle between two neighbouring grid lines. In our numerical study of the reflection and diffraction of SH waves at a thin layer the layer velocity is assigned to two grid lines. Hence, if we want to calculate the response of a layer of thickness h, Δx is approximately $h/2$, and this value is used in our calculations. This is an approximation which can be wrong perhaps by 10 to 20%; only within these limits the finite-difference seismograms can be considered as exact and thus as a reference for analytically calculated seismograms.

Results. The model and the receiver geometry are illustrated in Fig. 4. A vertically travelling plane SH wave leaves the receiver level at $t=0$ and is reflected and diffracted at a thin layer with thickness h, velocity $\beta_2 = 0.5\,\beta_1$ and density $\rho_2 = \rho_1$. For the analytical calculation (Fig. 4a) formula (6) is used with a line source at large distance R' vertically above the diffracting edge, and corresponding time shifts and amplitude corrections are applied. The input

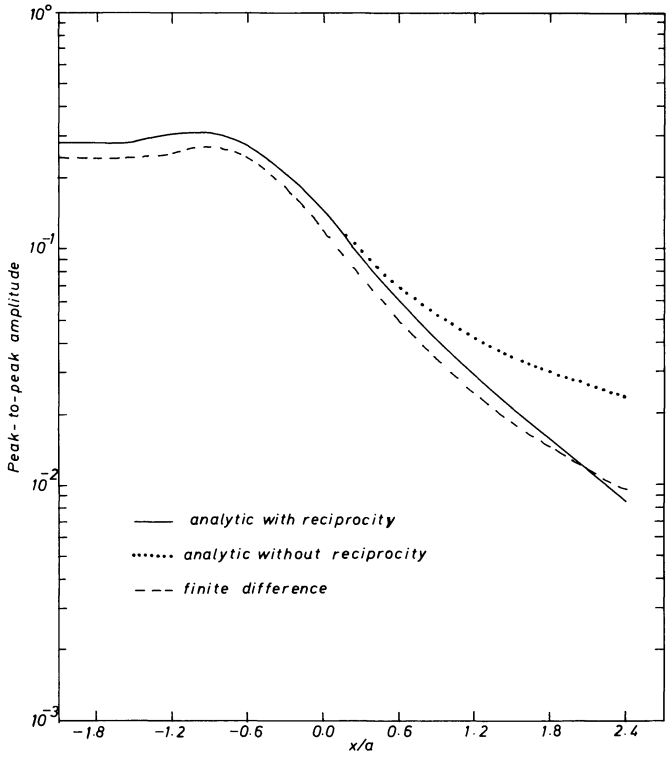


Fig. 5. Peak-to-peak amplitudes of the reflection and diffraction as a function of distance for the case shown in Fig. 4

signal has a simple bell-shaped form and unit maximum amplitude (see arrival I in Fig. 4b). The pulse duration corresponds to a length in the medium surrounding the layer of about 13 h and therefore is much larger than h. The distribution of receivers is symmetric with respect to the shadow boundary in reflection at $x = 0$. To the left of the shadow boundary one observes the reflection and the diffraction, to the right we have the diffraction alone. The finite-difference seismograms (Fig. 4b) display also the direct wave and the first multiple between the layer and the free surface, including the corresponding diffraction; the pulse forms of all three arrivals in Fig. 4b demonstrate nicely the differentiating effect of the reflection at thin layers. In the analytical calculation the receivers are assumed within the medium, whereas in the finite-difference calculation they are at the free surface which doubles the amplitudes of the reflections and diffractions. To correct for this difference the finite-difference seismograms were multiplied by 0.5. Hence, the reflections and diffractions in Fig. 4a and b are directly comparable.

The overall agreement of the pulse forms and amplitudes in this figure is quite good even to the largest distances x, i.e., to diffraction angles α up to 67°. A quantitative comparison of peak-to-peak amplitudes is given in Fig. 5. Taking into account the remarks made before on the accuracy of the finite-difference

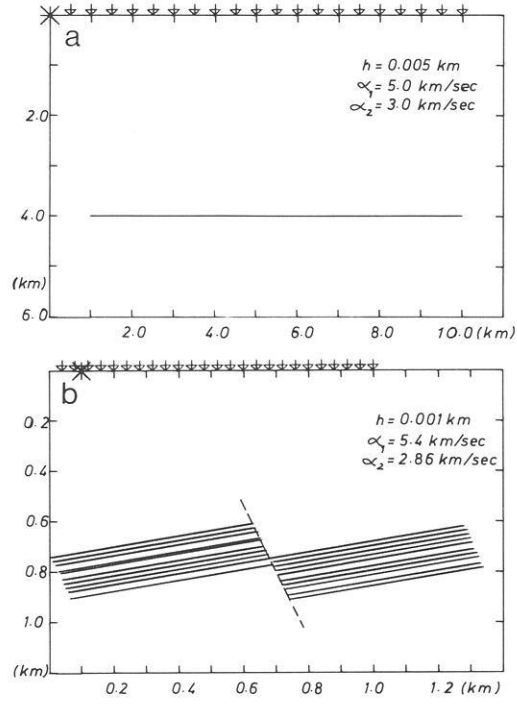


Fig. 6. a Thin layer of finite width embedded in a homogeneous medium. The positions of the line source and the receivers are marked at the top of the model. The density ratio is $\rho_2/\rho_1 = 0.85$.
b A sequence of coal seams which are offset by a fault. The density ratio is $\rho_2/\rho_1 = 0.82$

calculations, we conclude that the approximate diffraction formula (6) has a broad range of applicability and that the inclusion of reciprocity is a definite improvement. The limits of this formula are, however, reached when for fixed source position, i.e., for fixed angle of incidence φ' at the diffracting edge, and for variable receiver position, i.e., for variable angle φ, the averaged reflection response $\bar{r}(t)$ of the half-plane becomes zero or even changes its sign. [This cannot occur with the averaged transmission response $\bar{b}(t)$.] If the half-plane is a homogeneous layer whose P- and S-wave impedances are less than those of the surrounding medium, this can only happen in the case of SH waves. Here, the reflection coefficient r_0 in (9) changes its sign at the Brewster angle, and according to (7) or (10) r_0 determines the sign of the reflection response of the layer. The Brewster angle is

$$\varphi_B = \arcsin\left(\frac{\rho_1^2\beta_1^4 - \rho_2^2\beta_1^2\beta_2^2}{\rho_1^2\beta_1^4 - \rho_2^2\beta_2^4}\right)^{\frac{1}{2}} \quad (=63° \text{ in our example}).$$

In cases of practical interest, φ' is normally less than φ_B. If then φ increases and exceeds φ_B, $\bar{r}(t)$ will eventually drop to zero and change its sign. This trend is evident in Fig. 5 at the largest distances. Therefore, as a rule, φ should not exceed φ_B. No such basic limitation exists for acoustic P waves, since here r_0 has the same (negative) sign for all angles of incidence.

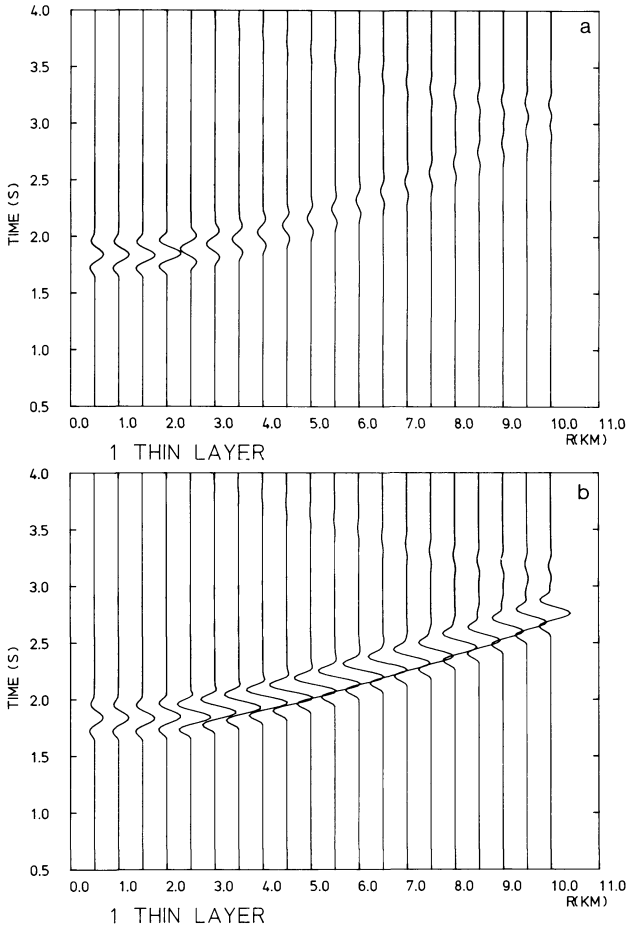


Fig. 7. a Acoustic diffraction response (vertical component) for the one-layer model of Fig. 6a. **b** The same as (**a**), but with the reflection response included

Applications

The first (purely synthetic) example is the diffraction and reflection response of a horizontal layer of finite width, excited by acoustic P waves from an explosive line source. The dimensions of the model, its parameters and the source and receiver geometry are given in Fig. 6a, and the seismograms for the vertical component in Fig. 7. The source pulse is one sine oscillation with smooth beginning and end. The dominant frequency is 2.5 Hz, corresponding to a ratio of wavelength (in the layer) to layer thickness of 240, i.e., the reflection from the layer is very well described by $r(t)$ in (10). Fig. 7a shows the two diffractions from the edges of the layer alone. The diffraction from the left edge is clearly visible, and, as expected, it changes its sign at the shadow boundary in reflection which is located at a horizontal distance of 2.0 km. The diffraction from the

right edge is rather weak, and its shadow boundary is beyond the profile section shown. Superposition of the reflection from the layer at distances greater than 2.0 km (Fig. 7b) makes the diffracted-reflected arrival continuous at the shadow boundary.

The model of the second example consists of 20 coal seams, each 1 m thick, which are offset by a fault (Fig. 6b). Such a model has been investigated by Dresen and Ullrich (1978) and Kerner (1978) with methods of model seismology. The dominant frequency is about 66 Hz, and the ratio of dominant wavelength to seam thickness is 43. The theoretical seismograms (Fig. 8) were again calculat-

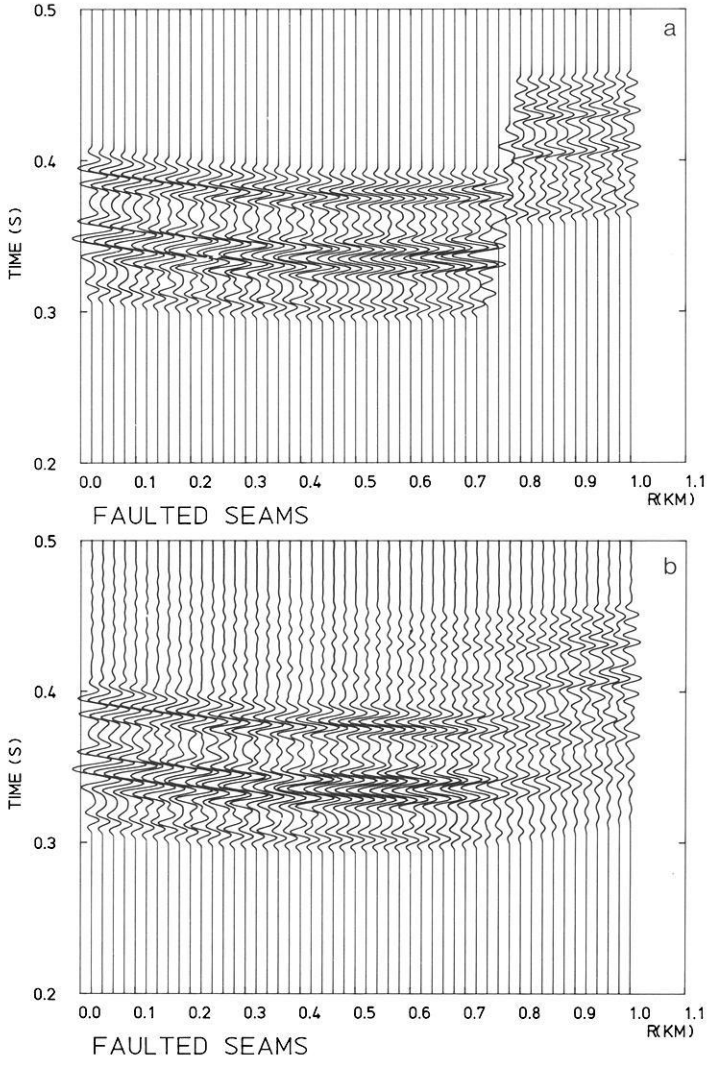


Fig. 8. a Acoustic reflection response (vertical component) for the model of faulted seams shown in Fig. 6b. **b** The same as (**a**), but with the diffractions from the seam ends at the fault included

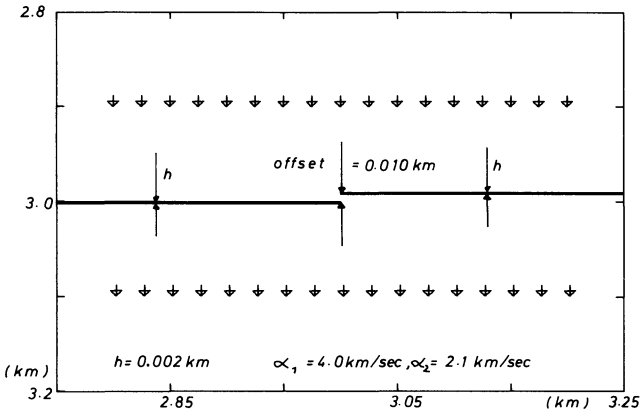


Fig. 9. Coal-seam model with an offset and two receiver profiles. For the upper profile the reflection and diffraction response is calculated (Figs. 10 and 12a), and for the lower profile the transmission and diffraction response (Figs. 11 and 12b). The density ratio is $\rho_2/\rho_1 = 0.5$. The line source is located 3 km above the offset, i.e., the incident wave is effectively plane

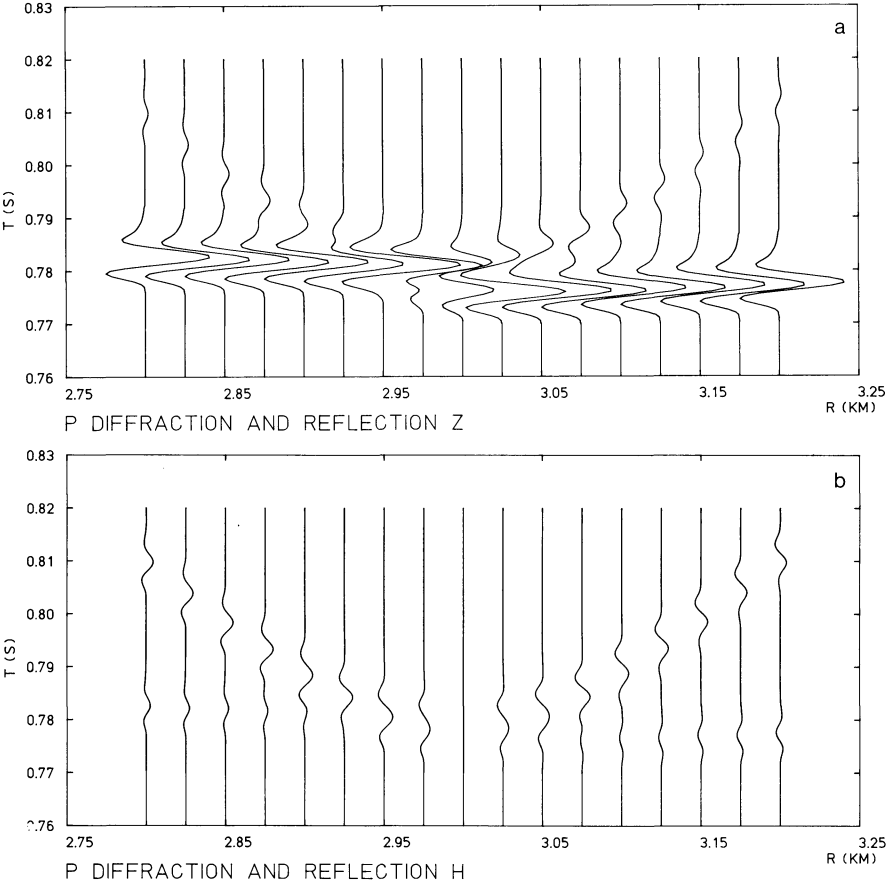


Fig. 10a and b. Acoustic reflection and diffraction response of the model in Fig. 9: **a** vertical component, **b** horizontal component. The amplitude scales in both sections are the same

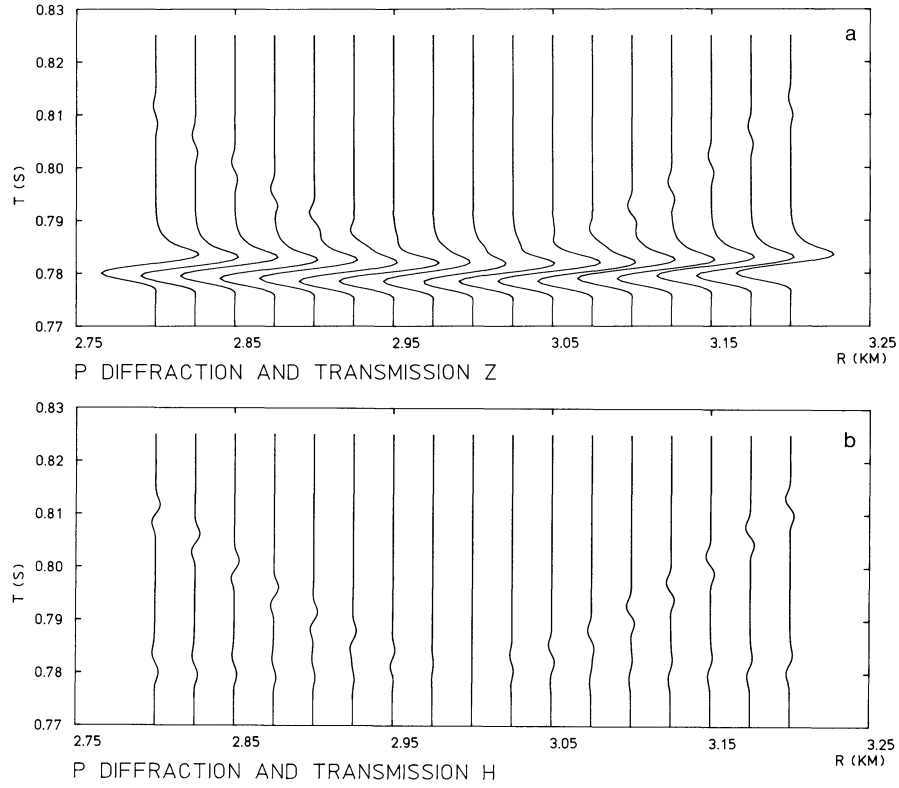


Fig. 11a and b. Acoustic transmission and diffraction response of the model in Fig. 9: **a** vertical component, **b** horizontal component. The amplitude scales in both sections are the same

ed for the acoustic case and the vertical component. Only primary reflections from the seams and primary diffractions from the seam ends were taken into account, and transmission effects at the seams were neglected. These approximations are justified for many cases of practical interest. In Fig. 8a the reflections are shown alone. They form a complicated interference pattern, consisting mainly of three wavegroups. The fault is very clear in this seismogram section. Including the diffractions from the seam ends at the fault masks the fault considerably (Fig. 8b), mainly because the reflections from the right block of seams are now weakened by the corresponding diffractions and do not stand out clearly against the diffractions from the left block which arrive earlier.

The last example deals with a realistic offset of a coal seam and the corresponding diffraction effects in the forward and backward direction for a source above the offset (Fig. 9). The situation would roughly correspond to an excitation of waves at the earth's surface and observations in tunnels close to the seam. Both the acoustic P wave case and the SH wave case have been treated for a source signal with a dominant frequency of 100 Hz (Figs. 10–12). As expected, the seam offset is most clearly seen in the reflected wave where it produces a time offset; this offset is larger for SH than for P waves because the *S*

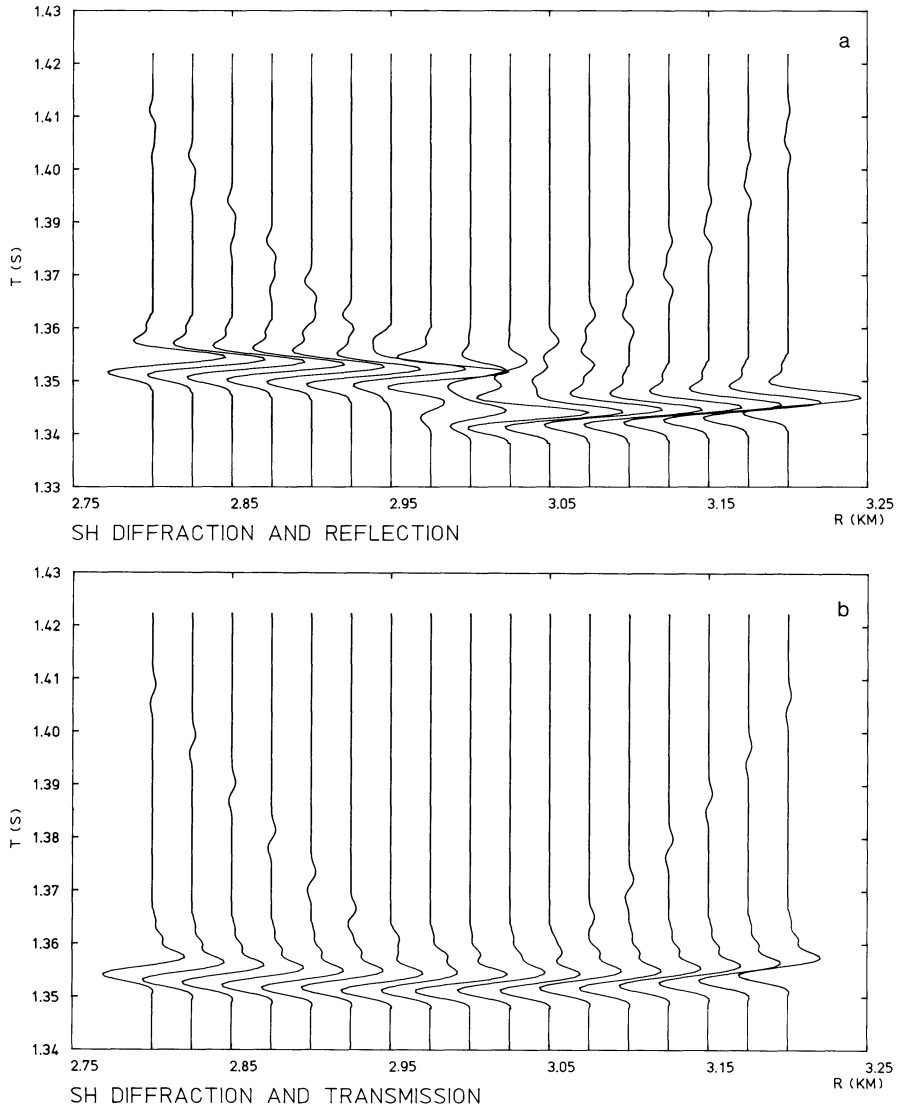


Fig. 12a and b. SH-wave reflection and diffraction response (**a**) and transmission and diffraction response (**b**) of the model in Fig. 9. The amplitude scales in both sections are the same

velocity is lower than the *P* velocity. Since in practice time offsets can also be generated by velocity heterogeneities along the whole wavepath and since they are not always eliminated by static corrections, it is reasonable to explore whether diffractions can be used, alone or additionally, to localize seam offsets. The seismograms for the horizontal component in the acoustic case (Figs. 10b and 11b) show that, indeed, diffractions may be prominent arrivals. Their prominence over the reflection or transmission in our example is due to the

large distance of the source above the seam which gives predominantly vertical polarization of the reflection and transmission. Such conditions would probably also have to exist in practical cases, but then horizontal-component recordings of P waves could actually help in identifying seam offsets. Seam depths less than the 3 km assumed in our example would require receiver profiles, which are correspondingly shorter and closer to the seam than in Fig. 9, and frequencies in excess of 100 Hz.

The applications of our computational method for theoretical seismograms show that it allows the modelling of subsurface structures typical for coal deposits; models more complicated than the one in Fig. 6b can easily be treated. Extensions of our method which have not yet been accomplished are the P–SV case and the inclusion of a layered overburden on top of the medium with the seams.

Conclusions

The reason for the relatively broad range of diffraction angles for which formula (6) is a good approximation is that the reciprocity principle has been incorporated, but it was shown that for large diffraction angles (6) is no longer applicable to SH waves. Similar restrictions could exist for other wave types, although not related to the Brewster angle. To find out where formula (6) really breaks down is a matter of further investigations and requires eventually finite-difference calculations also for the acoustic and the P–SV case.

The method for computation of theoretical seismograms presented in this paper, including both diffractions and reflected waves, allows fast calculations of the elastic response of relatively complicated structures. Therefore, it is possible to simulate effectively with this method the shooting and recording techniques of seismic prospecting such as the common-depth-point technique. The resulting seismogram sections could then be subjected to inversion procedures such as migration, and the latter would not just be the inverse operation of the modelling method. Thus, migration and other procedures could seriously be tested.

The generalization of the approximate diffraction theory for transparent half-planes to P–SV waves in solid media is straight-forward. The simplest cases are those of incident P waves and P diffractions and of incident SV waves and SV diffractions. Formula (6) can directly be applied, and the averaged plane-wave reflection and transmission responses, $\bar{r}(t)$ and $b(t)$, are those for the P–P and the SV–SV case, respectively. In the case of SV diffractions due to incident P waves (and similarly of P diffractions due to incident SV waves) the diffraction operators have to be included in the averaging process, and the geometrical spreading factors are slightly more complicated than the term $(R' + R)^{-\frac{1}{2}}$ in (6). The term of the averaging formula, which takes account of the P diffraction upon incidence of an SV wave, has to be set equal to zero for angles of incidence at the edge of the half-plane greater than the critical angle arc sin (β_1/α_1), since then no plane-wave reflection and transmission response exists.

Acknowledgments. This work was supported financially by Ruhrkohle AG and the government of Nordrhein-Westfalen. The computations were performed at the computing center, University of Karlsruhe. We thank Vlastislav Červený, Karl Fuchs, Michael Korn and Horst Stöckl for discussions on the subject of this paper, Karl Fuchs for reading the manuscript, and Ingrid Hörnchen for typing it. Contribution No. 181, Geophysical Institute, University of Karlsruhe.

References

Berryhill, J.R.: Diffraction response for nonzero separation of source and receiver. Geophysics **42**, 1158–1176, 1977

Boore, D.: Finite-difference methods for seismic wave propagation in heterogeneous materials. Methods in Computational Physics, Vol. 11. New York: Academic Press 1972

Dresen, L., Ullrich, G.: On the reflectivity of cyclically layered coal deposits – studies by means of two-dimensional models. Paper presented at 48th Annual Meeting of Society of Exploration Geophysicists, San Francisco 1978

Fertig, J., Müller, G.: Computations of synthetic seismograms for coal seams with the reflectivity method. Geophys. Prospect. **26**, 868–883, 1978

Kelly, K.R., Ward, R.W., Treitel, S., Alfred, R.M.: Synthetic seismograms: a finite-difference approach. Geophysics **41**, 2–27, 1976

Kerner, C.: Datenbearbeitung von Reflexionsseismogrammen, gemessen an Modellen des zyklisch geschichteten Steinkohlengebirges. Diploma Thesis, University of Bochum 1978

Pao, Y.-H., Mow, C.-C.: Diffraction of elastic waves and dynamic stress concentrations. 694 pp. New York: Crane Russak and London: Adam Hilger 1973

Trorey, A.W.: Diffractions for arbitrary source-receiver locations. Geophysics **42**, 1177–1182, 1977

Received August 28, 1979; Revised Version October 8, 1979

J. Geophys. 46, 369–384, 1979

Journal of Geophysics

Seismic Investigations of the Subcrustal Lithosphere Beneath Fennoscandia

B.R. Cassell [1] and K. Fuchs

Geophysikalisches Institut, Universität Karlsruhe, Hertzstr. 16, D-7500 Karlsruhe, Federal Republic of Germany

Abstract. A detailed *P*-velocity-depth model SCA1 of the structure of the subcrustal lithosphere beneath Scandinavia was derived using data from an approximately 600-km-long seismic profile running across the northern part of Scandinavia, the BLUE ROAD traverse. After examining the effect of a laterally varying Moho on later arrivals it was possible to reconstruct a *P*-velocity model containing five layers *P*1 to *P*5 of high velocity in the depth range from 50 to 100 km which appear to be embedded in material of normal to low seismic *P*-velocity. The amplitude ratios of these arrivals were compared with those from synthetic seismograms derived from the model.

These results were then used for a partial reinterpretation of Russian nuclear explosion observations recorded at NORSAR. The King and Calcagnile (1976) model KCA of monotonously increasing velocities in the subcrustal lithosphere produces significant discrepancies between observed and theoretical amplitudes which were found to be resolved by the insertion of the BLUE ROAD model. In addition, evidence for a further low velocity zone between 170.0 and 190.0 km depth was found.

Taking all findings into account the subcrustal lithosphere and upper mantle beneath Fennoscandia seem to consist of alternating layers of low and high seismic velocoty, their lateral extensions possibly not exceeding 100–160 km.

Key words: Lithosphere-Asthenosphere transition beneath Fennoscandia – Deep seismic sounding – Synthetic seismograms – Amplitude ratios.

[1] *Present address:* Department of Geodesy and Geophysics, University of Cambridge, Madingley Rise, Madingley Road, Cambridge CB3 OEZ, United Kingdom

Contribution no. 174, Geophysical Institute, Karlsruhe

0340-062X/79/0046/0369/$03.20

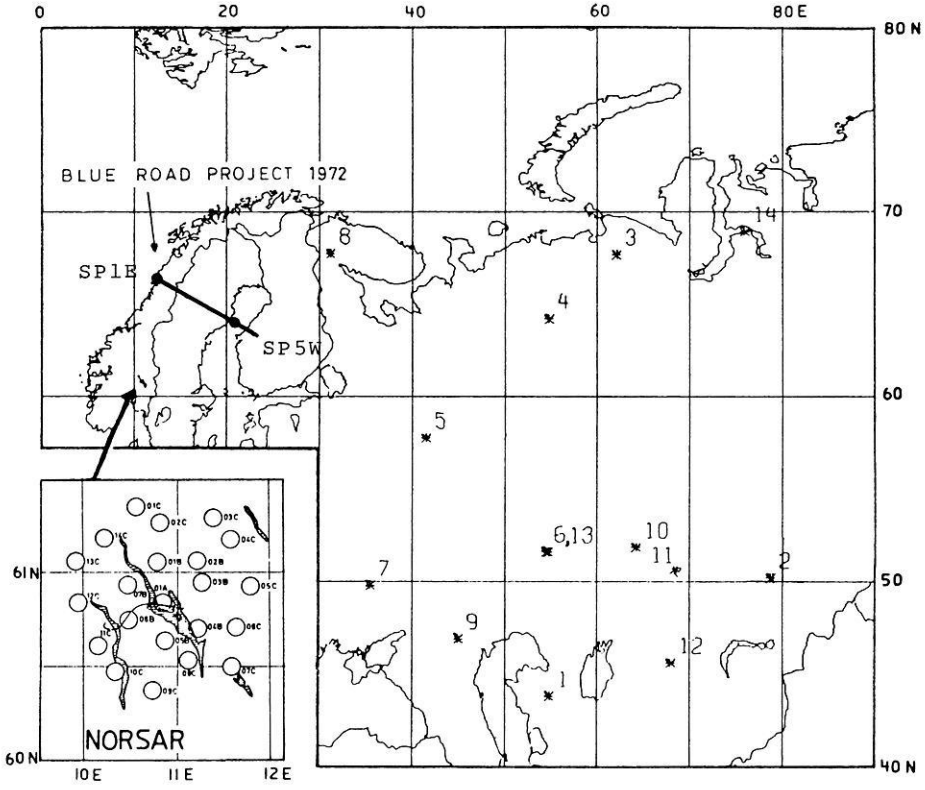


Fig. 1. Locations of seismic events in the USSR recorded at NORSAR (after King and Calcagnile 1976). *Top left*: BLUE ROAD traverse. *Inset*: NORSAR array location and subarray configuration

Introduction

Existing models of the postglacial uplift of the Baltic Shield incorporate an elastic layer superimposed over either a viscous half space (Cathles 1975) or a thick viscous asthenosphere (Lliboutry 1971; Bott 1971). In order to obtain a more thorough understanding of the dynamics responsible for the uplift a more accurate description of the structure and physical properties of the lithosphere and asthenosphere is required.

The seismic record sections used in the first part of this study were from the 'BLUE ROAD PROJECT 1972', comprising an explosion seismic profile which runs from the Norwegian coast near Mo i Rana over the Caledonides and across the Gulf of Bothnia between Umeå and Vasa before continuing for another 100 km into Finland (Hirschleber et al. 1975; Vogel 1976) (Fig. 1).

Five shots were located so as to make possible a detailed survey of the crustal structure along the profile (Hirschleber et al. 1975). The two longest record sections, SP 1 E (600 km) and SP 5 W (420 km) allow reversed coverage of the subcrustal lithosphere to a depth of about 100 km and both contain

clear first arrivals from the Moho. The data which has been filtered by a band pass of 1–20 Hz supplies ample evidence for a number of later arrivals.

The second part of this study constitutes a detailed examination of data recorded at NORSAR in the years 1970 to 1974. These recordings were provided by 14 nuclear explosions in the USSR within a distance range of $11° \lesssim \Delta \lesssim 40°$.

Taking observational gaps into account, King and Calcagnile (1976) constructed ar record section up to a distance of nearly 4,000 km. This was done by constructing profiles over the entire NORSAR area for each of the 14 events in various parts of the USSR (Fig. 1). These profiles cover a distance range between 1,250 and 4,295 km which enabled King and Calcagnile (1976) to calculate a *P*-velocity model to a depth of approximately 1,000 km. The resulting model, however, contains no low velocity zones. This fact is in contrast to existing models of the postglacial uplift of Fennoscandia, comprising viscous flow in the upper mantle which should indicate a region of low *P*-velocity.

For a more detailed description of the preparation of data used in both studies the reader is referred to Hirschleber et al. (1975) and King and Calcagnile (1976).

Interpretation of the Blue Road Data

For the interpretation the crustal structure derived by Hirschleber et al. (1975) was used. The crustal models consist of two homogeneous layers with a velocity contrast of 6.18 to 6.7 km/s at a depth of 18.6 km for SP 1 E and a velocity contrast of 6.08 to 6.5 km/s at a depth of 12.8 km for SP 5 W. Sellevoll (1973) has shown such a crustal model consisting of a granitic and a basaltic layer separated by the Conrad discontinuity to be a good first approximation for Scandinavia.

As can be seen in Figs. 2–4, first arrivals from 250 km onwards seem to undulate in a time interval of approximately 0.4 s. Using these arrivals Hirschleber et al. have applied a wave-front method (Meissner 1965) from which laterally varying Moho depths have been derived.

In order to show the effect of such an undulating Moho (P_n-velocity 8.1 km/s) on later arrivals, travel-times have been calculated using a ray tracing computer program allowing for the propagation of body waves in laterally inhomogeneous media with curved interfaces (Červený et al. 1974). During the course of this study several additions have been made to the program in order to enhance its efficiency (Cassell 1978). Not only are the undulating first arrivals reproduced but so are 'shadow zones' in the distance interval of 350–400 km (Fig. 2).

Due to the relatively constant Moho depths from a distance of 420 km onwards the undulating effect on arrivals later than *P*1 for SP 1 E is negligible, as is the effect on all later arrivals for SP 5 W. The depths of the subcrustal layers corresponding to *P*1 and *P*2 in Fig. 2 have been derived by means of a computer program developed by G. Müller which calculates travel-times for laterally homogeneous models (Figs. 3 and 4). In order to account for the non-linear first arrivals for SP 1 E and SP 5 W averaging P_n travel-time branches have been drawn through them. Although the amplitudes in the seismogram

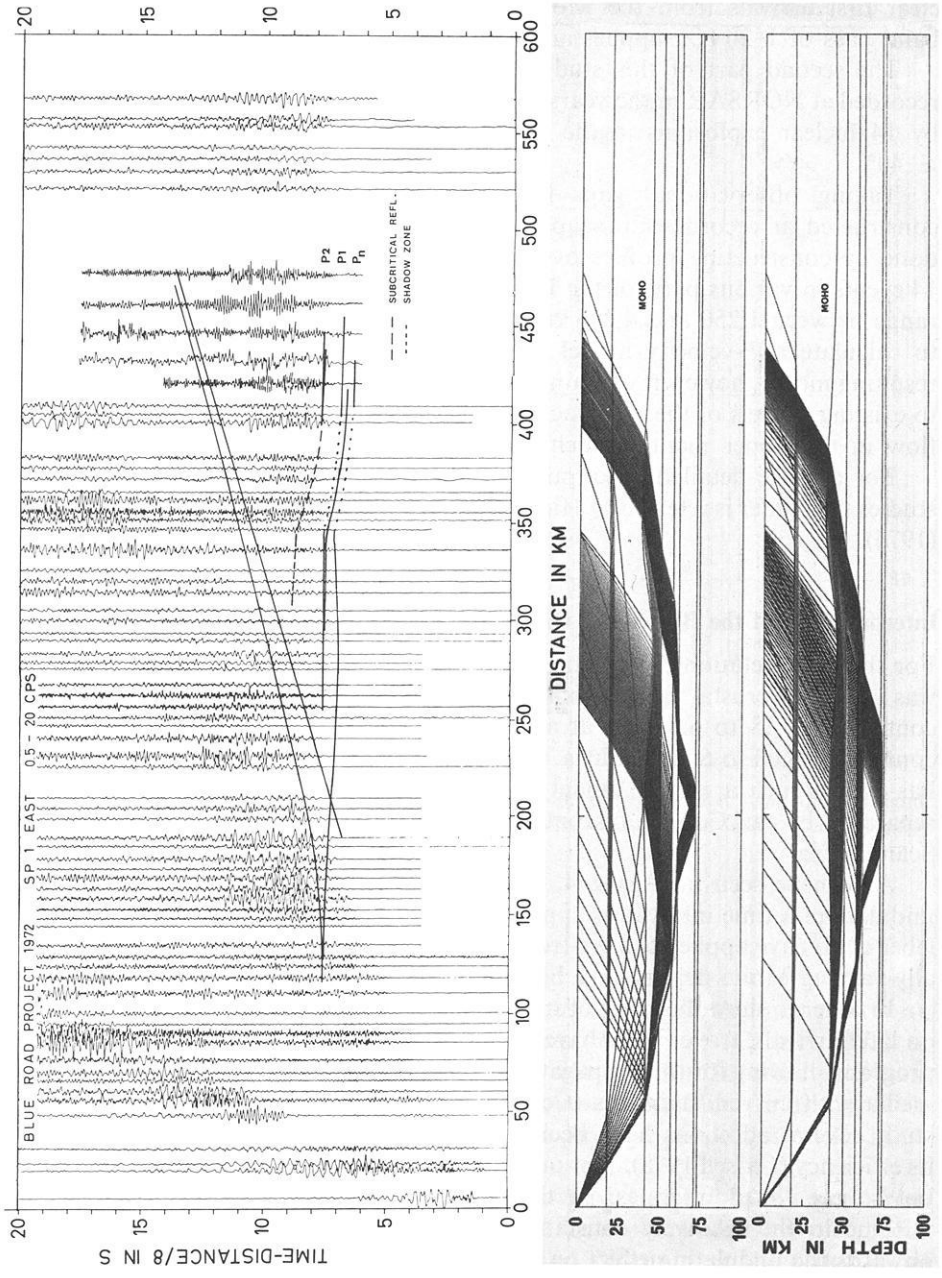


Fig. 2. Ray diagrams and travel-time curves showing the effect of an undulating Moho on later arrivals for SP 1 E. *Centre*: Reflections. *Bottom*: Refractions. The Moho relief was taken from Hirschleber et al. (1975) and the depth of the layers corresponding to *P*1 and *P*2 have been determined by calculating for a laterally homogeneous Earth. Depths are modified by allowing for a flatearth-transformation. The travel-time branches do not exceed 450 km due to the numberical method used

sections are not normalised, the relatively large-amplitude Moho reflections which should be near the inner cusp (Fig. 3) (Červený and Ravindra 1971) suggest the introduction of a 6-km-thick transition zone for the Moho at DP 1 E.

The large number of later arrivals for SP 1 E, especially in the recordings of the Gulf of Bothnia stations, suggested the correlation of a maximum of five nearly parallel travel-time branches with average velocity differences of 0.04 km/s (Fig. 3). These have been numbered $P1$ to $P5$. The possibility of these being multiple reflections has been ruled out by the unsuccessful attempt to reconstruct them by ray tracing and they could not be interpreted as such using the crustal structure as given by Hirschleber et al. (1975).

Due to the low velocity contrast between the subcrustal arrivals it was necessary to introduce low velocity zones in order to move the critical points to the small distances required. Owing to the sufficient velocity contrast between P_n and $P1$, $P1$ would be the *only* travel-time branch which could possibly be explained by a velocity increase beneath the Moho at a depth of 66.0 km without an inversion zone. The small critical distances and relatively strong amplitudes of the other arrivals make it necessary to strengthen the velocity contrasts at the interfaces by including low velocity layers ($\sim$7.7 km/s) between them. In the case of $P1$ the travel-time data could not resolve the presence of a first order discontinuity or a low velocity zone. For this study, layer $P1$ was taken to incorporate the same characteristics for use in further calculations, i.e., a thin layer embedded in low velocity material (Fig. 5). The average thicknesses of the layers is 1.8 km, in any case at least as large as a wavelength at the depths in question. The transition zones underneath the layers were introduced to reduce the effect of multiples in the computation of synthetic seismograms. For SP 5 W (Fig. 4) it was only possible to detect two later arrivals, $P1$ and $P2$ (Fig. 5). As explained earlier on for SP 1 E, arrival $P1$ for SP 5 W could also be modeled by either using a velocity increase at a depth of 66.0 km without an inversion zone or a thin layer embedded in low velocity material. In this case the latter model has again been used for reasons of consistency. Ray tracing tests (Fig. 2) show that the lateral extensions of the reflecting layers need not exceed 160 km.

The depths of layers $P1$ and $P2$ for SP 5 W coincide roughly with those of SP 1 E. In order to lessen the degree of non-uniqueness due to the inversion of the data, amplitude ratios have been calculated. The amplitude ratios of observed data and computed synthetic seismograms (Reflectivity method by Fuchs and Müller 1971) have been compared in order to verify the agreement between observation and theory (Fig. 6). For the computation of the synthetic seismograms the dominant frequency of the source signal was approximately 6.8 Hz, consisting of a roughly sinusoidal waveform with two extrema. This frequency is the average frequency of wave groups in the observational data. Since the uncertainty of the characteristics of the true source signal must also be taken into account a source signal with four extrema was tried without achieving an improvement. The observational amplitudes were determined by measuring the maximal amplitudes of wave groups directly behind arrivals. The main problem however, was the determination of critical distances since the non-normalised amplitudes made it impossible to distinguish between subcri-

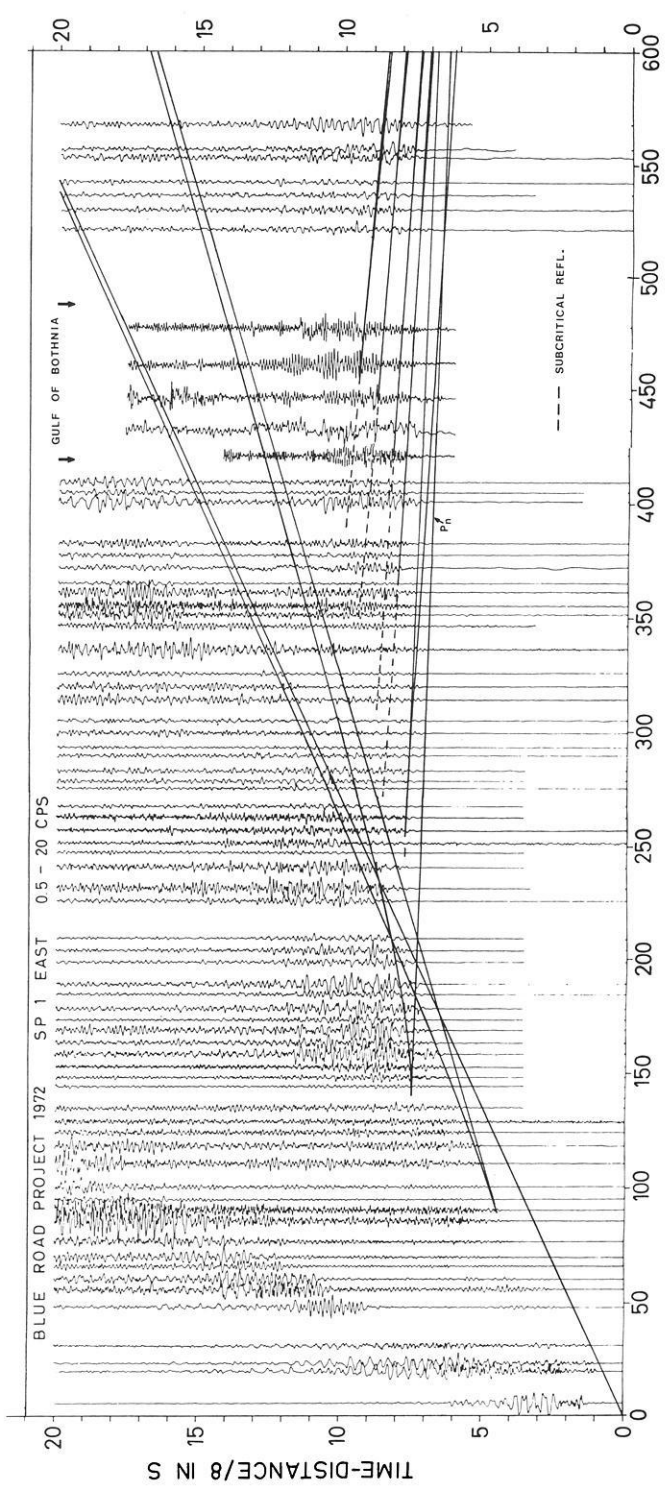


Fig. 3. Travel-time curves and record sections for SP 1 E. Arrivals later than P_n are labeled *P1* to *P5*. Crustal arrivals calculated after a model by Hirschleber et al. 1975. Moho and upper mantle arrivals correspond to the model presented in Fig. 5. P_n branch approximates undulating first arrivals

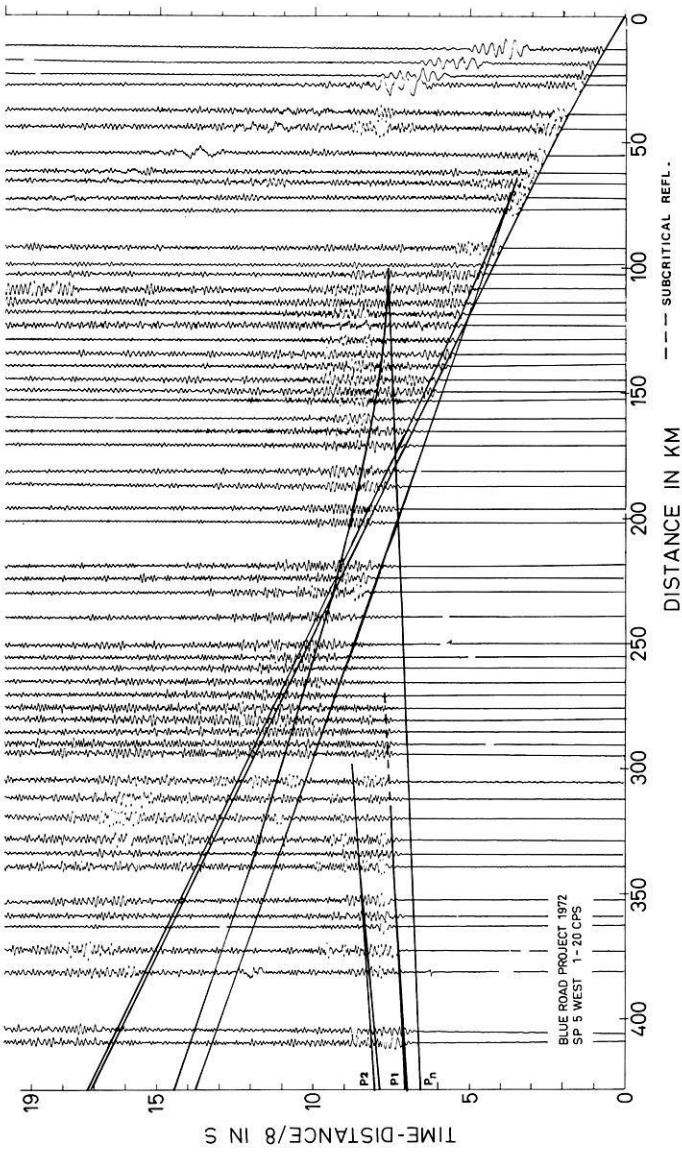


Fig. 4. Travel-time curves and record section for SP 5 W. Crustal arrivals calculated after a model by Hirschleber et al. 1975. Corresponding model presented in Fig. 5. P_n branch approximates undulating first arrivals

tical reflections and supercritical reflections. This was overcome by computing synthetic seismograms for small variations of the models until the best amplitude ratio correlations were found. To reduce the scatter of observed amplitudes, ratios of arrivals on the same seismogram were used. Figure 7 shows the final synthetic seismogram section and amplitude ratio fit for SP 5 W. Even with ratios the amount of scatter is considerable but the solid line follows the trend of the observations.

We have made no attempt to study the effect of long-wavelength heterogeneities in the Earth's crust on amplitude data. The short-wavelength scatter is certainly due to near-surface heterogeneities and may also reflect heterogeneities

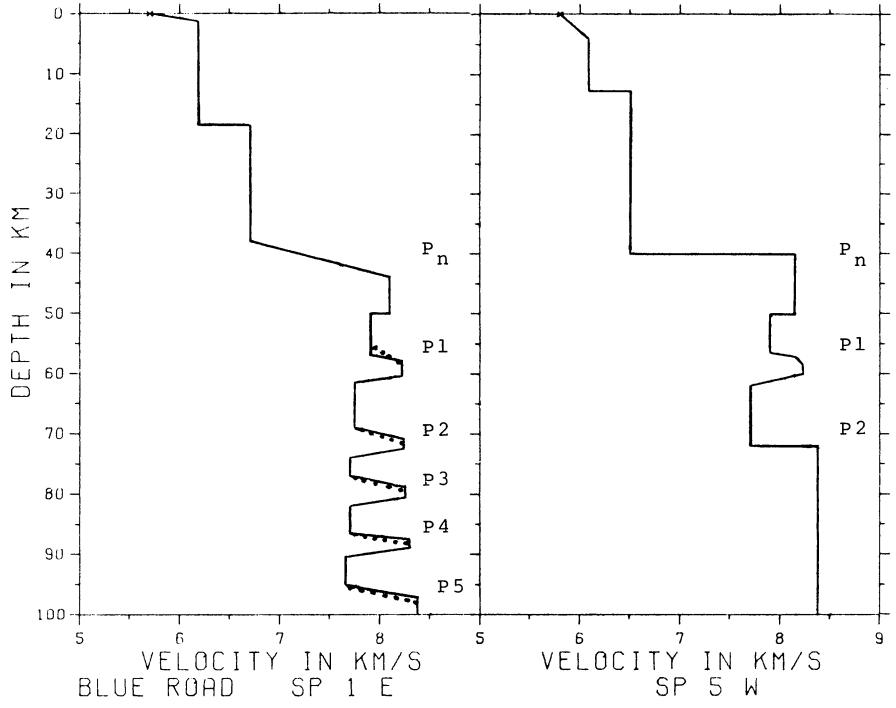


Fig. 5. Proposed velocity-depth models for SP 1 E and SP 5 W. Notations at the high velocity layers correspond to those of the travel-time branches in Figs. 3 and 4. *Dotted lines* in the left diagram specify deviations of the model which produce the dotted lines in Fig. 8. The Moho was found to consist of a 6-km-thick transition zone for SP 1 E

within the subcrustal lithosphere. Velocity-depth models are to be seen in Fig. 5. In Fig. 8 three possibilities for theoretical amplitude ratios for SP 1 E are shown. The deviations were achieved by varying the velocity gradients at the tops of the layers. The solid line in Fig. 8 is taken to be the best fit although discrepancies still remain which could not be resolved by calculating for a laterally homogeneous Earth. The corresponding model variation for the dotted line in Fig. 8 is to be seen as a dotted line in model SP 1 E in Fig. 5. The dashed line in Fig. 8 is produced by an assumed decrease of the velocity gradient on top of layer $P1$, not shown in Fig. 5. Since it was only possible to construct models by using methods allowing for laterally homogeneous structures a better fit for the amplitude ratios was not to be obtained.

At this point it should be emphasized that the detailed subcrustal structures produced by the correlations of the travel-time curves, especially by $P3$ to $P5$ for SP 1 E are in any case not the only possibilities but have been taken to be the most likely as evidence pointing to a complex subcrustal structure. Small variations of the correlation would not distort the main features of the model. A vertically homogeneous model of the upper mantle as suggested by Hirschleber et al. (1975) is not compatible with the observations. By using BLUE ROAD data, Lund (1979) has derived similarly structured P-velocity models

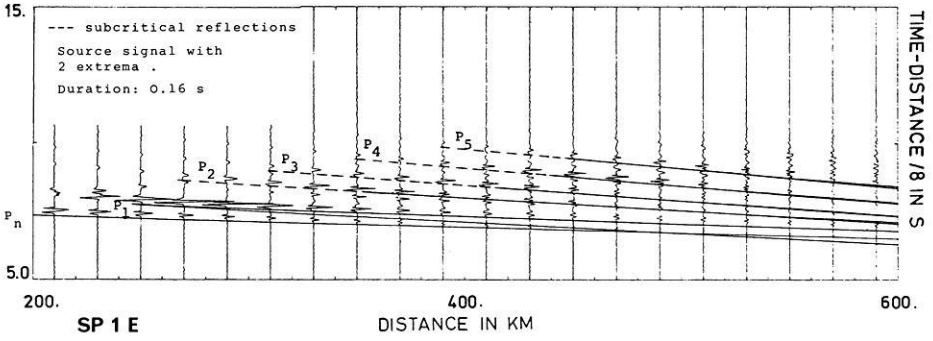


Fig. 6. Synthetic record section for SP 1 E. Travel-time braches calculated after model in Fig. 5. The dominant frequency of the source signal is approximately 6.8 Hz

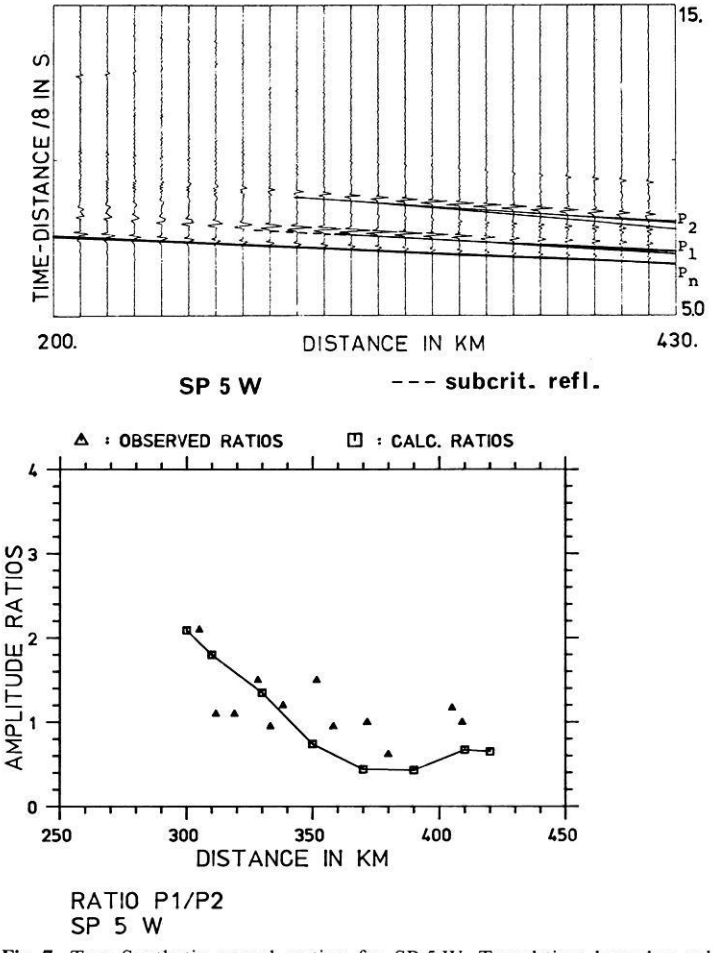


Fig. 7. *Top*: Synthetic record section for SP 5 W. Travel-time branches calculated after model in Fig. 5. The dominant frequency of the source signal is approximately 6.8 Hz and has two extrema. *Bottom*: Comparison of observed and theoretical amplitude ratios for SP 5 W

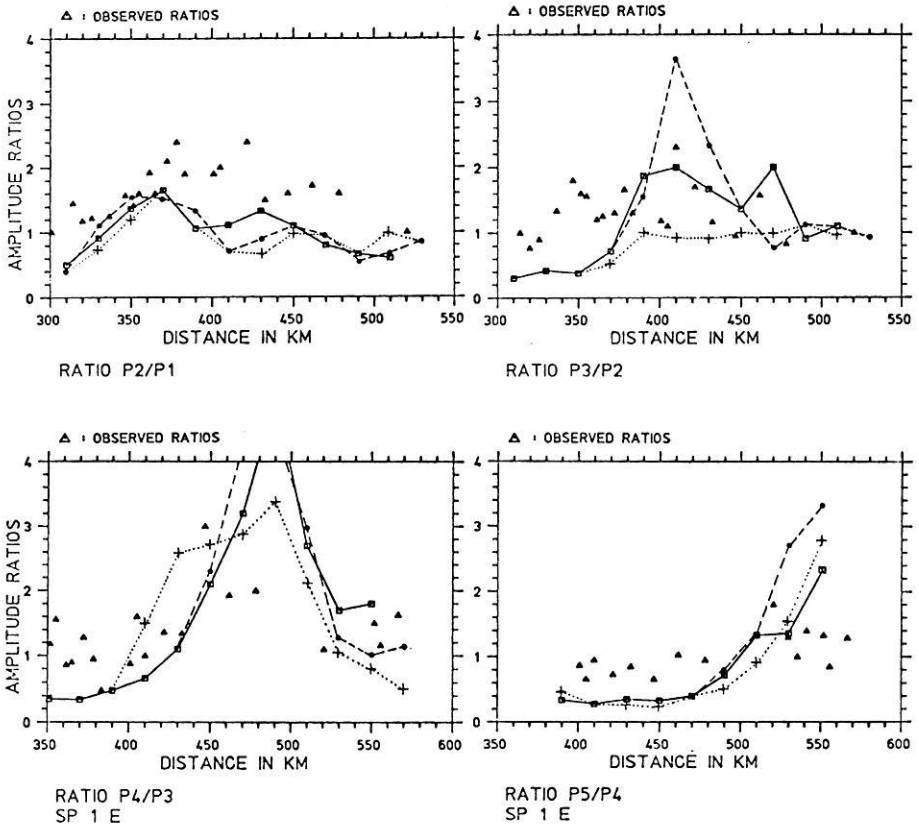


Fig. 8. Comparison of observed and theoretical amplitude ratios for SP 1 E. *Dotted lines* correspond to the *dotted lines* in Fig. 5. The *dashed curve* is produced by decreasing the velocity gradient at the top of layer P1 in Fig.5

of the subcrustal lithosphere. These were supplemented by an analysis of *S*-waves.

Reinterpretation of the NORSAR Data

King and Calcagnile (1976) constructed record sections using NORSAR data from 14 selected nuclear explosions in western Russia. The record sections with lengths of 110 km, the diameter of the array, could be put together in such a fashion as to obtain a total record section in a distance range 1,000 km $\lesssim \Delta \lesssim$ 4,500 km (Fig. 9). The amplitude variability across the array could partially be due to the effect of varying Moho depths (Berteussen 1975) and three-dimensional velocity variations of the lithosphere (Aki et al. 1977).

By subtracting approximately 4 s from the travel-times for event nos. 3 and

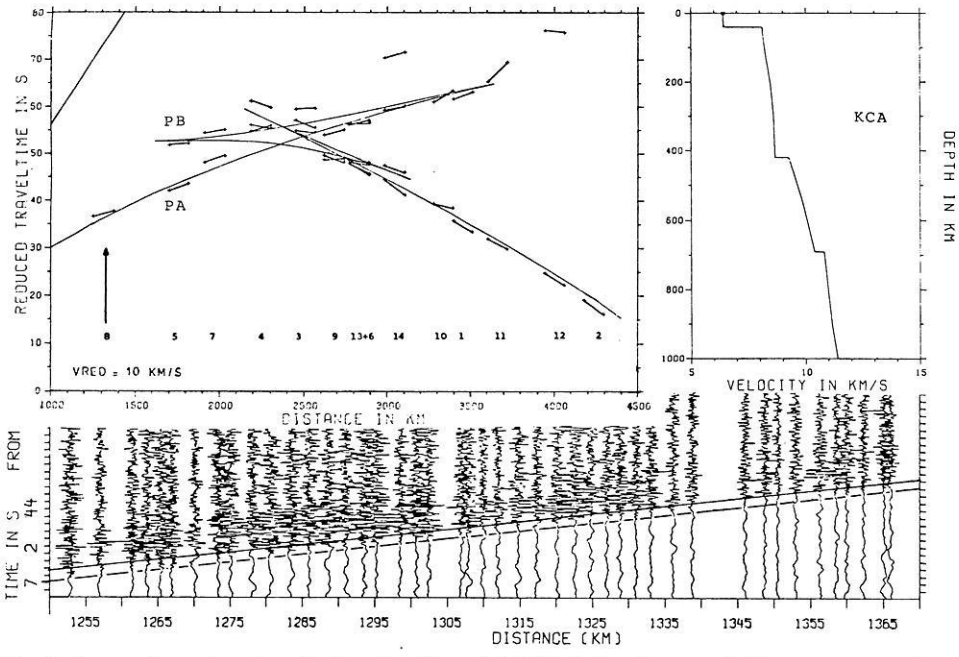


Fig. 9. *Bottom*: Record section of selected unfiltered NORSAR data for event 8. The strong secondary arrivals (*solid line*) were not considered by King and Calcagnile (1976). *Top right*: KCA model. *Top left*: Travel-time curves corresponding to the KCA model. *Short lines* represent arrivals for each numbered event. Event numbers as in Fig. 1. For a description of arrivals PA and PB see text

4 it was possible for King and Calcagnile to extrapolate continuous curves through the strongest arrivals in the data, these incorporating two (T, Δ) triplications (KCA model). The discrepancies in the correlation become evident by plotting the record sections at a reduced velocity of 10 km/s. These discrepancies are possibly due to varying precision in the determination of event locations and origin times. The azimuthal differences are probably also a contributing factor. The gaps which are incurred by the nature of the data make it difficult if not impossible to locate cusps with a reasonable degree of precision.

The KCA model consists of a relatively small positive velocity gradient below the Moho to 420 km depth followed by a first order discontinuity over a further positive velocity gradient leading to another first order discontinuity at 690 km depth.

Since there was no data available for the distance range necessary to draw conclusions on the nature of the Moho, an average depth of 40 km was chosen. Crustal effects were not considered in the calculations since their influence upon the deep structures of interest in this study is negligible. The data from event 8 (Fig. 9) suggests a strong secondary onset, arriving about 1–2 s after the first onset. This secondary arrival (solid line in Fig. 9, bottom), which has not been taken into account by King and Calcagnile gives rise to the

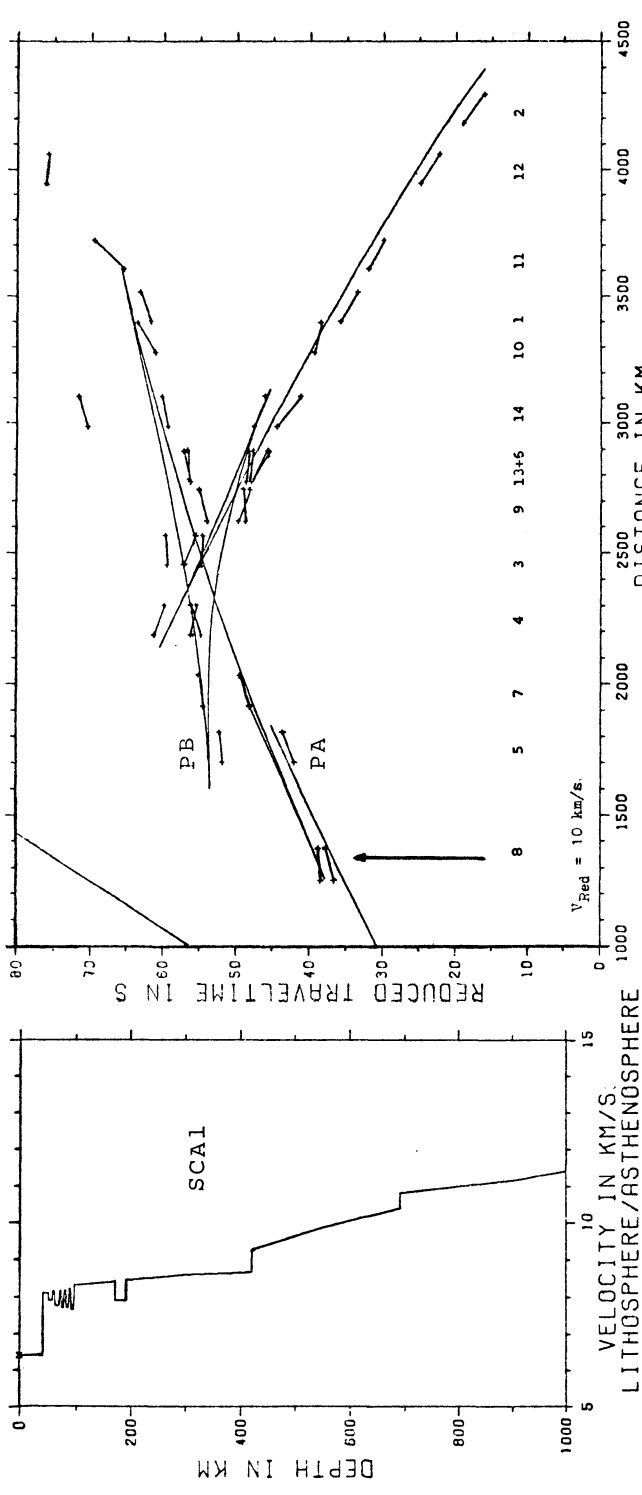


Fig. 10. *Left:* Modified version of the KCA model of the subcrustal lithosphere and upper mantle containing the BLUE ROAD results and a low velocity zone at 170 km depth. (SCA1). *Right:* Corresponding travel-time curve. *Short lines* as in Fig. 9. The secondary arrival for event 8 is to be seen in the bottom part of Fig. 9

Table 1. *P*-velocity model of the subcrustal lithosphere beneath Fennoscandia

z (km)	v(z) (km/s)	z (km)	v(z) (km/s)
0.0	6.40	86.5	7.70
40.0	6.40	87.5	8.30
40.0	8.10	89.0	8.30
50.0	8.10	90.5	7.65
50.0	7.90	95.0	7.65
57.0	7.90	97.0	8.32
58.0	8.22	100.0	8.32
60.5	8.22	170.0	8.40
61.5	7.75	170.0	7.90
69.0	7.75	190.0	7.90
71.0	8.24	190.0	8.45
72.5	8.24	200.0	8.45
74.0	7.70	200.0	8.59
77.0	7.70	420.0	8.66
78.8	8.26	420.0	9.27
80.5	8.26	550.0	9.88
82.0	7.70	690.0	10.38

construction of a low velocity zone between 170.0 and 190.0 km depth (Fig. 10). This is also in accordance with results published by Nolet (1977) who deduced a velocity reversal between 150 and 230 km by means of a Rayleigh wave dispersion analysis.

Although the travel-time curve produced by the relatively sharp structure of the low velocity zone satisfies the data in Fig. 9 its exact dimensions are not considered to be final.

The inferred low velocity zone might also explain the discontinuation of the observed first arrival branches for events 5 (PA) and 7 in the case that the origin times of these events have been determined with sufficient precision (Fig. 10).

After having calculated synthetic seismograms for the KCA model it was found that the amplitude ratios of first (PA) and secondary (PB) arrivals were not at all in agreement with those of the observational data for event 5.

In an attempt to reduce the amplitudes of the first arrivals, the BLUE ROAD model and the low velocity layer at 170 km depth were introduced to the KCA model which resulted in our final model SCA1 and which we believe to be representative for Fennoscandia according to the presently available data (Fig. 10, Table 1).

This resulted in an essential improvement of the amplitude ratios (Fig. 11). The incurred deviations from the KCA model travel-times are small as compared to the precision in the construction of the travel-time curves themselves.

The source signal for the synthetic calculations had a dominant frequency of 2.0 Hz.

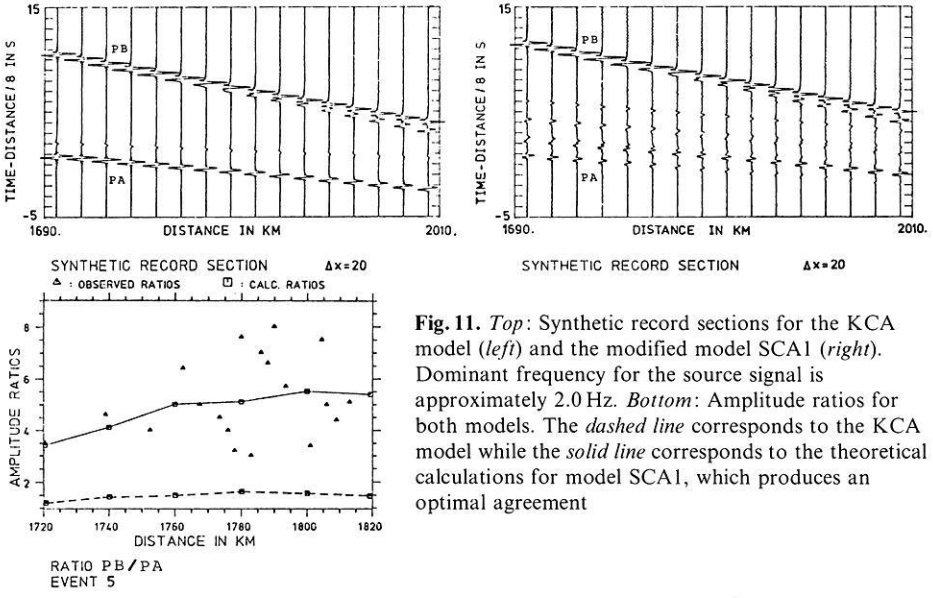


Fig. 11. *Top*: Synthetic record sections for the KCA model (*left*) and the modified model SCA1 (*right*). Dominant frequency for the source signal is approximately 2.0 Hz. *Bottom*: Amplitude ratios for both models. The *dashed line* corresponds to the KCA model while the *solid line* corresponds to the theoretical calculations for model SCA1, which produces an optimal agreement

Discussion

Other long range explosion seismic profiles have revealed the presence of relevant structural features in the lower lithosphere beneath Fennoscandia.

Massé and Alexander (1974) compared the upper mantle structure of the Canadian Shield with results of an interpretation of Russian nuclear explosion data recorded at NORSAR; in part the same data as later used by King and Calcagnile (1976). The Massé and Alexander model (MA) exhibits similar features to the KCA model, although the former includes a low velocity zone beneath the Moho.

Kulhánek and Brown (1974), using earthquake and explosion data recorded by the Uppsala array, proposed velocity depth models, including one with a low velocity zone between 50 and 150 km depth.

Sacks et al (1977), by examining SP conversions of incident teleseismic *S*-waves, found a velocity reversal at the lithosphere-asthenosphere boundary at a depth of 250 km. The data used in the present study provided no evidence to confirm their results.

The velocity-depth models shown in Fig. 5 give rise to a subcrustal lithosphere-asthenosphere consisting of alternating regions of low and high *P*-velocity.

This study did not take into account *S*-wave data and it is quite likely that the *S*-velocity-depth model would differ in detail. However, the complexities of the *P*-velocity structure require that the simple homogeneous model of a subcrustal lithosphere and asthenosphere as found by surface wave analysis must be modified for geodynamical calculations.

Using ray tracing methods (Cassell 1978) it has been shown that the minimum lateral extensions of the reflectors $P1$ and $P2$ are approximately 100–160 km which leads to the conclusion that they must not necessarily exceed that amount in order to produce the corresponding travel-time curves. A similar model of the lower lithosphere has been presented by Fuchs and Schulz (1976) who have investigated the penetration of tunnel waves through thin high-velocity layers.

Oxburgh and Parmentier (1978) propose a process allowing for the generation of continental lithosphere by diapiric accretion of low-density, depleted mantle bodies. This could lead to a plausible explanation of the alternating subcrustal structure suggested in this study.

Fuchs (1977) argues that shear stresses may create preferred orientations of olivine crystals which would cause maximum velocity orientation.

Conclusions

Beneath the Moho (~ 8.15 km/s) which was found to consist of a 6-km-thick transition zone under the Caledonides and a first order discontinuity to the east, the derived velocity-depth model pertaining to the BLUE ROAD data shows the subcrustal structure beneath Scandinavia to consist of a series of alternating high and low velocity layers down to a depth of 100 km.

The presence of critical reflections from the high velocity layers, which are to be observed at near distances, requires the insertion of low velocity zones of 7.9 to 7.65 km/s. Throughout the high velocity layers the P-wave velocity increases with depth within the range of 8.22 to 8.38 km/s.

The reinterpretation of the NORSAR long-range data infers the existence of an additional low velocity zone (~ 7.9 km/s) between 170.0 and 190.0 km depth.

Furthermore, the new model SCA1 of the upper mantle in Fennoscandia assures an optimal agreement between the observed and theoretical travel-times and amplitude ratios of first and secondary arrivals. Unfortunately the data were insufficient to provide information on the velocity-depth configuration in the depth range of 100 km to 170 km. This raises the question as to whether this depth range contains a continuation of the pattern of the above mentioned alternating layers.

With the presently available data it has not been possible to detect a sharp lithosphere-asthenosphere boundary. A homogeneous model does not agree with the observations. The main conclusion, however, is that the subcrustal lithosphere forms a region of clear vertical and possibly also horizontal variations characterised by high and low velocity material.

Acknowledgements. The authors wish to thank Professor Dr. R. Meissner, Dr. H.B. Hirschleber, and Dr. C.E. Lund for providing us with the data from the BLUE ROAD PROJECT 1972.

We are also grateful to Dr. E. Husebye for the NORSAR data. Computations were carried out on the UNIVAC 1108 of the University of Karlsruhe and on the RAYTHEON 500 of the Geophysical Institute.

References

Aki, K.; Christoffersson, A.; Husebye, E.S.: Determination of the three-dimensional seismic structure of the lithosphere. J. Geophys. Res. **82**, 277–296, 1977

Berteussen, K.A.: Crustal structure and P-wave travel-time anomalies at NORSAR. J. Geophys. **41**, 71–84, 1975

Bott, M.H.P.: The Interior of the Earth. pp. 235–247. London: E. Arnold Ltd. 1971

Cassell, B.R.: Explosionsseismische Untersuchungen der subkrustalen Lithosphäre unter Skandinavien. Diploma Thesis, Univ. Karlsruhe 1978

Cathles, L.M.: The viscosity of the Earth's mantle. 386 pp. Princeton, New Jersey: Princeton University Press 1975

Červený, V.; Langer, J.; Pšenčik, I.: Computation of geometric spreading of seismic body waves in laterally inhomogeneous media with curved interfaces. Geophys. J.R. Astron. Soc. **38**, 9–19, 1974

Červený, V.; Ravindra, R.: Theory of seismic head waves. Toronto: University of Toronto Press 1971

Fuchs, K.: Seismic anisotropy of the subcrustal lithosphere as evidence for dynamical processes in the upper mantle. Geophys. J.R. Astron. Soc. **49**, 167–179, 1977

Fuchs, K.; Müller, G.: Computation of synthetic seismograms with the reflectivity method and comparison with observations. Geophys. J.R. Astron. Soc. **23**, 417–433, 1971

Fuchs, K.; Schulz, K.: Tunneling of low-frequency waves through the subcrustal lithosphere. J. Geophys. **42**, 175–190, 1976

Hirschleber, H.B.; Lund, C.E.; Meissner, R.; Vogel, A.; Weinrebe, W.: Seismic investigations along the Scandinavian 'Blue Road' traverse. J. Geophys. **41**, 135–148, 1975

King, D.W.; Calcagnile, G.: *P*-wave velocities in the upper mantle beneath Fennoscandia and western Russia. Geophys. J.R. Astron. Soc. **46**, 407–432, 1976

Kulhánek, O.; Brown, R.J.: *P*-wave velocity anomalies in the Earth's mantle from the Uppsala array observations. PAGEOPH **112**, 597–617, 1974

Lliboutry, L.A.: Rheological properties of the asthenosphere from Fennoscandian data. J. Geophys. Res. **76**, 1433–1446, 1971

Lund, C.E.: The fine structure of the lower lithosphere underneath the Blue Road profile in northern Scandinavia. Tectonophysics **56**, 111–122, 1979

Massé, R.P.; Alexander, S.S.: Compressional velocity distribution beneath Scandinavia and western Russia. Geophys. J.R. Astron. Soc. **39**, 587–602, 1974

Meissner, R.: Wellenfrontverfahren für die Refraktionsseismik. Berichte des Institutes für Meteorologie und Geophysik der Universität Frankfurt/Main. Nr. 9. Im Eigenverlag des Institutes 1965

Nolet, G.: Upper mantle under western Europe inferred from the dispersion of Rayleigh waves. J. Geophys. **43**, 265–285, 1977

Oxburgh, E.R.; Parmentier, E.M.: Thermal processes in the formation of continental lithosphere. Philos. Trans. R. Soc. London, Ser. A: **288**, 415–429, 1978

Sacks, I.S.; Snoke, J.A.; Husebye, E.S.: Lithosphere thickness beneath the Baltic Shield. Carnegie Institution. Annual Report of the Director. Department of Terrestrial Magnetism. Carnegie Inst. Washington, Yearb. **76**, 805–813, 1977

Sellevoll, M.A.: Mohorovičić discontinuity beneath Fennoscandia and adjacent parts of the Norwegian Sea and the North Sea. Tectonophysics **20**, 359–366, 1973

Vogel, A.: The Blue Road Geotraverse: The aim of the interdisciplinary research. Vol. 98, pp. 251–254. Deep seismic sounding in 1972. Vol. 98, pp. 257–260. Geologiska Föreningens i Stockholm Förhandlingar 1976

Received June 5, 1979; Revised Version November 22, 1979

J. Geophys. 46, 385–395, 1979

Journal of Geophysics

Northern Part of the Tonga Region: A Complicated Subduction Closure

V. Hanuš[1] and J. Vaněk[2]

[1] Institute of Geology and Geotechnics
[2] Geophysical Institute, Czechoslovak Academy of Sciences,
141 31 Praha 4-Spořilov, Boční, 1401, Czechoslovakia

Abstract. A detailed investigation of the morphology of the Wadati-Benioff zone based on the geometry of earthquake distribution allowed us to interpret the deep structure of the northern part of the Tonga region as a result of the complicated interplay of three different subduction systems (Tonga, Lau, Horne) and two deep-seated fracture zones (Niuafo'ou, Peggy). The Tonga system represents the active part of the recent subduction of the Pacific plate, and the Lau and Horne systems are considered to be buried paleoplates, activated by a deep collision with the Tonga system. The Niuafo'ou fracture zone forms the northern tectonic closure of the active Tonga subduction and is acting as a transform fault. The function of the Peggy fracture zone might be similar, forming the northern closure of the Lau paleosubduction. The active parts of both fracture zones reach a depth of more than 400 km.

Key words: Seismicity – Plate tectonics – Tonga island arc – Deep transform faults.

Introduction

The Tonga-Kermadec subduction seems to be one of the most impressive phenomena that substantially influenced the development of the plate concept of recent tectonic evolution of the Earth (Isacks et al., 1968). Although studied by various seismological methods in considerable detail (Sykes, 1966; Oliver and Isacks, 1967; Mitronovas et al., 1969; Sykes et al., 1969; Mitronovas and Isacks, 1971; Barazangi et al., 1972), the structure of this subduction zone appears to be much more complicated than these investigators have proposed, especially in some areas of high intermediate and deep seismic activity (Hanuš and Vaněk, 1978a and b). In this connection, the northern part of the Tonga region, with an interplay of three different subduction systems and two deep-seated fracture zones, can be considered as the most complicated area of the

0340-062X/79/0046/0385/$02.20

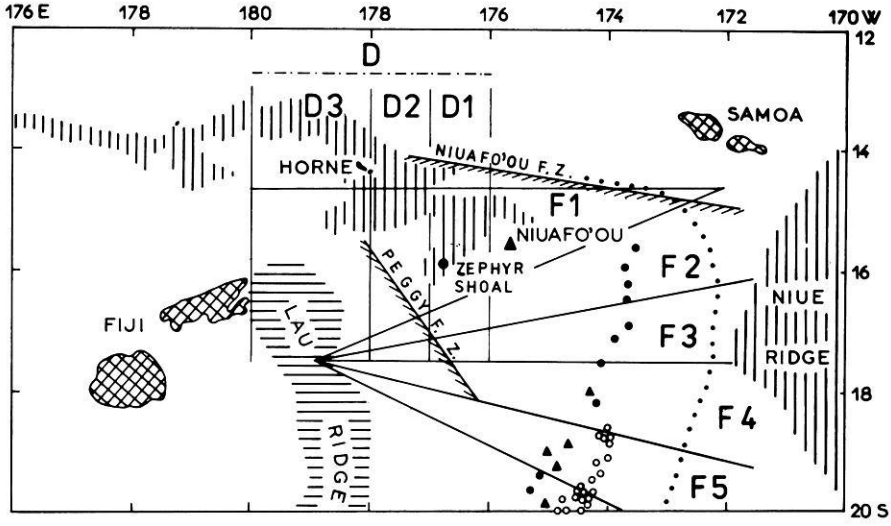


Fig. 1. Main structural elements and scheme of sections used for the study of the northern closure of the Tonga subduction zone. The Tonga-Kermadec trench is denoted by a *dotted line*, active volcanoes by *full triangles*, islands and submarine elevations associated with active volcanic chain by *full circles*, coral islands and reefs with associated submarine elevations by *open circles*

Tonga-Kermadec island arc, and, at the same time, as the northern closure of the recent Tonga subduction. The aim of the present paper is to investigate the detailed morphology of the Wadati-Benioff zone and to interpret the geometry of earthquake distribution in the area limited by 14°–19° S and 172° W–180°.

General Description

ISC data (Regional Catalogue of Earthquakes) for the seven years' period 1967–1973 were used as basic materials for studying the morphology of the Wadati-Benioff zone in the Tonga region. The ISC determinations, being based on world-wide observations, are burdened by the smallest random and systematic errors from all seismological data available. It appears that the accuracy given in the ISC bulletins is realistic and fully sufficient for studying the geometrical distribution of earthquakes in the regions of convergent plate margins.

The Tonga region was covered by a system of 22 sections approximately perpendicular to the axis of the Tonga-Kermadec trench, the scheme of which can be found in Hanuš and Vaněk (1978a). An analysis of the complete set of vertical sections, showing the depth distribution of earthquake foci in relation to the distance from the trench, reveals that the complicated geometry of the earthquake distribution cannot be interpreted by applying an elementary model of subduction. Assuming that the subduction zones preserve their simple plate-

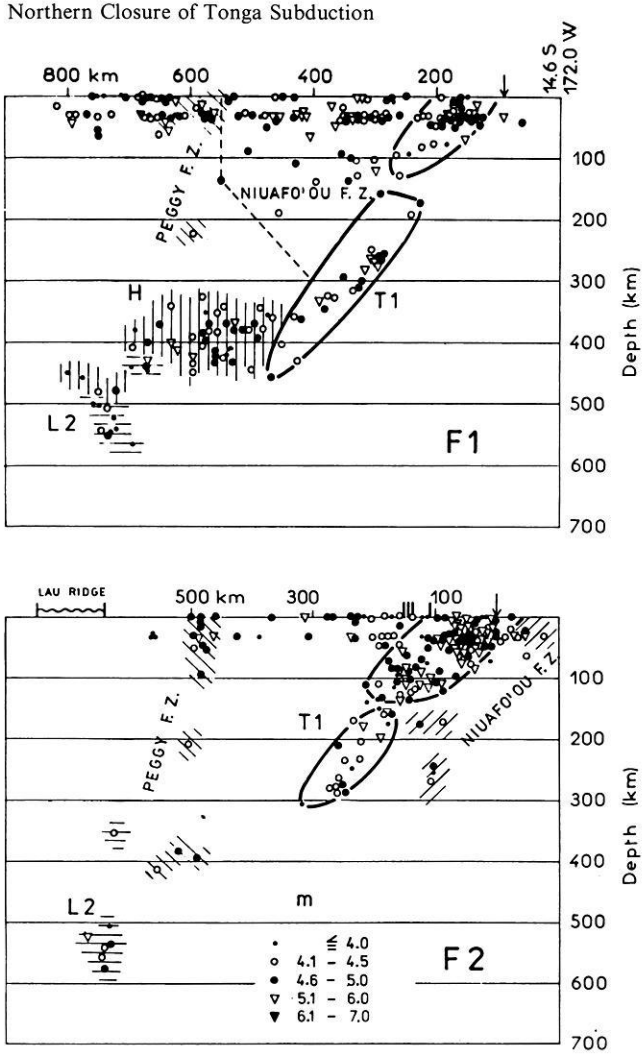


Fig. 2. Vertical Sects. *F*1 and *F*2 giving the distribution of earthquake foci in relation to the distance from the trench. For *F*1 the distance is plotted from the point 14.6° S, 172.0° W; m=ISC magnitude; active volcanoes are denoted by *full triangles*, islands and submarine elevations associated with active volcanic chain by *short vertical lines*, coral islands and reefs with associated submarine elevations by *short dotted lines*, position of the trench by *arrows*, the Lau Ridge by *wave-line*, the Wadati-Benioff zone *T* 1 by *full-line contour*, zones *L* 2, *H*, and fracture zones by *different hatching*

like shape the earthquake foci in the Tonga region can be associated with three interacting systems of subducted lithospheric plates: the recently subducting Tonga system *T* 1 and *T* 2 in the east, the remnant Lau system *L* 1 and *L* 2 in the west, and the remnant Horne system *H* in the north (Hanuš and Vaněk, 1978a).

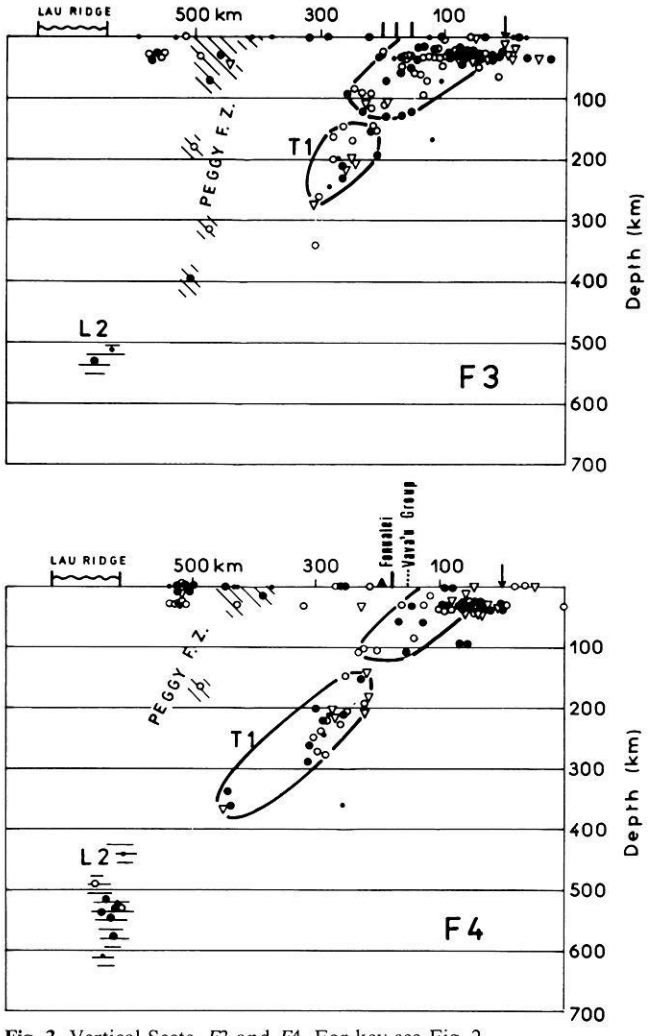


Fig. 3. Vertical Sects. *F*3 and *F*4. For key see Fig. 2

From the sequence of vertical sections, four northern sections *F*1–*F*4 of triangular shape are considered in the present paper. The position of these sections, together with the main structural elements of the northern part of the Tonga region, is given in Fig. 1, the depth distribution of earthquake foci being shown in Figs. 2 and 3.

It appears that the earthquake foci occurring in sections *F*1–*F*4 can be co-ordinated to the recently active Tonga zone *T* 1, to the Lau paleoplate *L* 2, and to the Horne system *H*. In addition, a group of shallow and intermediate earthquakes, westwards of the zone *T* 1, exists in section *F*1. Applying the principle of simple plate-like shape of lithospheric plates we conclude that these earthquakes cannot be geometrically associated with any of the above subduction

zones. The specific distribution of the latter foci indicates that they belong to two active deep-seated fracture zones, the positions of which are plotted in Fig. 1; we denote them as the Peggy and Niuafo'ou fracture zones, respectively. The earthquakes associated with the Peggy fracture zone can be also found in sections *F*2, *F*3, and *F*4. In the next paragraphs the above-mentioned units will be treated in detail.

Systems of Subduction

(a) Tonga System

The general picture shown by the sequence of vertical sections confirms the existence of a well-defined Wadati-Benioff zone *T* 1 beginning in the vicinity of the Tonga-Kermadec trench. This zone is divided by an intermediate aseismic gap into two distinct seismically active parts. The gap appears to be spatially connected with the occurrence of active andesitic volcanism and can be interpreted as a zone of partial melting in the subducted Pacific plate (for details see Hanuš and Vaněk, 1978a).

The depth penetration of the zone *T* 1 changes along the trench, reaching a depth of 455 km in the northern part of the Tonga region (Sect. *F*1). Then it substantially decreases due to the hampering effect of the Niue Ridge (Sects. *F*2, *F*3) with a subsequent southward increase (Sect. *F*4) to a maximum depth of 565 km (see Fig. 4 in Hanuš and Vaněk, 1978 a). Both the intermediate and deep seismic activity are abruptly terminated in the north by the Niuafo'ou fracture zone, which evidently forms the northern tectonic closure of the Tonga subduction (Isacks et al., 1969). The Wadati-Benioff zone *T* 1 represents, in our interpretation, the active part of the recent subduction of the Pacific plate along the Tonga-Kermadec trench.

(b) Lau System

In the western part of the Tonga region numerous deep earthquakes occur, which cannot be geometrically co-ordinated to the Wadati-Benioff zone of the Tonga subduction. The epicenters of these deep earthquakes are arranged in a meridional belt, the course of which substantially differs from that of the Tonga-Kermadec trench. These earthquakes were co-ordinated to the Lau Ridge, a remnant arc according to Milsom (1970) and Karig (1972), which runs in the same meridional direction and is situated exactly above the belt of deep earthquakes in question (Hanuš and Vaněk, 1978a and b).

Of two zones distinguished as *L* 1 and *L* 2, only earthquakes belonging to *L* 2 occur in the northern part of the Tonga region (sections *F*1–*F*4). The Lau system is interpreted as a pair of buried paleoplates activated by a deep collision with the recently downgoing Pacific plate of the Tonga system of subduction (for details see Hanuš and Vaněk, 1978b).

(c) Horne System

In the complicated seismic pattern of the vertical section *F*1 a horizontal strip of earthquakes with focal depths between 300 and 500 km can be observed westwards of the zone *T* 1 (see Fig. 2), which belongs neither to the zones *T* 1 and *L* 2, nor to any of the fracture zones occurring in the northern part of the Tonga region.

After projecting the foci of the above-mentioned strip of deep earthquakes in the proper direction, the system appears to form a plate-like body dipping to the south. This is demonstrated by the meridional vertical sections *D*1–*D*3 in Fig. 4. For comparison, the foci belonging to the Tonga zone *T* 1 are also plotted in Section *D*1, as well as those co-ordinated to the Peggy fracture zone and to the Lau zone *L* 2 in sections *D*2 and *D*3, respectively. The average dip of this zone, denoted as zone *H*, is about 40° to the south, its thickness ranges between 50 and 80 km, and the focal depth of earthquakes located in the zone varies between 325 and 480 km.

Zone *H* may be tentatively interpreted as a remnant subduction zone belonging perhaps to one of the manifestations of the supposed extinct Pacific-Phoenix spreading centre (Winterer, 1976). The activation of zone *H* is caused by a deep collision with the recently subducting Tonga zone *T* 1, as demonstrated by the vertical Sect. *D*1 in Fig. 4 and by the horizontal projection of zones *T* 1, *H*, and *L* 2 in Fig. 5. The clustering of earthquake foci in section *D*1 does not allow a detailed co-ordination of individual events to respective zones *T* 1 and *H*. The position of the collision is marked by the hatched area in Fig. 5, its centre of gravity being at 15.9° S and 176.4° W at a depth of about 400 km.

If the above interpretation is correct, some traces of the corresponding ancient island arc should be found in the physiographic pattern of the Pacific ocean between Samoa and the New Hebrides island arc. However, geologic and petrologic data on the islands of the area in question are very scarce. According to Macdonald (1945) the rocks of the Wallis Islands undoubtedly belong to the alkaline suite of the central Pacific volcanoes. The Horne Islands consisting of deeply weathered andesite lavas and breccias, and minor amounts of Miocene limestones, marls, sandstones, and conglomerates (Aubert de la Rüe, 1935) might be a part of the supposed ancient island arc, this assumption being not in contradiction to chemical analyses of Lacroix (1941). Also the bathymetry of the ocean floor shows a large elevation in the east-west direction, the exposed part of which is represented by the Horne Islands (World Map, 1968; see the scheme in Fig. 1). Therefore, the activated paleoplate observed in the northern part of the Tonga region is provisionally denoted as the Horne system *H*. In this context it is not surprising that Sclater et al. (1972) found andesitic rocks on Zephyr Shoal, which is situated in the south-eastern part of the supposed Horne island arc.

It is not excluded that the Horne system may continue to the northern part of the New Hebrides island arc, being responsible for isolated deep shocks in the Northern Fiji Plateau and the andesitic volcanism of the Tikopea Island (Fryer, 1974).

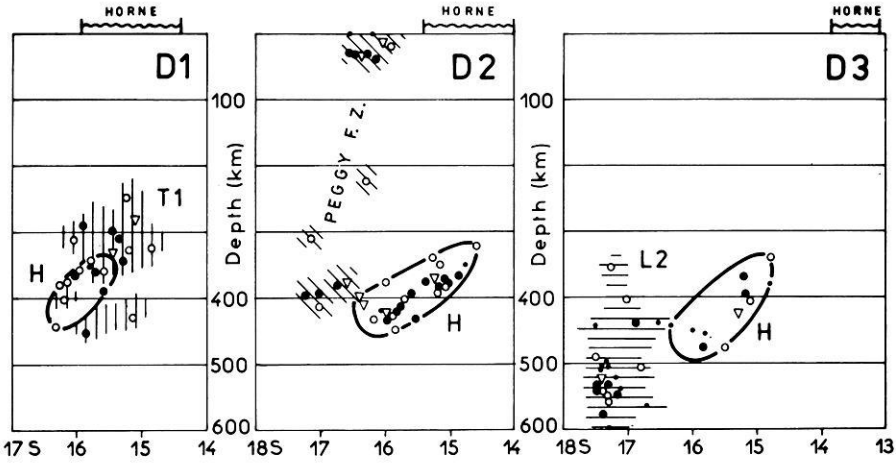


Fig. 4. Vertical Sects. *D*1–*D*3 for the Horne system, giving the distribution of earthquake foci in dependence on latitude; symbols as in Fig. 2. Zone *H* is denoted by *full-line contour*, zones *T* 1, *L* 2, and Peggy fracture zone by different hatching. For positions of Sections see Fig. 1

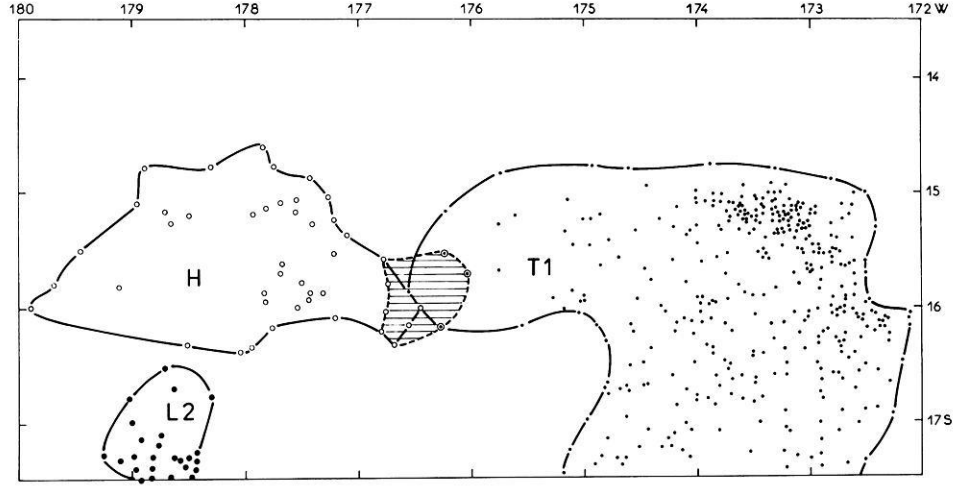


Fig. 5. Map of earthquake epicenters with delineation of three subduction systems in the northern part of the Tonga island arc. *T* 1: recent Tonga subduction; *L* 2: activated Lau paleosubduction; *H*: activated Horne paleosubduction; position of the deep collision is denoted by *horizontal hatching*

Deep Fracture Zones

In addition to the well-defined zones *T* 1, *L* 2, and *H*, two groups of shallow, intermediate and deep earthquakes can be observed westwards of the zone *T* 1 in the northern part of the Tonga region (see Figs. 2 and 3). In our interpretation, these earthquakes belong to two deep-seated fracture zones; they are denoted

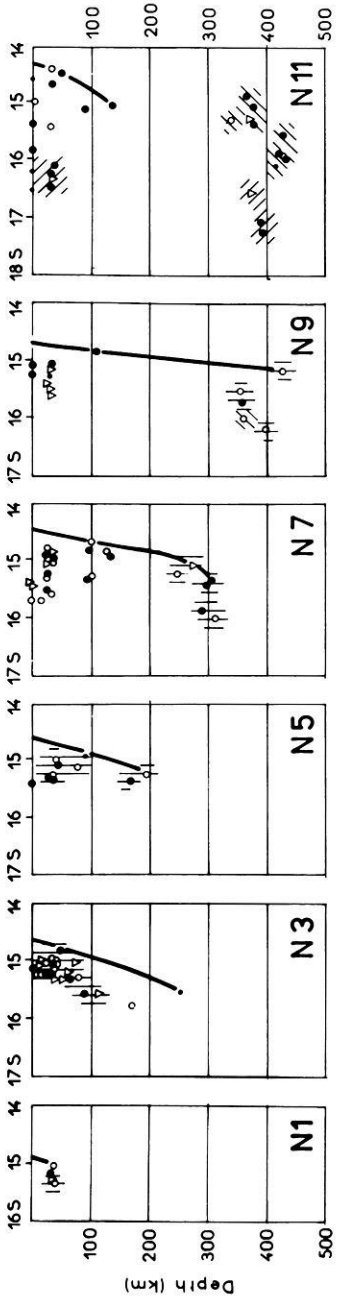


Fig. 6. Vertical Sects. *N*1–*N*12 across the Niuafo'ou fracture zone. Zone *T*1 is denoted by vertical hatching, zone *H* by NE-SW hatching, Peggy fracture zone by NW-SE hatching, and northern boundary of the Niuafo'ou fracture zone by a *full line*. For positions of Sections see Fig. 7

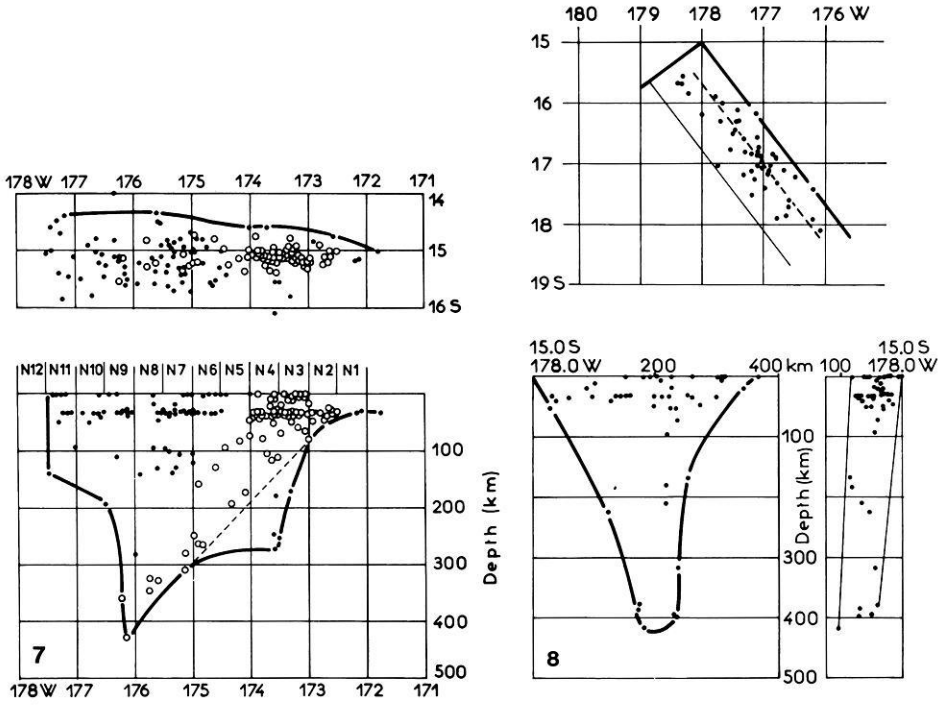


Fig. 7. Surface projection of the northern boundary and depth section along the Niuafo'ou fracture zone with associated earthquakes. Earthquakes belonging to both the fracture zone and subduction zone *T* 1 are denoted by *open circles*

Fig. 8. Surface projection, and vertical sections along and across the Peggy fracture zone, with associated earthquakes

as the Niuafo'ou and Peggy fracture zones, respectively, and their positions are shown in Fig. 1.

The *Niuafo'ou fracture zone,* the strike of which is almost in the east-west direction, evidently forms the northern tectonic closure of the active Tonga subduction zone *T* 1 and is playing the role of a transform fault (Isacks et al., 1969). In order to study the shape of the active part of the Niuafo'ou fracture zone, the associated activity is plotted in a sequence of meridional Sects. *N*1–*N*12 in Fig. 6 (for positions of sections see Fig. 7). After co-ordinating the individual shocks to appropriate zones, vertical sections *N*1–*N*12 give a clear picture of the northern boundary of the fracture zone. The southward dip ranges from 60° to 80°. The activity connected with the Niuafo'ou fracture zone reaches a depth of 430 km and disappears in Sect. *N*12. The surface projection of the northern boundary and the depth section along the fracture zone are shown in Fig. 7. It appears that in the eastern part the fracture zone cannot be distinguished from the subduction zone *T* 1, forming the northern closure of the downgoing Pacific plate. However, intermediate earthquakes in Sect. *N*3 and *N*4 indicate the occurrence of a mantle fracture with a northward dip under the zone of subduction. In the west, the active part of the Niuafo'ou fracture

zone is abruptly terminated by a surface steeply dipping eastwards to the deepest part of the northern Tonga subduction.

The Niuafo'ou fracture zone was named after the Niuafo'ou Island, an active volcano situated in the area of the fracture zone. It follows from the conventional silica alkali plot of existing chemical analyses of Lacroix (1941) and Macdonald (1948) that the Niuafo'ou basalts can hardly be correlated with the alkaline suite of the central Pacific volcanoes, found on the Wallis Islands (Macdonald, 1945), or with the andesitic suite of the Tonga island arc (Ewart et al., 1973). The chemical composition of the Niuafo'ou basalts, which is similar to that of the East Pacific Rise tholeiites (Engel et al., 1965), seems to show a genetic relationship with the fracture zone.

The *Peggy fracture zone*, the position of which coincides with the Peggy Ridge (see the bathymetric charts in Sclater et al., 1972, and in Hawkins, 1974), is another active deep-seated fracture zone found in the northern part of the Lau Basin. The earthquakes associated with the Peggy fracture zone can be distinguished in all the vertical Sects. *F*1–*F*4 (see Figs. 2 and 3). The surface projection of the active part and two vertical sections along and across the Peggy fracture zone are given in Fig. 8. The fracture zone dips steeply to the south-west at an angle of about 85° and its seismic activity reaches a depth of more than 400 km. The chemical composition of rock samples dredged from the Peggy Ridge shows a marked affinity between these basaltic rocks and ocean-ridge type tholeiites (Sclater et al., 1972), indicating a possible genetic relationship with the Peggy fracture zone, a phenomenon observed also for the Niuafo'ou fracture zone.

The function of the Peggy fracture zone is disputable, but it might be similar to that of the Niuafo'ou fracture zone, forming the northern closure of the Lau paleosubduction. In this case, the original Lau subduction zone would have been inclined to the east, with a later tilting of the buried Lau paleoplates to the present position caused by a deep collision with the subsequent Tonga subduction (Hanuš and Vaněk, 1978b). This mechanism could also explain the process of formation of the uplifted Lau Basin.

References

Aubert de la Rüe, E.M.: La constitution géologique des îles Wallis et Futuna. C.R. Acad. Sci. Paris **200**, 328–330, 1935

Barazangi, M., Isacks, B., Oliver, J.: Propagation of seismic waves through and beneath the lithosphere that descends under the Tonga island arc. J. Geophys. Res. **77**, 952–958, 1972

Engel, A.E.J., Engel, C.G., Havens, R.G.: Chemical characteristics of oceanic basalts and the upper mantle. Geol. Soc. Am. Bull. **76**, 719–734, 1965

Ewart, A., Bryan, W.B., Gill, J.B.: Mineralogy and geochemistry of the younger volcanic islands of Tonga, S.W. Pacific. J. Petrol. **14**, 429–465, 1973

Fryer, P.: Petrology of some volcanic rocks from the Northern Fiji Plateau. Geol. Soc. Am. Bull. **85**, 1717–1720, 1974

Hanuš, V., Vaněk, J.: Structure of the Wadati-Benioff zone in the Tonga region. Čas. Mineral. Geol. (Prague) **23**, 5–23, 1978a

Hanuš, V., Vaněk, J.: Tonga-Lau system: Deep collision of subducted lithospheric plates. J. Geophys. **44**, 473–480, 1978 b

Hawkins, J.W., Jr.: Geology of the Lau Basin, a marginal sea behind the Tonga arc. In: The Geology of Continental Margins: pp. 505–520. Berlin, Heidelberg, New York: Springer 1974
Isacks, B.L., Oliver, J., Sykes, L.R.: Seismology and new global tectonics. J. Geophys. Res. **73**, 5855–5899, 1968
Isacks, B., Sykes, L.R., Oliver, J.: Focal mechanisms of deep and shallow earthquakes in the Tonga-Kermadec region and the tectonics of island arcs. Geol. Soc. Am. Bull. **80**, 1443–1470, 1969
Karig, D.E.: Remnant arcs. Geol. Soc. Am. Bull. **83**, 1057–1068, 1972
Lacroix, A.: Composition minéralogique et chimique des laves des volcans des îles de l'Océan Pacifique. Acad. Sci. Paris Mém. **63**, 1–97, 1941
Macdonald, G.A.: Petrography of the Wallis Islands. Geol. Soc. Am. Bull. **56**, 861–872, 1945
Macdonald, G.A.: Notes on Niuafo'ou. Am. J. Sci. **246**, 65–77, 1948
Milsom, J.S.: The evolution of the Lau Ridge, Fiji Islands. Earth Planet. Sci. Lett. **8**, 258–260, 1970
Mitronovas, W., Isacks, B.L.: Seismic velocity anomalies in the upper mantle beneath the Tonga-Kermadec island arc. J. Geophys. Res. **76**, 7154–7180, 1971
Mitronovas, W., Isacks, B., Seeber, L.: Earthquake locations and seismic wave propagation in the upper 250 km of the Tonga island arc. Bull. Seismol. Soc. Am. **59**, 1115–1135, 1969
Oliver, J., Isacks, B.: Deep earthquake zones, anomalous structures in the upper mantle, and the lithosphere. J. Geophys. Res. **72**, 4259–4275, 1967
Regional Catalogue of Earthquakes, Vol. 4–10. Edinburgh: Int. Seismol. Centre, 1967–1973
Sclater, J.G., Hawkins, J.W., Mammerickx, J., Chase, C.G.: Crustal extension between the Tonga and Lau Ridges: petrologic and geophysical evidence. Geol. Soc. Am. Bull. **83**, 505–518, 1972
Sykes, L.R.: The seismicity and deep structure of island arcs. J. Geophys. Res. **71**, 2981–3006, 1966
Sykes, L.R., Isacks, B.L., Oliver, J.: Spatial distribution of deep and shallow earthquakes of small magnitudes in the Fiji-Tonga region. Bull. Seismol. Soc. Am. **59**, 1093–1113, 1969
Winterer, E.L.: Anomalies in the tectonic evolution of the Pacific. In: The geophysics of the Pacific Ocean Basin and its margin. AGU Geophys. Monograph **19**, 269–278, 1976
World Map (1:2500000): Samoa Islands, New Hebrides. Moscow: Main Admin. Geod. Cartogr. 1968

Received June 18, 1979; Revised Version August 3, 1979

J. Geophys. 46, 397–411, 1979

Journal of Geophysics

Palaeomagnetic and Rockmagnetic Properties of the Permian Volcanics in the Western Southern Alps*

C. Heiniger

Institut für Geophysik, ETH Hönggerberg, CH-8093 Zürich, Switzerland

Abstract. A palaeomagnetic and rockmagnetic investigation has been carried out in the western Southern Alps. At 31 sites in the Permian volcanics from four regions 349 samples were drilled: 1. the region of Arona (SW of Lago Maggiore), 2. the porphyry district of Lugano and Ganna, 3. the Valle Brembana (N of Bergamo) and 4. the Auccia volcanics (N of Brescia).

AF-cleaning as well as continuous and progressive thermal demagnetization reveal, in most of the igneous rocks studied, the presence of a stable characteristic remanent magnetization (ChRM) representing the magnetization acquired at the time of formation of these rocks. Microscopic observations, IRM-acquisition curves and Curie point measurements indicate the ChRM to be carried by Ti-poor magnetite and titanohematite. Stable secondary magnetizations due to oxidation may be present. Their directions, however, are very similar to the primary thermoremanent magnetization (TRM). Therefore it is inferred that the oxidation probably took place shortly after acquisition of the primary TRM.

The magnetization directions within individual sites are well grouped (α_{95} usually $< 10°$), but the site mean directions are dispersed, due to regional and local tectonic complications. At Arona, Ganna and Auccia a suitable tilt correction can be made. Since the consistent directions from these sites are very similar to the well defined results from the Bolzano porphyries (Zijderveld et al., 1970), it is suggested that the western and eastern Southern Alps have, on a large scale, behaved as a single tectonic unit. The Southern Alpine block has been rotated anticlockwise by about 50° relative to extra-alpine Europe since the Early Mesozoic.

The Permian Southern Alpine palaepoles are situated close to the Permian part of the African polar wander path. Therefore the palaeomagnetic data support geological and sedimentological arguments which consider the Southern Alps as originating on the southern margin of Tethys and forming a parautochthonous extension of the African plate since the Early Mesozoic.

* Institut für Geophysik, ETH Zürich, Contribution No. 250

0340-062X/79/0046/0397/$03.00

Key words: Rock magnetism – Palaeomagnetism – Southern Alps – Plate tectonics.

1. Introduction

For more than a decade the Alpine and Mediterranean area has been of much interest for palaeomagnetic studies. Earlier investigations in the Southern Alps – the region to the South of the Insubric Line (Fig. 1) – give palaeomagnetic evidence for tectonic rotations (e.g. Zijderveld and Van der Voo, 1973). Most of the palaeomagnetic data from the Southern Alps however, come from regions to the east of the Giudicaria fault and concentrate on the Bolzano quartz-porphyries. Many of these data are based on purely palaeomagnetic work with little or no examination of the magnetic mineralogy.

The aim of the present palaeomagnetic and rockmagnetic study was to obtain more evidence related to the large scale tectonic history of this area.

2. Geological Outline

The northern and western margins of the Southern Alps are sharply defined by the Insubric fault line. The Po and Venetian plains represent the southern boundary. The tectonic style, with folds, small 'nappes' and faults contrasts markedly with the 'nappe'-style characteristic of the Alps north of the Insubric Line. In terms of Alpine tectonic units they represent the 'autochthonous' hinterland of the Austro-Alpine nappes, from which they have been separated by the Late Alpine Insubric fault.

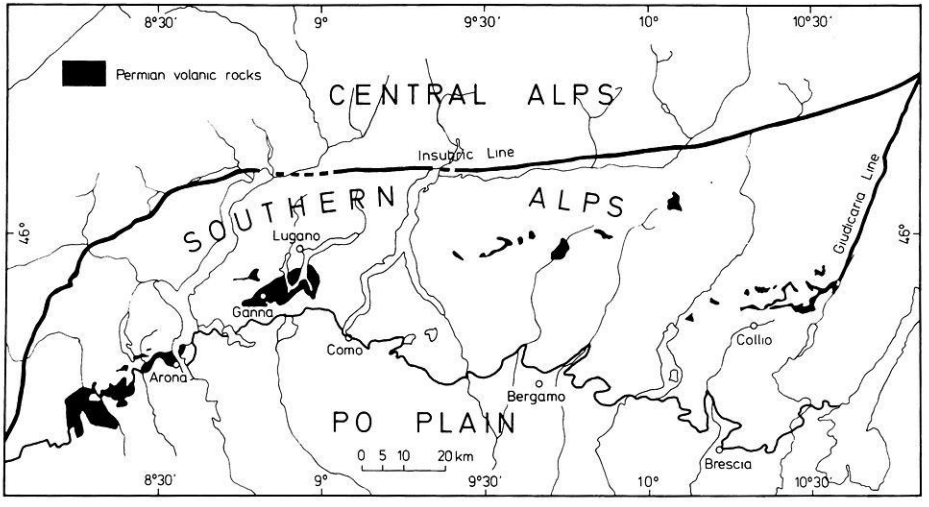


Fig. 1. Geographic situation of the four sampling regions and main outcrops of Permian volcanics in the western Southern Alps

The stratigraphic sequence of the Southern Alps comprises Hercynian and Pre-hercynian crystalline basement rocks, Palaeozoic sediments and volcanics and a highly differentiated Mesozoic-Tertiary sedimentary sequence.

The Permian volcanics and sediments, together with the Triassic formations, constitute the lower part of the sedimentary sequence. They rest unconformably on a basement that was subject to Hercynian orogenesis in Carboniferous time.

With the exception of the Bolzano quartz-porphyries (eastern Southern Alps) all the main outcrops of the Permian volcanics are situated to the west of the Giudicaria Line. The Bolzano quartz-porphyries have been thoroughly investigated geologically (cf. Pichler, 1959) as well as palaeomagnetically (Zijderveld et al., 1970), and their palaeomagnetic results will be compared with those from the western Southern Alps.

Geological data from the Permian igneous rocks from the western Southern Alps are sparse. There are no radiometric ages for the volcanics. However, the Baveno granite which is considered to be genetically equivalent to the Lugano volcanics, has been dated to 270 to 290 m.y. (Jaeger and Faul, 1959).

On the basis of stratigraphic evidence, the Bolzano quartz-porphyries can be considered to be of Permian age (Pichler, 1959). In the case of the Lugano porphyries and those from the region of Arona, the Lower Permian age follows from their stratigraphic position which is younger than the Westphalian of Manno (Graeter, 1952) and older than the Upper Permian 'Verrucano Alpino' of the Bergamasc Alps.

The volcanic sequencies are most completely preserved in the region of Lugano where they have been examined in detail by Kuenen (1925) and De Sitter (1939). The most complete petrological study of the Arona sampling area is given by Kaech (1903) and for the Bergamasc Alps by De Sitter and De Sitter-Koomans (1949). For the Auccia volcanics it is referred to the studies of Cassinis (1966; 1968).

Petrologically all these Permian igneous rocks show a large variety of porphyries, porphyrites, quartz-porphyries, and tuffs. In all sampling areas the rocks show varying degrees of alteration due to weathering. The primary mineralogical composition often cannot be reconstructed in detail. As a consequence unambiguous identification of the rock type is sometimes not possible.

The tectonic situation differs from sampling area to sampling area and sometimes even from site to site within each area. This deserves much attention, because only a reliable tectonic correction of the remanence directions allows for a palaeomagnetic interpretation of the results.

3. Natural Remanent Magnetization

Between 5 and 13 cores were drilled at each of 31 sites in the Permian volcanics (cf. Table 1). Each gave 1 to 5 cylindric specimens, 2.54 cm i.d. and 2.2 cm high. The resultant ratio of height to diameter of 0.85 gives an optimal reduction of sample shape anisotropy (Porath et al., 1966). Altogether 349 samples were investigated to establish the palaeomagnetic and rock magnetic properties of the Permian volcanics.

Table 1. Mean direction of ChRM before and after tectonic correction

Region Site	ChRM after demagnetization					Tectonic Reference		Tectonically corrected direction of ChRM			
	n	D	I	α_{95}	kF	Az	Dip	D	I	α_{95}	kF
Auccia											
AU1	14	133.2	−44.7	6.9	34.2	304–320	18–42	131.3	−26.8	6.9	34.3
AU2	25	138.3	−22.5	3.0	96.5	8	18	141.2	−12.1	3.0	96.5
AU3	9	142.9	−20.7	3.5	215.7	300	10	141.8	−11.5	3.5	215.7
AU4	13	140.4	−21.4	2.7	237.4	320	8	139.4	−12.0	2.7	337.4
AU5	13	142.7	−16.2	1.0	>1 000	4	12	144.0	− 8.2	1.0	>1 000
Mean	74	139.4	−25.1	2.6	39.9			139.8	−14.1	2.2	54.7
Regional mean (N=5):								139.7	−14.1	8.2	88.8
Valle Brembana											
BR2	12	209.4	−36.3	2.6	276.3	}		(193	−2	2.6	276.3)
BR3	8	197.0	−49.7	4.8	131.7	} (330	50)	(178	−9	4.8	131.7)
BR6	10	209.0	−38.5	11.5	15.4	}		(191	−4	11.5	15.4)
Mean	30	206.4	−40.4	4.7	32.4			(189.5	−5.4	4.7	32.4)
BR7	7	238.7	−26.8	17.3	13.2	(120–150	3	255	−21	17.3	13.2)
BR8	8	222.2	18.4	5.3	111.4	–	–	–	–	–	–
Lugano											
1	11	108.7	34.5	3.6	158.0	–	–	–	–	–	–
2a	3	125.1	16.3	7.9	247.3	–	–	–	–	–	–
2b	15	92.6	32.4	2.8	191.7	–	–	–	–	–	–
3a	4	125.9	22.4	10.3	80.4	–	–	–	–	–	–
3b	12	102.2	39.1	6.5	45.4	–	–	–	–	–	–
4	10	120.0	−36.4	14.1	16.4	–	–	–	–	–	–
5	56	106.6	21.7	3.8	25.8	–	–	–	–	–	–
6	6	131.5	−2.8	11.6	34.2	–	–	–	–	–	–
Ganna											
GA1	13	147.1	26.7	8.4	25.6	150	45	147	−22	8.4	25.6
GA2	11	151.6	5.2	4.6	99.2	150	30	151	−25	4.6	99.2
VHZ1	3	135.8	31.4	8.6	207.3	146	42	137	−11	8.6	207.3
VHZ2	4	135.0	30.3	7.0	171.7	154	56	137	−23	7.0	171.7
Mean	31	146.3	20.1	5.5	22.9			146.3	−20.5	4.3	36.5
Regional mean (N=4):								142.9	−20.4	10.4	78.5
GA3	10	189.9	−4.0	5.0	95.5	–	–	–	–	–	–
GA4	7	169.7	−4.0	5.8	109.8	–	–	–	–	–	–

Table 1 (continued) Mean direction of ChRM before and after tectonic correction

Region site	ChRM after demagnetization					Tectonic reference		Tectonically corrected direction of ChRM			
	n	D	I	α_{95}	kF	Az	Dip	D	I	α_{95}	DkF
Arona											
AR1	11	138.3	41.4	2.9	254.0	140	45	140	−5	2.9	254.0
AR3	7	146.9	22.0	6.7	83.0	145	35	146	−13	6.7	83.0
AR4	3	124.8	27.9	12.9	92.0	145	35	128	−16	12.9	92.0
BI21	11	124.1	29.8	10.4	20.4	130	52	125	−22	10.4	20.4
Mean	32	134.2	32.3	5.1	26.0			134.8	−11.1	5.8	19.9
Regional mean (N=4):								134.9	−14.2	13.6	46.5
AR2	3	116.7	58.9	15.9	60.8	(140	45	130	16	15.9	60.8)
BI23	8	102.4	22.1	12.1	22.1	–	–	–	–	–	–
BI24	22	106.9	10.9	4.1	41.6	(50	20	105	0	4.1	41.6)

Az, Dip: Azimuth and dip of tectonic reference plane used for correction
n: Number of specimens
N: Number of sites
α_{95}: Semi angle of Fisher's 95% cone of confidence (Fisher, 1953)
kF: Fisher precision parameter
Directions in brackets are not used for palaeomagnetic interpretation

The characteristic remanence (ChRM), acquired at the time of formation of the volcanics, has been evaluated by means of stepwise alternating field demagnetization as well as continuous and progressive thermal demagnetization. About 90% of the samples were AF-cleaned, the remainder was demagnetized by thermal methods.

The intensities of the initial NRM and of the ChRM are distributed over wide ranges. They were measured with a Digico complete results magnetometer (Molyneux, 1972) and with a ScT cryogenic magnetometer (Goree and Fuller 1976). The mean value of initial NRM-intensity of all samples investigated is $6.03*10^{-5}$ Gauss, the mean value after optimum demagnetization is $2.95*10^{-5}$ Gauss. On average the ChRM intensity represents about half the initial NRM intensity. No correlation between intensity-distribution and rock type or sampling area was found.

The ChRM is taken to be the most stable remanence component, as determined with the aid of a palaeomagnetic stability index obtained by computing the change of direction of the remanence vector for each demagnetization step. The minimum directional change is associated with the most stable direction and is tabulated in Fig. 2 for a representative sample together with plots of remanence intensity and a stereographic projection of the direction. For this example the characteristic remanence direction is found at AC-field amplitudes between 120 and 300 Oe where the smallest changes in direction occur. The optimum field is chosen at 200 Oersted because the index parameter PSI is minimum and amounts to only about 47 mdeg/Oe. Because the rock magnetic

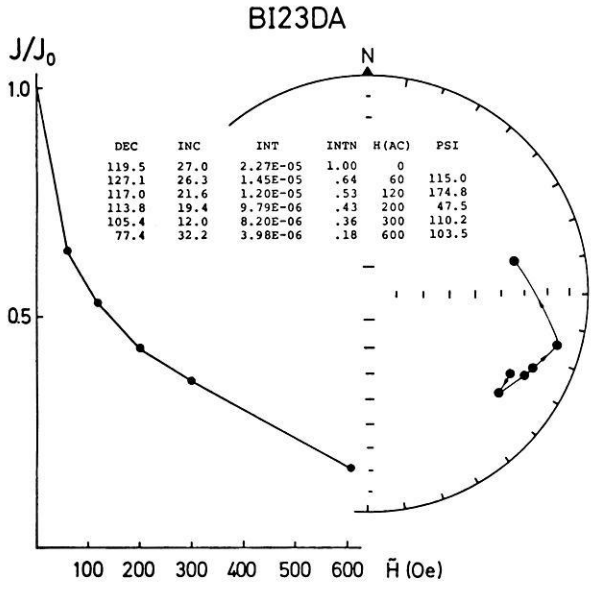


Fig. 2. Change of direction (PSI, in millidegrees per Oersted) and normalised intensity (INTN) of natural remanent magnetization (NRM) of a porphyry sample during AF-demagnetization

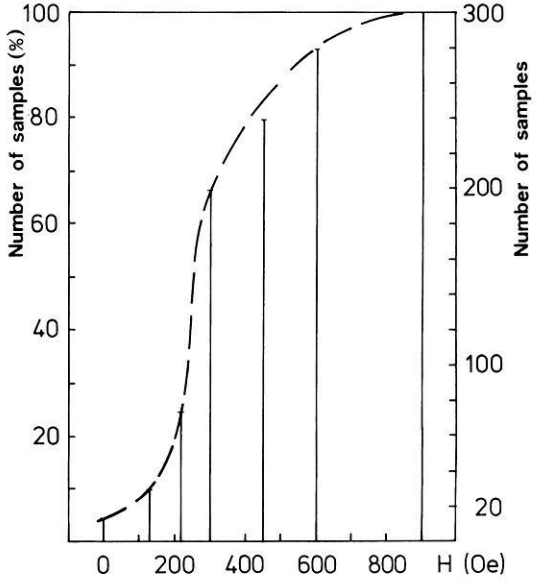


Fig. 3. Cumulative frequency distribution of maximum coercivities of the 'unstable' component of the Permian volcanics

properties of the volcanics vary not only from region to region, but also from site to site, it was necessary to demagnetize all specimens in at least three steps in order to determine the optimum demagnetizing field.

In about 40% of all demagnetized samples the characteristic remanence is found after demagnetization with 300 Oersted; 5% show an initial NRM

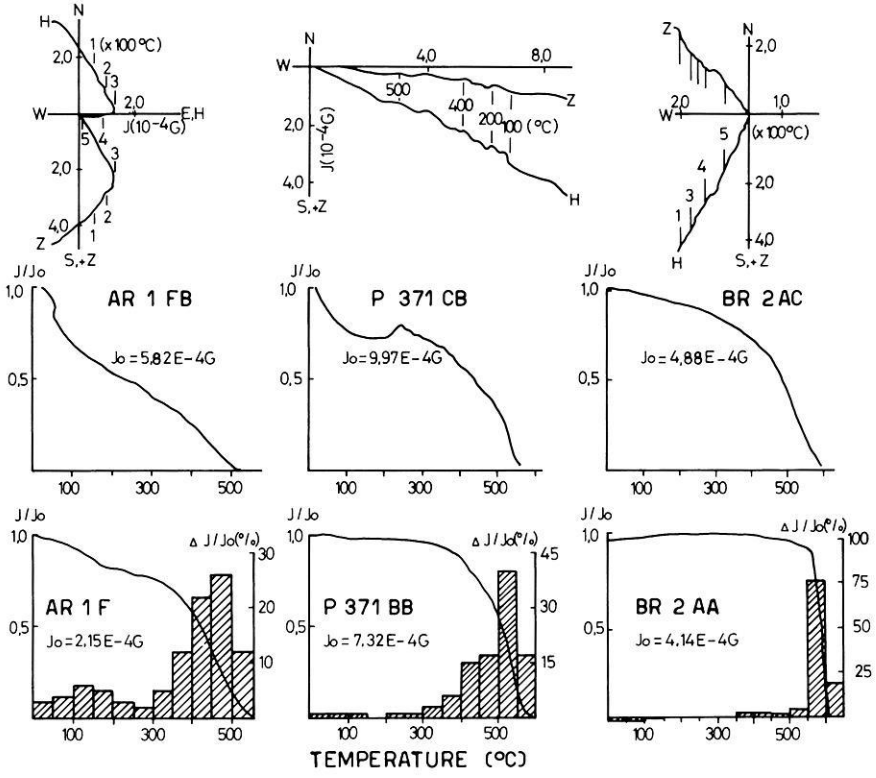


Fig. 4. Results of thermal demagnetization of three rock samples. *Top*: Orthogonal projections of the NRM directions as a function of temperature. *H* denotes the horizontal magnetization component, defined by N-S and E-W axis. *Z* is the projection of magnetization vector on the E-W vertical plane. *Middle*: Corresponding normalised intensity in terms of initial remanence. *Bottom*: Normalised intensity and histogram of blocking temperatures of progressive thermally demagnetized samples from the same core

direction practically identical to that after demagnetization with high fields, whereas another 5% achieve a stable direction only upon cleaning in fields between 600 and 900 Oe. For the total of the samples tested the non-characteristic remanence is mainly carried by coercivities between 200 Oersted and 300 Oersted, as can be seen from the steepest part in the graph of Fig. 3.

To examine the directional distribution of the stable components and to get an idea about the blocking temperature distribution of the remanence, both continuous and progressive thermal demagnetizations have been carried out. For the latter a special spinner magnetometer was built to record continuously the direction and intensity of the remanence vector during heating (Heiniger and Heller, 1976).

The vector plot of the directions (Fig. 4) sometimes shows an initial secondary component which is removed below 300° C. However, above 300 °C the remanence decreases towards the origin along an almost straight line.

This observation is valid for the continuous as well as for the progressive thermally demagnetized samples. Therefore we conclude that, if secondary remanence directions are present in the high blocking temperature range, they do not differ from any primary directions.

The most striking feature of the progressive thermal demagnetization curves of many samples is the extreme stability at lower temperatures (Fig. 4: samples P371BB and BR2AA). The blocking temperature spectra are often restricted to temperatures between 400 °C and 600 °C. In the case of sample BR2AA the remanence is partly carried by a mineral fraction with blocking temperatures above 580 °C. These high values together with the square shaped intensity vs. temperature curve may indicate that titanohematite is the main carrier of remanence. In sample P371CB the peculiar NRM intensity increase without change of direction around 200 °C during continuous thermal demagnetization may also be caused by the presence of titanohematite (cf. Heller and Egloff, 1974). In the case of samples AR1F and AR1FB the blocking temperatures are distributed over a much wider range and reach maximum values of only 500 °C to 550 °C. Also the stability is less pronounced. Therefore the NRM may be carried mainly by Ti-poor magnetite.

4. Magnetic Mineralogy

About 25 polished sections have been studied. Ore minerals occur only accessorily. Their concentration is always smaller than one percent by volume. Both hematite and magnetite have been identified. Paramagnetic ore minerals such as rutile and pyrite are rare, but they have been observed in samples from the Valle Brembana region. The ore minerals are generally observed along large grains of hornblende and pyroxene. Often they also are clustered in small (former) cracks. Very small opaque grains with diameter around 1 μm occur in the rock matrix originating from altered mafic minerals. The grain diameters vary from about 200 μm down to submicroscopic size. Magnetite can be observed in idiomorphic, primarily crystallized grains but also as a pseudomorphic product of pyroxene. Ilmenite exsolutions within the magnetite grains are frequently found. Magnetite grains are also observed in the neighbourhood of titanohematite which shows exsolution lamellae of ilmenite. These exsolutions indicate high temperature oxidation of the original hemoilmenite.

The temperature dependence of spontaneous magnetization of powdered rock samples confirm the results of thermal demagnetization; Curie points characteristic for titanohematite and Ti-poor magnetite have been detected.

Following Dunlop (1972), acquisition curves of isothermal remanent magnetization (IRM) can be used to establish the distribution of coercivities in a rock sample by plotting the increment of change of magnetization ΔJ against change of applied dc-field ΔH. Using superconductive coils magnetic fields of up to 50 kOe can be produced. This is sufficient to saturate most of the rockforming ferromagnetic minerals. Dunlop's method allows the identification of these minerals by determining their complete coercivity-spectrum without causing irreversible chemical changes of mineralogy, which is an ever present danger during thermal demagnetization.

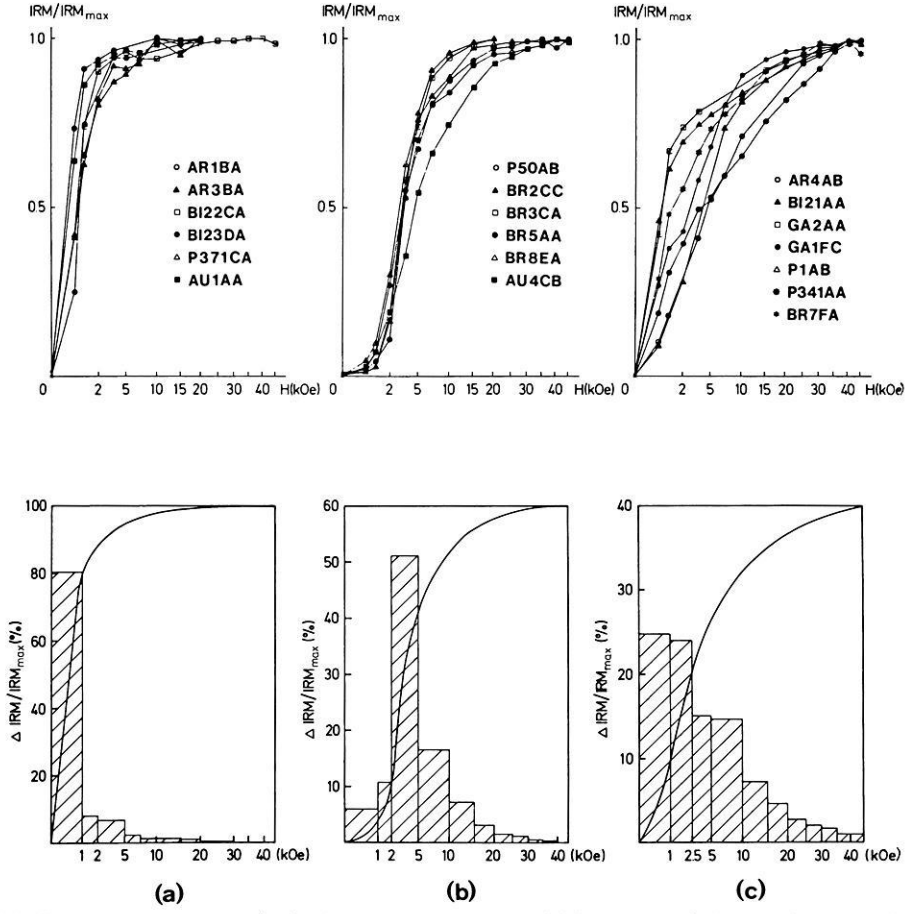


Fig. 5a–c. *Top*: Representative isothermal remanence acquisition curves of the Permian volcanics. Three different types (**a**), (**b**), and (**c**) can be distinguished. *Below*: Generalised coercivity distribution for each type

From most of the sampling sites one to two samples have been investigated using this method. The IRM acquisition curves and the results of the AF-cleaning experiments show that many of the Permian volcanics contain a considerable amount of high coercivity NRM-components. Three different types of coercivity spectra can be distinguished (Fig. 5). The first group consists of samples with predominantly low coercivity components (Fig. 5a). These are dominated by magnetite. The second group consists of samples without low coercivity components below 2 kOe and saturating between 10 kOe and 20 kOe. At $H=5$ kOe only two third of the saturation remanence is acquired. These samples contain predominantly titanohematite (Fig. 5b). A third group (Fig. 5c) comprises samples containing low- as well as very high coercivity components. The low coercivity magnetic minerals are again magnetite, but the high coercivity fraction is more difficult to identify from such data. On the basis of the minerals

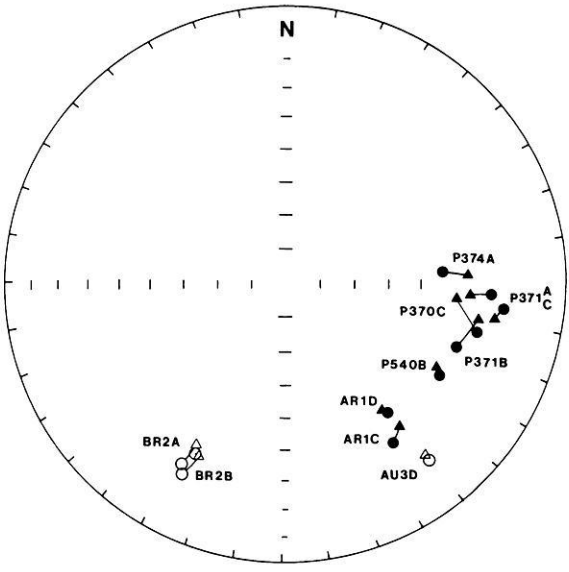


Fig. 6. Comparison of the directions of stable remanence of thermally demagnetized samples (●) with those of corresponding AF-demagnetized samples (▲). *Open symbols* denote negative inclination (upper hemisphere); full symbols denote positive inclination (lower hemisphere)

identified microscopically and by Curie-temperature measurement, the most likely mineral is titanohematite. The nature of the ferromagnetism and magnitudes of the spontaneous magnetization, anisotropy and magnetostriction constants are rather poorly known for natural hematite. However, Roquet (1954) has reported experimentally determined coercivities up to 30 kOe in hematite.

5. Palaeomagnetic Directions

A comparison of thermally cleaned samples with corresponding AF-cleaned samples from the same core (Fig. 6) shows that the cleaned directions are very similar with both methods, whether the magnetic properties are due to magnetite or titanohematite. Therefore, the ChRM directions are considered to be representative for the primary NRM also in those samples which contain a large titanohematite fraction and which have been shown to be very stable against AF-demagnetization, e.g. from the Valle Brembana sampling area and also from the Lugano porphyry district.

As the thermal and alternating field demagnetization show only one single stable directional component, the direction of any stable secondary magnetization must be similar to those of the characteristic primary remanence, and may therefore also have been acquired in the Permian.

The mean ChRM directions within individual sites are well grouped. The semi angle α_{95} of Fisher's cone of confidence is mostly smaller than 10°. The site mean directions, however, are dispersed, due to local and regional tectonic complications (Table 1).

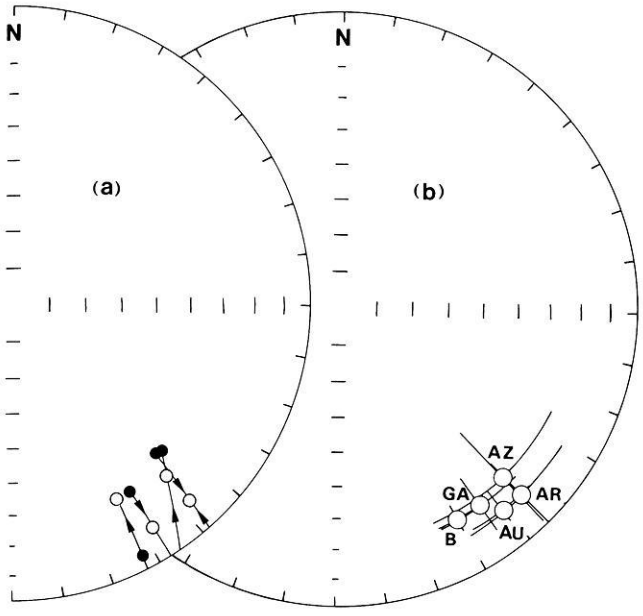


Fig. 7. a Stereographic projection of site-mean directions of ChRM for the sampling area of Ganna before and (*arrow*) after tectonic correction. For symbols see Fig. 6. **b** Regional mean directions after tectonic correction with main axes of the α_{95}-confidence oval. *AR*: Arona; *GA*: Ganna (western part of the Lugano sampling area); *AU*: Auccia volcanics compared with (*AZ*) the results given by Zijderveld and De Jong (1969) and (*B*) the Bolzano quartz porphyries (Zijderveld et al., 1970)

The (Alpine) tectonic movements have considerably affected the relative position of the Permian volcanics. At some places geological data do not allow a reconstruction of the tectonic movements. This is especially the case for the volcanics at the upper Valle Brembana and also for those from the Eastern part of the Lugano district, where the local tectonic situation has turned out to be more complicated than expected.

The usual dip correction about horizontal fold axes fails in these two regions. It would require unrealistically steep dip angles (between 80° and even more than 180°) to get site mean directions characteristic for the Southern Alps (as e.g. Bolzano, Fig. 7b). At the Valle Brembana region there is no coincidence of the observed orientation of stratification planes (cf. Table 1) and the hypothetically required unfolding. In the eastern part of the Lugano porphyry district, a lack of sediments or recognizable stratification due to erosion and weathering again does not allow a reliable tectonic correction. In both cases several phases of differential (Alpine) tectonic movements may have caused the strongly dispersed site mean directions within these regions and prevent a simple dip correction.

At Arona, Ganna, and Auccia the tectonic situation can be accounted for. The mean remanence directions for the sites around Ganna are plotted in Fig. 7a before and after tectonic correction. The correction was done by unfold-

Table 2. Mean direction of ChRM of the sampling areas with palaeomagnetically relevant remanences. Basis of statistic calculation: N; (cf. Table 1)

Region	Mean site pos.		Direction of ChRM after tectonic correction						Virtual palaeomagnetic pole position			
	Lat.N	Long.E	N	n	D	I	α_{95}	kF	Lat.N	Long.E	dm	dp
Auccia	45.8	10.4	5	74	139.7	−14.1	8.2	88.8	38.1	244.9	8.4	4.3
Ganna	45.8	8.8	4	31	142.9	−20.4	10.4	75.5	42.7	242.6	10.9	5.7
Arona	45.7	8.5	4	32	134.9	−14.2	13.6	46.5	35.4	248.0	13.9	3.1

ing the SE dipping overlying layered Triassic dolomites. Before correction all site means have a positive inclination; after unfolding, the inclinations become negative as expected for Permian rocks in this region. The data confirm the results of an early study carried out by Van Hilten and Zijderveld (1966). In the region of Arona most site mean directions can again be tectonically corrected by means of the bedding orientation of the overlying Triassic dolomites. The site mean directions of the Auccia volcanics were tectonically corrected by means of the bedding orientation of the underlying Collio formation. The mean directions of these three sampling areas after dip correction are shown in Fig. 7b. The data agree with those given by Zijderveld and De Jong (1969) and also with those from the Bolzano quartz-porphyries (Zijderveld et al., 1970) although this area is on the eastern side of the Giudicaria fault line.

6. Discussion

At all sites where the local tectonic situation allows a correction of the characteristic site mean directions by means of simple unfolding of formerly horizontal boundary surfaces, there is an improvement of their grouping. Moreover, there is not only a reduction of dispersion of a certain sampling area; there is also a close grouping of these corrected mean directions between each sampling area (Table 2). Besides a fold test, carried out by Pavoni et al. (1969) in the Lugano area, this may be considered as further evidence that the ChRM directions represent a stable initial remanence of the volcanics and therefore record the direction of the Permian geomagnetic field at the sampling sites.

The tectonically corrected remanence directions of all the volcanics investigated in the western Southern Alps (cf. Table 2) can also be compared with those of the Bolzano porphyries and with Upper Palaeozoic directions from extra-Alpine Europe (Fig. 8). The similarity of the mean directions of all these Permian volcanics suggests that the western Southern Alps have, on a large scale, behaved as a single tectonic block.

As the directions coincide with the well defined results from the Bolzano porphyries (eastern Southern Alps), it can be concluded, that the Southern Alpine area as a whole constitutes one single crustal block since the Permian.

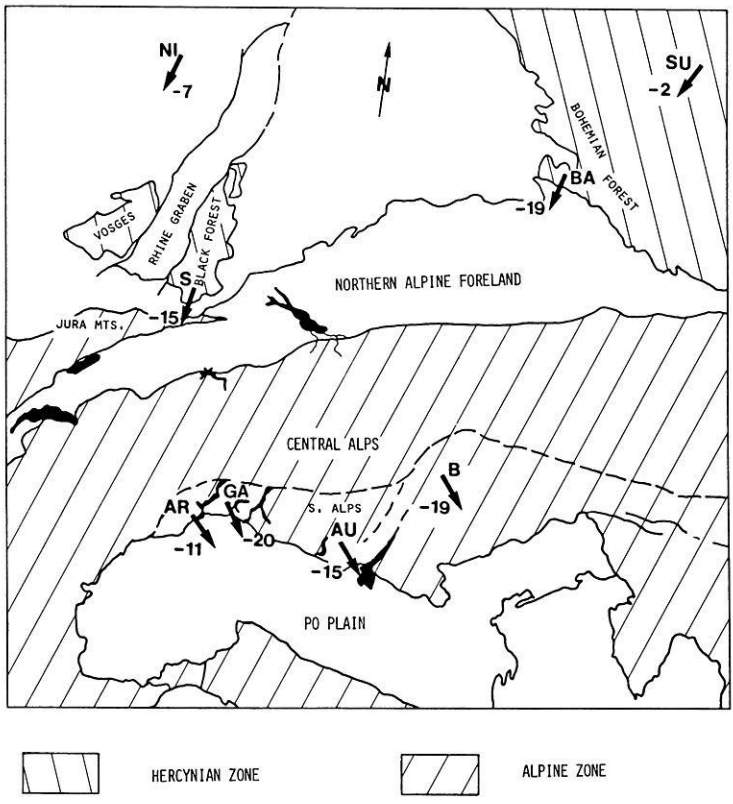


Fig. 8. *Arrows* show the remanence declinations of Southern Alpine Upper Palaeozoic rocks compared with those of the same age from extra-Alpine Europe (numbers denote the appropriate site mean inclinations). *AR*: Arona, *GA*: Ganna and *Au*: Auccia (cf. Table 2); *B*: Bolzano (Zijderveld et al., 1970); *Ni*: Nideck (Nairn, 1957); *S*: Black Forest (Konrad and Nairn, 1972); *BA*: Bavarian Forest (Heller, personal communication); *SU*: Inner Sudetic Basin (Birkenmajer et al., 1968)

The magnetization directions are very different from extra-alpine European directions. The declinations indicate that this block has been rotated by about 50° anticlockwise relative to extra-Alpine Europe since the Early Mesozoic.

In fact the Southern Alpine directions are very similar to African directions for that period. The Permian Southern Alpine palaeopoles are situated close to the Permian part of the African polar wander path of Van der Voo and French (1974) and nowhere near the polar wander path of stable Europe (Fig. 9). The palaeomagnetic data support other geophysical and sedimentological arguments that the Southern Alps may be considered a parautochthonous extension of the African plate since the Early Mesozoic (Channell and Horváth, 1976). Therefore, it seems realistic to assume that the Southern Alpine region has moved together with the African lithospheric plate which has rotated anticlockwise by about 50° relative to the European plate since the Early Mesozoic.

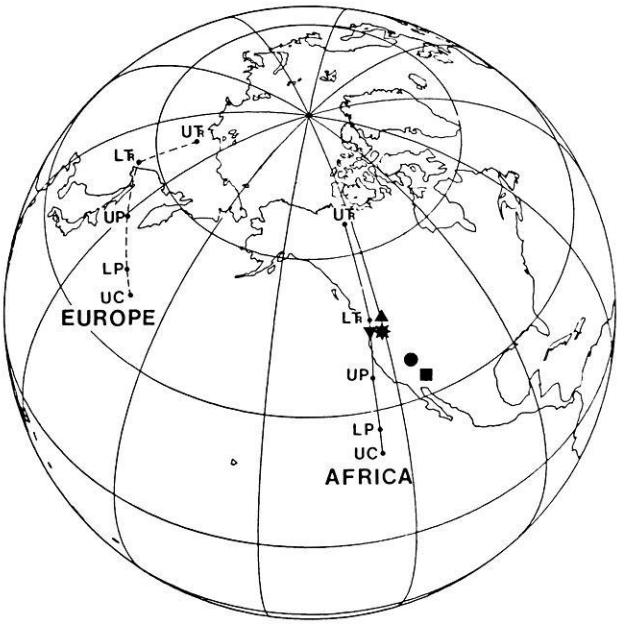


Fig. 9. Virtual palaeomagnetic pole positions for Auccia (●), Ganna (★) and Arona (■) as listed in Table 2, compared with the African and European polar wander path from Upper Carboniferous (*UC*) to upper Triassic (UT_R) from the data of Van der Voo and French (1974), and the Southern Alpine pole position of Bolzano (▼, Zijderveld et al., 1970) and the (Upper Triassic) Dolomites (▲, Manzoni, 1970). Modified after Channell et al. (1979)

Acknowledgements. I would like to thank to Professor Dr. W. Lowrie for his interest and advice and also for his help in writing up the English manuscript. My thanks are also due to PD Dr. F. Heller for many useful comments and invaluable discussions.

References

Birkenmajer, K., Krs, M., Nairn, A.E.M.: A palaeomagnetic study of upper Carboniferous Rocks from the Inner Sudetic Basin and the Bohemian Massif. Geol. Soc. Am. Bull. **78**, 589–608, 1968

Cassinis, G.: Rassegna delle formazioni permiane dell'alta Val Trompia (Brescia). Atti Ist. Geol. Univ. Pavia **17**, 50–66, 1966

Cassinis, G.: La sezione-tipo delle Vulcaniti di Auccia (Permico bresciano). Atti Ist. Geol. Univ. Pavia **19**, 40–49, 1968

Channell, J.E.T., D'Argenio, B., Horváth, F.: Adria, the African Promontory, in Mesozoic Mediterranean Palaeogeography. Earth Sci. Rev. (in press) 1979

Channell, J.E.T., Horváth, F.: The African/Adriatic Promontory as a palaeogeographical premise for Alpine Orogeny and plate movements in the Carpatho-Balkan Region. Tectonophysics **35**, 71–101, 1976

Dunlop, D.J.: Magnetic mineralogy of unheated and heated red sediments by coercivity spectrum analysis. Geophys. J.R. Astron. Soc. **27**, 37, 1972

Fisher, R.A.: Dispersion on a sphere. Proc. R. Soc. London, Ser. A: **217**, 295–305, 1953
Goree, W.S., Fuller, M.: Magnetometers using RF-driven squids and their applications in rock magnetism and palaeomagnetism. Rev. Geophys. Space Phys. **14**, 4, 591–608, 1976
Graeter, P.: Geologie und Petrographie des Malcantone (Suedliches Tessin). Schweiz. Mineral. Petrogr. Mitt. **31**, 2, 1952
Heiniger, C., Heller, F.: A high temperature vector magnetometer. Geophys. J.R. Astron. Soc. **44**, 281–288, 1976
Heller, F., Egloff, R.: Multiple Reversals of Natural Remanent Magnetization in a Granite-Aplite Dyke of the Bergell Massif (Switzerland). J. Geomagn. Geoelectr. **26**, 499–505, 1974
Hilten, D. Van, Zijderveld, J.D.A.: The magnetism of the Permian Porphyrics near Lugano (Northern Italy/Switzerland). Tectonophysics **3**, 429–446, 1966
Jaeger, E., Faul, H.: Age measurements on some granites and gneisses from the Alps. Geol. Soc. Am. Bull. **70**, 12, 1553–1558, 1959
Kaech, M.: Geologisch-petrographische Untersuchungen des Porphyrgebietes zwischen Lago Maggiore und Val Sesia. Eclogae Geol. Helv. **8**, 1, 47–164, 1903
Konrad, H.J., Nairn, A.E.M.: The Palaeomagnetism of the Permian Rocks of the Black Forest, Germany. Geophys. J.R. Astron. Soc. **27**, 369–382, 1972
Kuenen, Ph.D.: The porphyry district of Lugano West of Valganna. Leidse Geol. Meded. **1**, 129–185, 1925
Manzoni, M.: Palaeomagnetic data of Middle and Upper Triassic age from the Dolomites (Eastern Alps, Italy). Tectonophysics **10**, 411–424, 1970
Molyneux, L.: A complete result magnetometer for measuring the remanent magnetization of rocks. Geophys. J.R. Astron. Soc. **24**, 429–433, 1972
Nairn, A.E.M.: Observations palaeomagnetiques en France: roches Permiennes. Bull. Soc. Geol. Fr. **7**, 721–727, 1957
Pavoni, N., El Mikacher, S., Weber, M.: Gesteinsmagnetische Untersuchungen an permischen Porphyriten der Morcote Halbinsel suedlich Lugano. Schweiz. Mineral. Petrogr. Mitt. **49**, 1, 103–107, 1969
Pichler, H.: Neue Ergebnisse zur Gliederung der unterpermischen Eruptivfolge der Bozener Porphyr-Platte. Geol. Rundsch. **48**, 112–131, 1959
Porath, H., Stacey, F.D., Aik, S.C.: The choice of specimen shape for magnetic anisotropy measurements on rocks. Earth Planet. Sci. Lett. **1**, 92, 1966
Roquet, J.: Sur les rémanences des oxides de fer et leur intérêt en géomagnétisme. Ann. Géophys. **10**, 226–247, 282–325, 1954
Sitter De, L.U.: Les porphyres luganois et leurs envelopes. Leidse Geol. Meded. p. 11, 1939
Sitter De, L.U., Sitter-Koomans, De, C.M.: The geology of the Bergamasc Alps, Lombardia, Italy. Leidse Geol. Meded. **14**B, 1949
Voo, R. Van der, French, R.B.: Apparent polar wandering for the Atlantic bordering continents: Late Carboniferous to Eocene. Earth Sci. Rev. **10**, 99–119, 1974
Zijderveld, J.D.A., Jong, K.A.: Palaeomagnetism of some Late Palaeozoic and Triassic rocks from the eastern Lombardic Alps, Italy. Geol. Mijnbouw **48**, 6, 559–564, 1969
Zijderveld, J.D.A., Voo, R. Van der: Palaeomagnetism in the Mediterranean area. In: Implications of continental drift to the earth sciences, D.H. Tarling, S.K. Runcorn, eds.: pp. 133–161. New York: Academic Press 1973
Zijderveld, J.D.A., Hazeu, G.J.A., Nardin, M., Voo, R. Van der: Shear in the Tethys and the Permian palaeomagnetism in the Southern Alps, including new results. Tectonophysics **10**, 639–661, 1970

Received September 7, 1979; Revised Version October 29, 1979

J. Geophys. 46, 413–427, 1979

Journal of Geophysics

Palaeomagnetism of Upper Cretaceous Limestones From the Münster Basin, Germany*

F. Heller and J.E.T. Channell

Institut für Geophysik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

Abstract. Marine, mainly flat-lying sediments of Late Cretaceous age are exposed throughout a wide area of the Münster Basin (NW-Germany). During the Cenomanian, Turonian, and Campanian fine-grained, grey, marly to pure limestones were deposited. The Campanian limestones carry magnetization components of unknown age due to the presence of secondary goethite and haematite. However, in the Cenomanian and Turonian rocks the natural remanent magnetization (NRM) is due to detrital magnetite and can be associated with the time of deposition. Fold tests confirm a Late Cretaceous age of magnetization in the magnetite-bearing limestones, since the NRM pre-dates latest Cretaceous deformation along the northern margin of the basin. The Münster Basin limestones provide one of the first reliable Cretaceous pole positions (Lat.: 76° N, Long.: 181° E) from stable Europe.

Key words: Palaeomagnetism – Limestones – Cretaceous – Stable Europe.

1. Introduction

In recent years two important polar wander curves for Europe have been compiled. That by Van der Voo and French (1974) covers the last 300 m.y. and that by Irving (1977) spans the last 375 m.y. Both curves show the same general trend with the pole moving apparently from low latitudes to its present high latitude position. However, the two polar wander paths differ significantly in their longitudinal position especially during the Jurassic and Cretaceous. This divergence is a result of the different compilation techniques used for the two polar wander curves. Mesozoic palaeomagnetic data from stable Europe are very scarce and a Mesozoic polar wander path can not be compiled from existing European data. The Van der Voo and French (1974) polar wander path for Europe relies heavily on North American palaeomagnetic data and

* Contribution Nr. 264, Institut für Geophysik, ETH-Hönggerberg, Zürich, Switzerland

0340-062X/79/0046/0413/$03.00

the fit of Europe and North America using the Atlantic magnetic lineations. On the other hand, Irving's (1977) European polar wander path incorporates new Russian palaeomagnetic data, and relies on the premise that data from the Caucasus and east of the Urals are applicable to stable Europe.

The aim of the present study is to establish a reliable Cretaceous pole position for central Europe and to compare this pole position with those inferred for Europe by Van der Voo and French (1974) and by Irving (1977). The Mesozoic in Europe is documented by a great variety of sedimentary rocks, especially carbonates. Many recent investigations have shown that limestones can be reliable recorders of the ancient geomagnetic field, although their remanent magnetization is usually very weak. Nevertheless, since the development of superconducting rock magnetometers (Goree and Fuller 1976), it is possible to measure certain limestones with sufficient precision. Late Jurassic limestones from Germany (Heller 1977, 1978a) have recently given a well defined Jurassic pole position for stable Europe. The Upper Cretaceous limestones of the Münster basin (Fig. 1) are usually flat-lying and tectonically undisturbed, they outcrop over a large area of Westphalia, and are extensively quarried affording good fresh outcrop.

2. Geology

The development of the Upper Cretaceous Münster Basin (Fig. 1) commenced during the Albian and Cenomanian due to subsidence of the northern part of the Rhenish Massif. At this time the sea transgressed the Münster Basin from the Lower Saxony Basin in the north (Arnold 1964). Between Cenomanian and Campanian a huge pile of marine sediments – up to nearly 2,000 m – was deposited in water depths not exceeding 200 m (Arnold 1964). The facies vary from transgressional conglomerates, glauconitic sands, marls, and limey marlstones to pure limestones. The basinal facies is characterized by a marly-calcareous sedimentation with increasing carbonate content in the Cenomanian (Thiermann and Arnold 1964) ranging from interlayered marls and marl-limestone sequences ('Plänerkalke') to pure limestones. The Turonian is also represented by light-coloured, grey and mostly pure limestones with interlayers of calcareous marlstones. Beginning in the uppermost Turonian clayey marls were deposited continuously throughout the Coniacian, Santonian and Lower Campanian. During the Upper Campanian marl sedimentation continued, with intercalations of calcareous marls and limestones.

The sediments are flat-lying and tectonically undisturbed throughout most of the Münster Basin. Towards the margins, especially along the northern margin, movements occurred during the so-called 'Subhercynian' orogenic phase (Stadler and Teichmüller 1971) in latest Cretaceous times. This event caused steep tilting of the sedimentary layers, flexures, and even small scale overthrusts up to 500 m (Lotze 1953; Rosenfeld 1963) due to uplift of the North-Westphalian Block (Fig. 1). However, local fold axes are essentially horizontal and block rotations unlikely (Rosenfeld, personal communication).

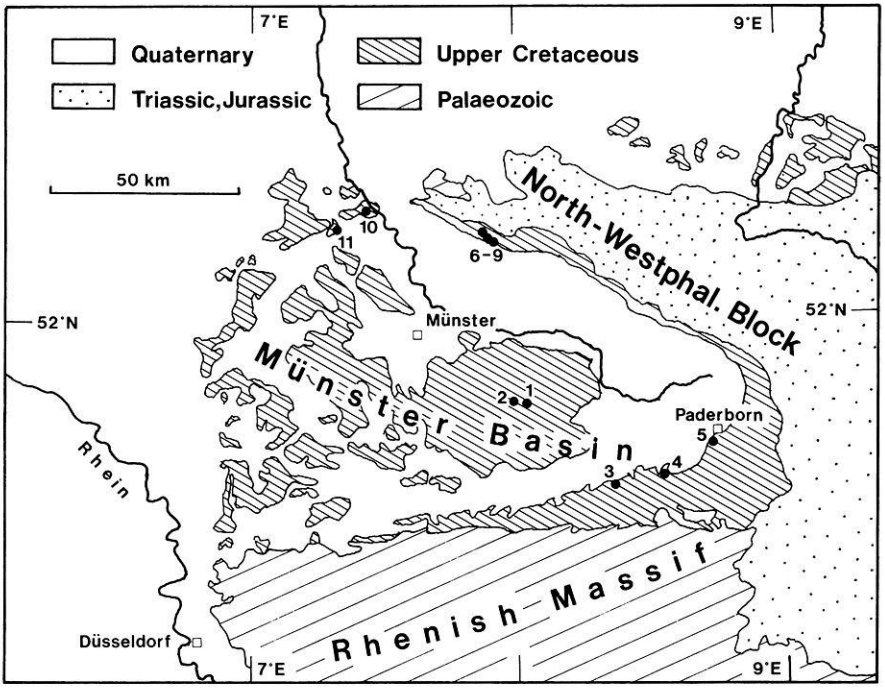


Fig. 1. Location of sampling sites (Nos. 1–11) in the Münster Basin

Since the marls were not accessible to palaeomagnetic sampling and fresh sandstone outcrops are rare, our investigations concentrate on the carbonate rocks which are quarried at many places. Eleven sites (Fig. 1) have been sampled which expose limestones of Cenomanian, Turonian, and Upper Campanian age.

3. Magnetic Properties

The natural remanent magnetization (NRM) of more than 500 limestone samples from the Münster Basin has been investigated. The initial NRM intensities vary between $4 \cdot 10^{-9}$ G and $5 \cdot 10^{-7}$ G with a mean value around $1 \cdot 10^{-7}$ G (Fig. 2). Although weak in intensity most of the NRM vectors could be determined with sufficient accuracy using the cryogenic magnetometer.

Depending on the NRM intensity, each sample was measured either in three or six different positions with respect to the instrument's triaxial detector system, thus allowing for three or six independent estimates of the remanence vector. These estimates were combined to give a mean value of the sample's magnetization vector together with some precision parameters (Heller 1978b; Lowrie et al. 1979) such as standard deviation of intensity and angular standard deviation of direction. Besides operator errors, the precision of measurement is influenced by the signal/noise ratio, sample shape defects, instability and

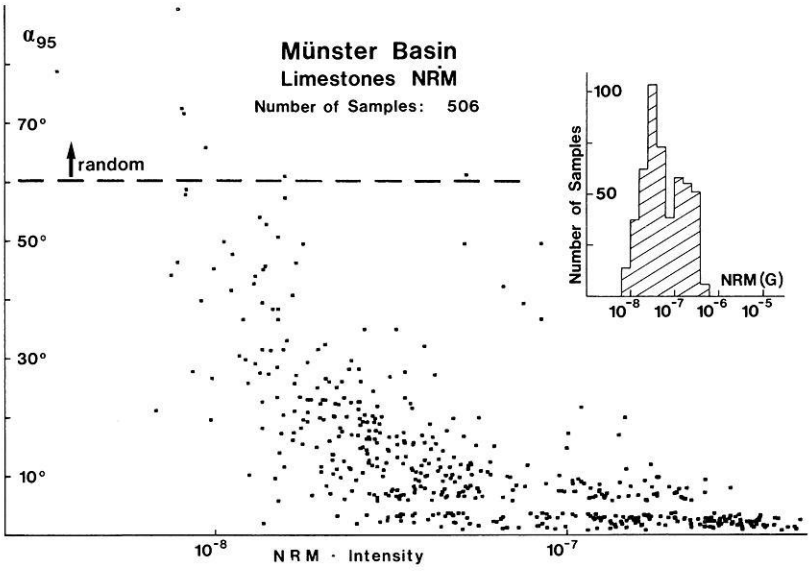


Fig. 2. The angular accuracy α_{95} from 3-position measurements for each individual sample plotted against the NRM intensity of that sample

inhomogeneity of remanence. When measuring in six positions, the inhomogeneity effects are minimized due to the completely symmetric measurement scheme.

Figure 2 shows the angular accuracy of measurement (α_{95}) of the NRM of each limestone sample to be largely intensity dependent. With decreasing NRM intensity (i.e., with decreasing signal/noise ratio) α_{95} increases, and signifies randomness (Watson 1956) at the 95% confidence level when values reach 57.2° for 6-position measurements and 60.2° for 3-position measurements. These values representing randomness usually occur in samples with intensities below $1 \cdot 10^{-8}$ G. Less than 2% of the NRM measurements had to be rejected by this criterion but much higher percentage of the collection had to be eliminated after demagnetization, when the intensities often fell below the critical limit.

For the palaeomagnetic investigations more than 300 samples have been subjected to AF-cleaning in fields up to 400 Oe. Many of these samples were subsequently magnetized progressively in fields up to 45 kOe in order to provide additional information on their magnetic mineral content.

Type 1 Limestone. The magnetic mineralogy of the limestones is highly variable. The predominant magnetic limestone type is characterized by the behaviour illustrated in Fig. 3. The NRM intensity steadily decays during AF-cleaning. A secondary component is removed by AC fields between 50 Oe and at most 200 Oe. Exceeding these field amplitudes, the projections of declination- and inclination-components run within experimental accuracy towards the origin. Thus no higher coercivity components are found in these samples. The IRM acquisition curves are saturated around $H=2$ kOe and subsequent progressive

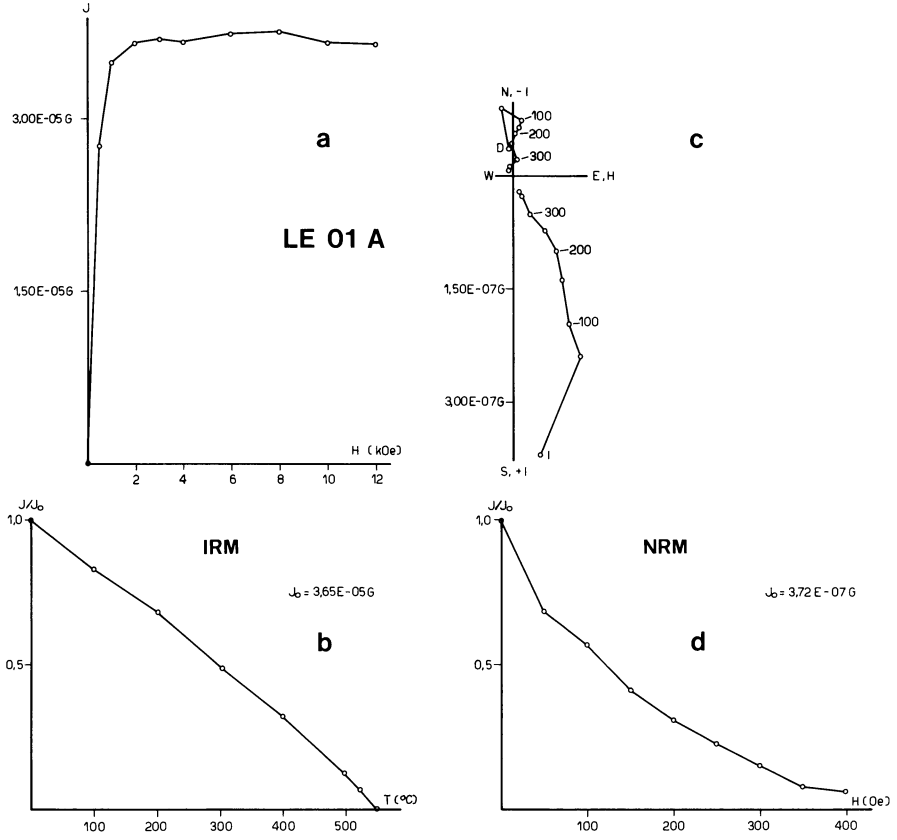


Fig. 3a–d. Type 1 limestones. **a** Acquisition of isothermal remanent magnetization (IRM) in increasing dc fields. **b** Stepwise thermal demagnetization of that IRM. **c** Vector diagram of declination (*D*) and inclination (*I*) during progressive AF demagnetization of the natural remanent magnetization (NRM). **d** The decrease in NRM intensity during progressive AF demagnetization

thermal demagnetization yields maximum blocking temperatures between 550° C and 575° C. All this evidence points to *magnetite* as the main carrier of remanence. Most of the limestones from sampling localities 5–11 (Fig. 1) qualify for this low coercivity type.

Type 2 Limestone. Magnetic high coecivity minerals predominate in a second type of limestone from the Münster Basin. The sample in Fig. 4 has an extremely stable NRM with a minor soft component which is removed by alternating fields of only 25 Oe. Up to 150 Oe (in other samples up to 400 Oe) no essential change of the NRM vector takes place. The IRM acquisition curve does not reach saturation even in fields as strong as $H=45$ kOe. Upon thermal treatment 90% of the IRM is lost below 150° C. The extremely high coercivity and the low maximum blocking temperatures are attributed to *goethite*. The spontaneous magnetization of goethite occurring in limestones is very sensitive to temperature

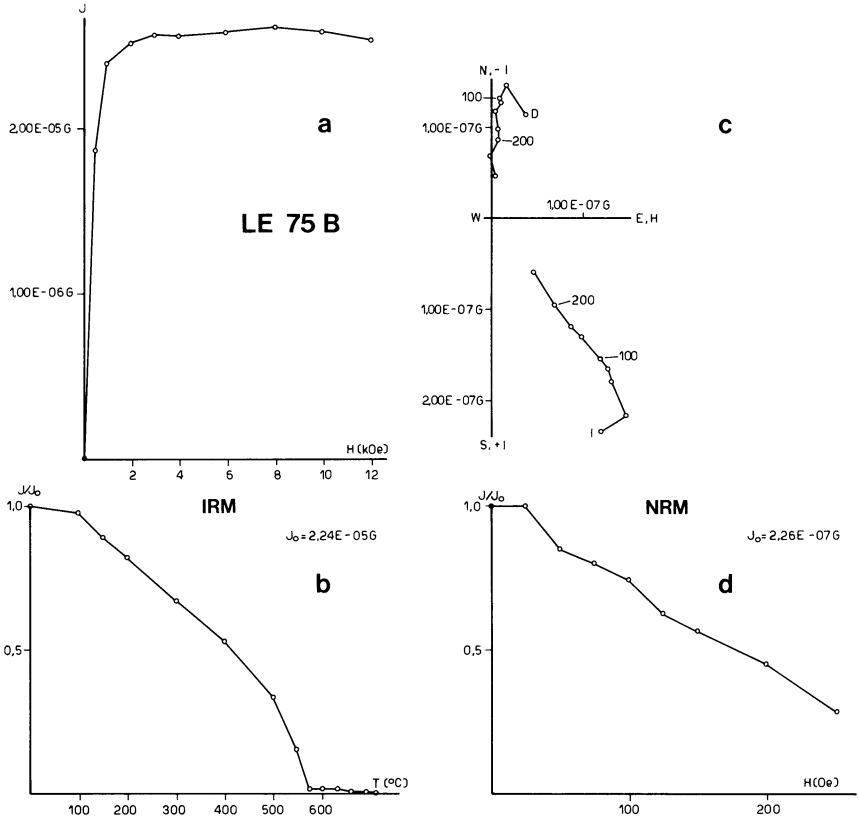


Fig. 3

fluctuation (Heller 1978a). This explains the partially erratic behaviour of the IRM acquisition curve. The IRM remaining after thermal demagnetization above 150° C may be due to the presence of either magnetite or hematite.

Type 3 Limestone. The sample in Fig. 5 again has a very stable NRM with only very little change of the magnetization vector during AF-cleaning. Having passed a small, very low intensity plateau below $H=5$ kOe, the IRM acquisition curve again is not saturated at $H=45$ kOe, the maximum field available. In this case, a much higher proportion of IRM is left after heating to 150° C. Since the maximum blocking temperatures of this portion lie well above 600° C (near to 700° C), it is concluded that besides *goethite, hematite* is a main contributor to the remanence of the sample. The stable level of the IRM curve between 2 kOe and 4 kOe as well as a small inflexion of the IRM thermal demagnetization curve near 550° C indicate a minor amount of magnetite to be present. Limestones of this type have been found mainly at the sampling localities 1–4 (Fig. 1).

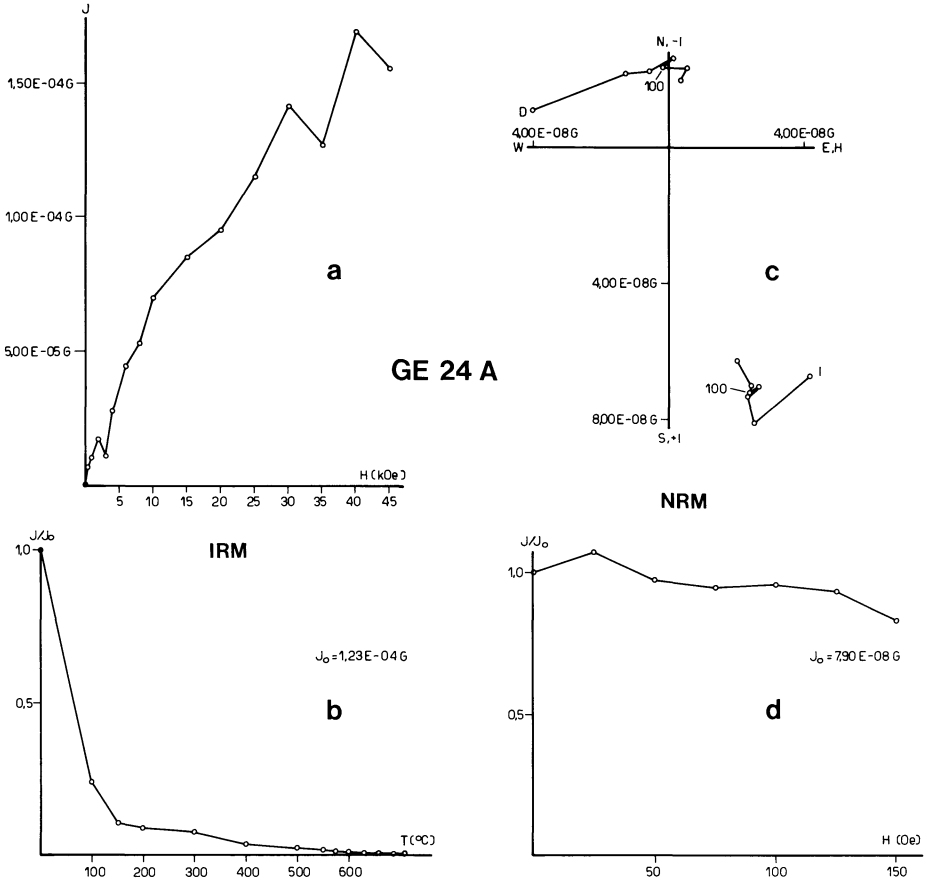


Fig. 4a–d. Type 2 limestones. **a**, **b**, **c**, and **d** see Fig. 3

Type 4 Limestone. The sample in Fig. 6 contains mainly *magnetite,* since the IRM intensity is mainly acquired in fields below $H=2$ kOe and a strong inflexion of the thermal demagnetization curve of IRM is found at 550° C. However, the IRM acquisition curve does not saturate below $H=3$ kOe, the maximum coercivity to be expected for magnetite, and a distinct IRM component survives heating to 575° C being totally obliterated above 625° C. Thus a substantial amount of IRM is contributed by *hematite.* The NRM contains again a soft component which is removed well below 100 Oe. Otherwise the directional stability of NRM is very high, but the trends of the projections of declination and inclination, although being straight lines within the limits of experimental error (note the low intensities), do not pass through the origin. This is caused by different magnetization directions in magnetite and hematite respectively.

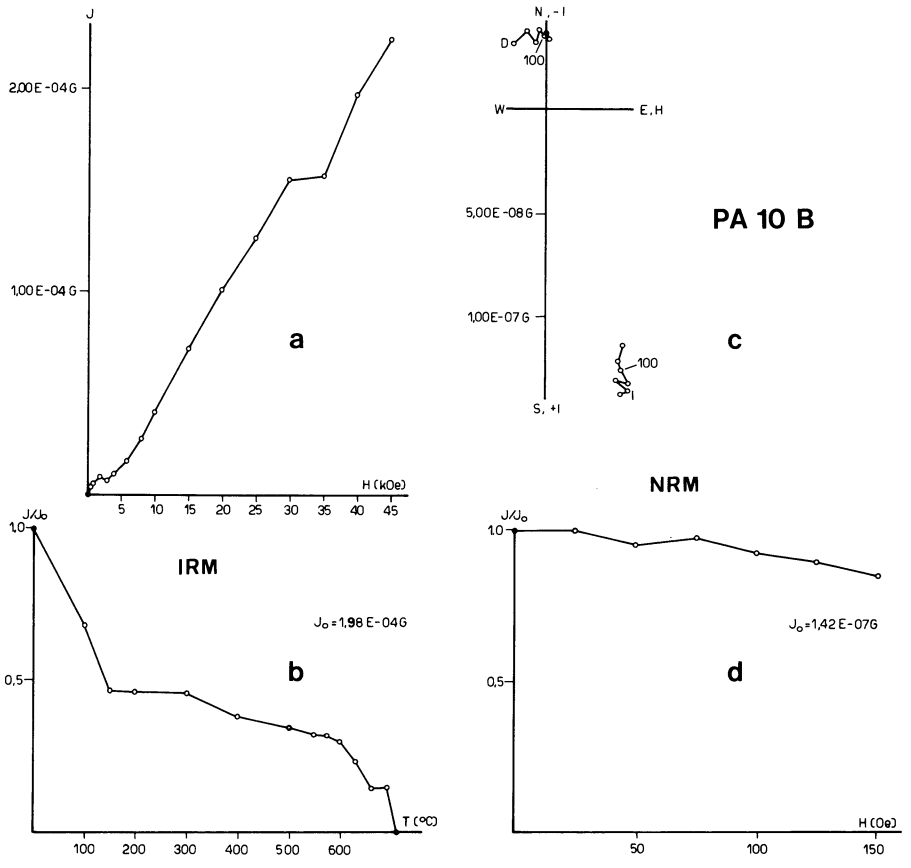


Fig. 5a–d. Type 3 limestones. **a**, **b**, **c**, and **d** see Fig. 3

4. Age of NRM and Selection of Reliable Palaeomagnetic Data

The magnetization history of a rock is often difficult to ascertain. Viscous remagnetization, oxidation and exsolution of thermodynamically unstable ferromagnetic minerals or secondary precipitation or recrystallization of magnetic minerals may lead to a complete overprint of more primary magnetizations.

The three ferromagnetic minerals magnetite, goethite and hematite which have been identified in the Münster Basin limestones by IRM studies, have been formed under different conditions. By analogy with deep sea sediments (Løvlie et al. 1971) and due to the fact that magnetite cannot precipitate by purely chemical reactions in seawater, it is generally agreed that most magnetite is of detrital, primary origin in carbonate sediments. The limestone NRM carried by magnetite therefore is assumed to be of synsedimentary origin, although viscous overprint cannot always be ruled out. However, two positive fold tests

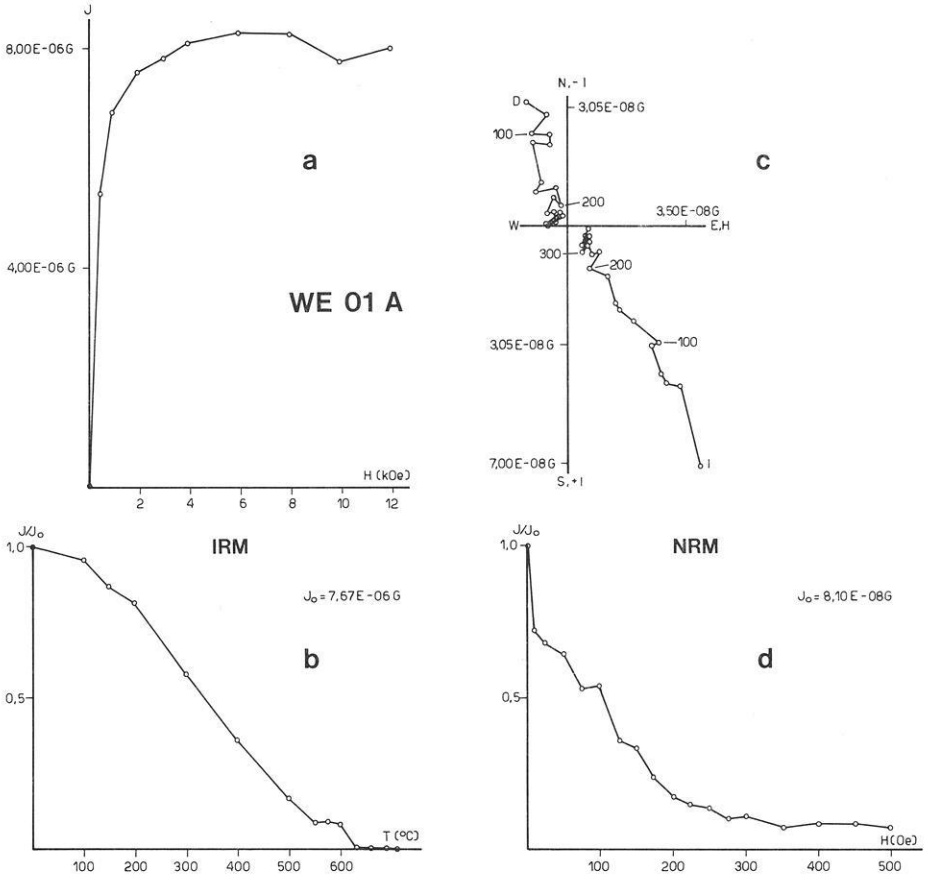


Fig. 6a–d. Type 4 limestones. **a**, **b**, **c**, and **d** see Fig. 3

(Fig. 7) allowed us to attest an early magnetization to the Münster Basin Type 1 limestones. The scatter of AF-cleaned, stable NRM directions is considerably reduced by unfolding of the limestone beds, and therefore a pre-folding origin of NRM is indicated. Since the main tectonic events in the area occurred during the so-called 'Subhercynian' orogenic phase in the latest Cretaceous, an Upper Cretaceous age of NRM can be established for this limestone type.

The age of the high coercivity magnetization components in limestone types 2–4 remains ambiguous, because no suitable fold tests could be conducted. There is evidence for later formation of goethite and hematite, the latter mineral being considered as a dehydration product of goethite (Berner 1969). During demagnetization of tilted mixed coercivity (Type 4) limestones, the magnetization components attributed to hematite often, but not always, have slightly different directions to those attributed to low coercivity magnetite (cf. Fig. 6). Although the NRM directions of the high coercivity samples (Types 2 and 3) are extremely stable during AF-cleaning and plot close to the low coercivity

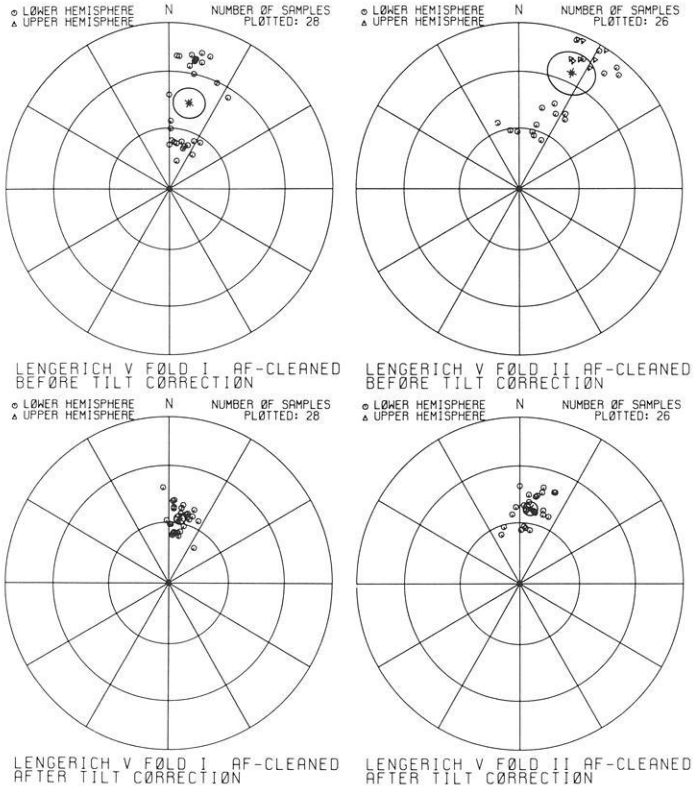


Fig. 7. Two fold tests on limestones from Lengerich (location 6,9; Fig. 1), indicating a pre-folding age for the AF-cleaned magnetization

directions, all samples of Type 2 and 3 have been discarded because of the likelihood that the high coercivity magnetizations were formed at a time much later than the deposition of the sediment. In the case of the mixed coercivity rocks (Type 4), vector diagrams (e.g., Fig. 6) were plotted for each individual sample in order to pick the direction of the low coercivity magnetization component. The higher coercivity components in these samples were considered to be not representative of the Upper Cretaceous geomagnetic field.

Some low coercivity magnetization components were shown to be unstable when applying a directional stability index (Heller 1977; Lowrie and Alvarez 1977) and had to be rejected. The data selection criteria together with the test on the accuracy of each individual measurement forced us to discard about one third of the demagnetized sample collection. The remaining samples which are considered to carry characteristic Upper Cretaceous NRM directions uncontaminated by later magnetization components, have been used for the calculation of the eleven site mean directions (Table 1).

Table 1. Palaeomagnetic data from the Münster Basin

Site (Location in Fig. 1)	Number of samples measured	Number of stable samples	Polarity %		Site mean direction after tectonic correction			Pole position			Stratigraphic position
			+	−	D	I	α_{95}	Lat. N	Long. E	α_{95}	
Neubeckum (1)	42	11	91	9	358.2	53.6	6.9	71.6	193.8	8.1	Upper Campanian
Enniger (2)	39	10	40	60	4.1	61.3	8.3	79.5	172.3	11.2	Upper Campanian
Erwitte (3)	23	14	100	–	8.3	61.3	6.5	78.5	156.8	9.2	Upper Turonian
Gesecke (4)	36	10	100	–	1.8	52.7	7.0	70.7	184.4	8.8	Upper Turonian
Paderborn (5)	17	14	100	–	3.3	59.3	5.6	77.4	177.3	6.7	Upper Turonian
Lengerich (6) (Fold I)	28	28	100	–	9.9	57.9	2.7	74.4	158.9	3.3	Middle Turonian
Lengerich (7)	21	21	100	–	355.7	57.1	3.9	74.9	202.2	5.2	Upper (?) Cenomanian
Lengerich (8) (Railway)	40	35	100	–	0.3	60.9	2.5	79.4	187.7	3.5	Middle Turonian
Lengerich (9) (Fold II)	26	26	100	–	8.2	52.6	3.5	69.7	168.7	4.0	Lower Turonian
Rheine (10)	25	24	100	–	0.2	63.1	3.2	82.1	188.0	4.7	Upper Cenomanian
Wettringen (11)	19	19	100	–	354.8	53.9	2.7	71.6	202.6	3.3	Upper Cenomanian
Mean (N=9)			100	–	2.5	57.8	3.1	75.8	181.1	3.8	Cenomanian-Turonian

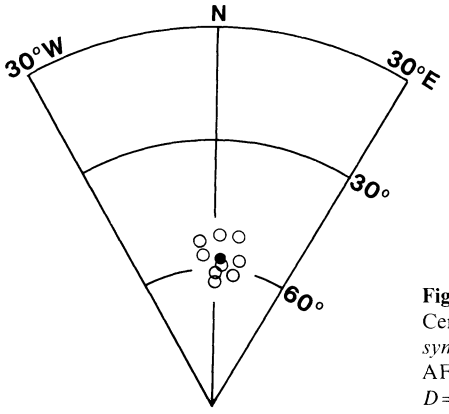


Fig. 8. Stereographic plot of the nine Cenomanian-Turonian site mean directions (*open symbols*; Table 1) after tectonic correction and AF-cleaning. The mean direction (*full circle*; $D=2.5°$, $I=57.8°$) has α_{95} of 3.1°

5. Palaeomagnetic Results and Conclusions

Table 1 shows that all the limestone samples of Cenomanian and Turonian age from the Münster Basin without exception carry NRM directions of normal polarity. The sections which we have sampled, certainly do not cover the entire time interval between 100 m.y. and 86 m.y., since our sections are spread across one quarter of the sediment pile at most. However, the normal polarity is consistent with our knowledge of the geomagnetic field during the Late Cretaceous. The Cenomanian and Turonian belongs to the Cretaceous quiet zone of normal polarity lasting from Early Aptian to the Santonian-Campanian (Lowrie and Alvarez 1977; Channell et al. 1979).

The Upper Campanian limestones from the Münster Basin are extremely weakly magnetized (initial NRM mean intensity: $2.7 \cdot 10^{-8}$ G) and only a small percentage of samples carried a characteristic low coercivity NRM (Table 1). Extremely stable, secondary magnetization components usually dominate the NRM of these limestones. The frequent negative polarity of the NRM indicates that this magnetization component significantly post-dates deposition because it has been established (Lowrie and Alvarez 1977) that the polarity of the geomagnetic field was normal during the Late Campanian when these limestones were deposited.

Therefore, the evaluation of an Upper Cretaceous mean direction for the Münster Basin is confined to the Cenomanian and Turonian limestones (Table 1, Fig. 8) which always carry a well-defined low coercivity magnetization component. The site mean directions (Fig. 8) cluster very closely ($\alpha_{95}=3.1°$) around a mean direction of $D=2.5°$ E and $I=57.8°$. This provides a reliable, well dated Upper Cretaceous pole for stable Europe (Long.$=181°$ E, Lat.$=76°$ N, $\alpha_{95}=3.8°$, $N=9$). The site pole positions and their mean have been plotted in Fig. 9.

Irving's (1977) polar wander curve for Eurasia shows an apparent minor loop during the Cretaceous (Fig. 9). Since Irving's Cretaceous pole positions have 95% confidence circles with radii of at least 5°, one might argue about

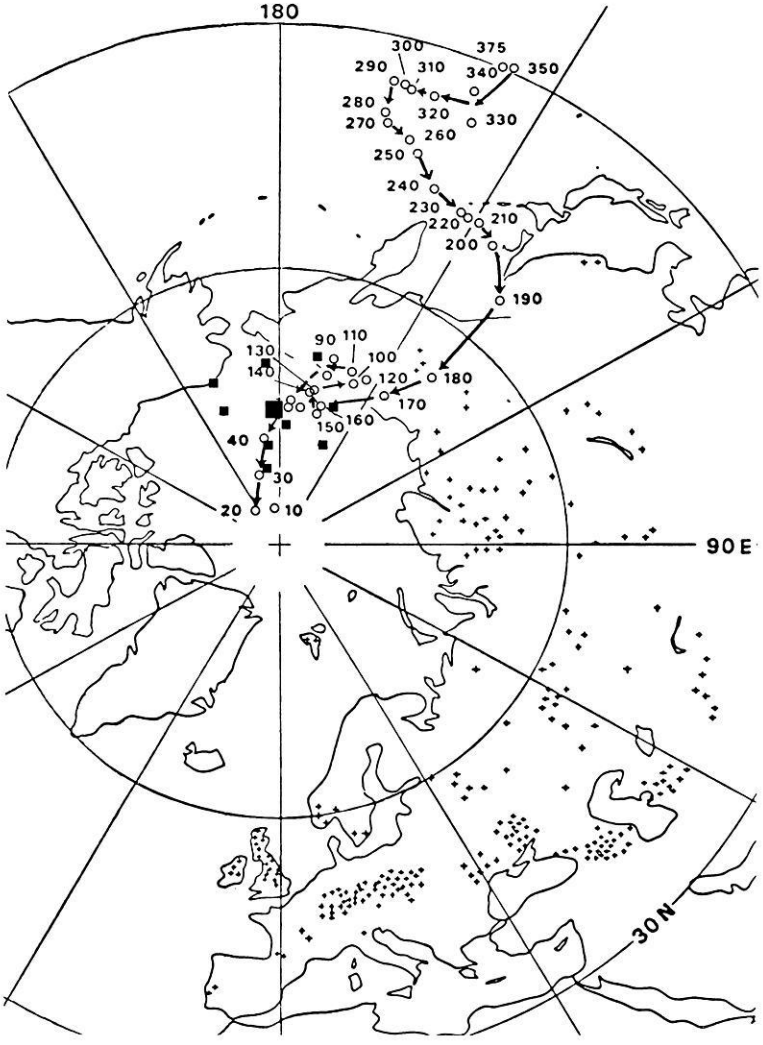


Fig. 9. The nine Upper Cretaceous pole positions from the Münster Basin (*small squares*) give a mean pole (*large square*) which does not coincide with the Upper Cretaceous portion of Irving's (1977) time-averaged polar wander path for Eurasia (*open circles*). The numbers give the age (m.y.) of the Eurasian poles

the significance of the loop. However, it produces a significant separation of our mean pole from Irving's time-averaging contemporaneous poles. The Münster Basin palaeopole indicates a simpler polar wander path for Europe than that derived by Irving. As mentioned above, Mesozoic palaeomagnetic data from Europe are almost non-existent and Irving's European APW curve relies heavily on Siberian poles and may be biased by data taken from orogenic belts (Caucasus).

The Upper Cretaceous Münster Basin palaeopole allows two tectonic conclusions to be drawn. The Upper Cretaceous Sintra granite in Spain (Van der Voo 1969) yields the same palaeopole position. Thus, Van der Voo's suggestion can now be confirmed that the anticlockwise rotation of the Iberian peninsula was complete in the Late Cretaceous. The Southern Alps have been interpreted by Channell and Tarling (1975) and Van den Berg and Wonders (1976), on the basis of palaeomagnetic measurements, to belong to an extension of the African plate. Cenomanian and Turonian sediments from this area carry an 'African' mean NRM direction $D=340.8°$ and $I=38.8°$. When comparing this value with the Münster Basin result, the minimum amount of rotation between Europe and Africa can be established to be about 20° since the Late Cretaceous. Also the former latitude of the Southern Alps can be estimated. Since the former latitude position of the Southern Alps is derived to be 9.5° ($\pm 3.8°$) more southerly than at present, the minimum shortening of the Tethys and its margins since the Late Cretaceous was of the order of 1,000 km.

References

Arnold, H.: Fazies und Mächtigkeit der Kreidestufen im Münsterländer Oberkreidegebiet. Fortschr. Geol. Rheinld. Westf. **7**, 599–610, 1964

Berner, R.A.: Goethite stability and the origin of red beds. Geochim. Cosmochim. Acta **33**, 267–273, 1969

Channell, J.E.T., Lowrie, W., Medizza, F.: Middle and Early Cretaceous magnetic stratigraphy from the Cismon section, Northern Italy. Earth Planet. Sci. Lett. **42**, 153–166, 1979

Channell, J.E.T., Tarling, D.H.: Palaeomagnetism and the rotation of Italy. Earth Planet. Sci. Lett. **25**, 177–188, 1975

Goree, W.S., Fuller, M.: Magnetometers using RF-driven squids and their application in rock magnetism and palaeomagnetism. Rev. Geophys. Space Phys. **14**, 591–608, 1976

Heller, F.: Palaeomagnetism of Upper Jurassic limestones from Southern Germany. J. Geophys. **42**, 475–488, 1977

Heller, F.: Rockmagnetic studies of Upper Jurassic limestones from Southern Germany. J. Geophys. **44**, 525–543, 1978a

Heller, F.: Die magnetischen Eigenschaften der Oberjurakalke Süddeutschlands. Habilitationsschrift ETH Zürich, pp. 75, 1978b

Irving, E.: Drift of the major continental blocks since the Devonian. Nature **270**, 304–309, 1977

Lotze, F.: Die Tektonik der Oberkreidezone bei Lengerich. Geotektonische Forsch. **9/10**, 18–50, 1953

Løvlie, R., Lowrie, W., Jacobs, M.: Magnetic properties and mineralogy of four deep-sea cores. Earth Planet. Sci. Lett. **15**, 157–168, 1971

Lowrie, W., Alvarez, W.: Late Cretaceous geomagnetic polarity sequence: Detailed rock- and palaeomagnetic studies of the Scaglia rossa limestone at Gubbio, Italy. Geophys. J. **51**, 561–582, 1977

Lowrie, W., Channell, J.E.T., Heller, F.: On the credibility of remanent magnetization measurements. Geophys. J. (in press, 1979)

Rosenfeld, U.: Bau und Entwicklung einer Überschiebungszone bei Lengerich (Westf.). Neues Jahrb. Geol. Palaeontol., Abh. **117**, 315–339, 1963

Stadler, G., Teichmüller, R.: Zusammenfassender Überblick über die Entwicklung des Bramscher Massivs und des Niedersächsischen Tektogens. Fortschr. Geol. Rheinld. Westf. **18**, 547–564, 1971

Thiermann, A., Arnold, H.: Die Kreide im Münsterland und in Nordwestfalen. Fortschr. Geol. Rheinld. Westf. **7**, 599–610, 1964

VandenBerg, J., Wonders, A.A.H.: Palaeomagnetic evidence of large fault displacement around the Po basin. Tectonophysics **33**, 301–320, 1976

Van der Voo, R.: Palaeomagnetic evidence for the rotation of the Iberian peninsula. Tectonophysics **7**, 5–56, 1969

Van der Voo, R., French, R.B.: Apparent polar wandering for the Atlantic bordering continents: Late Carboniferous to Eocene. Earth Sci. Rev. **10**, 99–119, 1974

Watson, G.S.: A test for randomness of directions. Mon. Not. R. Astron. Soc. Geophys. Suppl. **7**, 160–161, 1956

Received October 1, 1979; Revised Version November 22, 1979

J. Geophys. 46, 429–450, 1979

Journal of Geophysics

A Two-Dimensional Magnetometer Array for Ground-Based Observations of Auroral Zone Electric Currents During the International Magnetospheric Study (IMS)

F. Küppers, J. Untiedt, W. Baumjohann, K. Lange, and A.G. Jones

Institut für Geophysik der Universität Münster,
Gievenbecker Weg 61, D-4400 Münster, Federal Republic of Germany

Abstract. A two-dimensional magnetometer array was progressively installed in Scandinavia, from 1974 onwards, for operation during the IMS period (1977–1979). The 36 instruments, which are buried in the ground, are of the Gough-Reitzel type, i.e., classical magnetometers, with wire-suspended magnets and optical recording. The time-variation period range observable is from about 50 s to several days. In northern Scandinavia, the spacing between the stations is about 100–150 km in both the north-south and east-west directions.

For presentation and analysis of the data a special Cartesian coordinate system has been introduced (the 'Kiruna system'). It is derived by mapping the local earth's surface onto a tangential plane, with its origin close to Kiruna (geographic coordinates 67.8 N, 20.5 E).

First results from some stations show that internal contributions to the recorded horizontal geomagnetic variations are small and possibly negligible at lower frequencies. However, at frequencies above about 2 mHz the variations of the vertical component display a strong amplification near the coast, and indicate the existence of conductivity anomalies at some inland locations.

In order to demonstrate the observational capabilities of the array, equivalent overhead current configurations are presented which characterize a substorm recorded on October 7, 1976.

Key words: Magnetometer arrays – Auroral zone – Electric current systems – International Magnetospheric Study.

1. Introduction

Since the pioneering work of Birkeland (1908, 1913) geophysicists have become increasingly aware that temporal variations of the geomagnetic field within and near to the auroral zone generally display a very inhomogeneous and often rather complex spatial structure. In order to study these variations in

0340-062X/79/0046/0429/$04.40

greater detail, there has been a tendency to place more and more permanent or temporary magnetic observatories at high latitudes. This was true especially during the Second International Polar Year 1932–1933 and the International Geophysical Year 1957–1958. By such activities it became possible to recognize the basic larger-scale features of high-latitude magnetic disturbances (e.g., Chapman and Bartels 1940; Fukushima 1953; Akasofu 1968; Rostoker 1972).

A further important step made within the last decade was the installation of permanent or temporary north-south chains of densely spaced magnetometers. Before the IMS (International Magnetospheric Study 1977–1979), such meridian chains were installed in Alaska (Akasofu et al. 1971), Canada (Kisabeth and Rostoker 1971; Chen and Rostoker 1974), Greenland (Wilhjelm and Friis-Christensen 1976; Wilhjelm et al. 1978), north-eastern Scandinavia (Maurer and Theile 1978), and along two geomagnetic meridians in the northern part of the Soviet Union (Loginov et al. 1978). Whereas the Greenland chain led to a considerable improvement in our knowledge of polar cap magnetic variations, results from the Canadian meridian chain have been especially important in revealing the detailed temporal and spatial behaviour of auroral electrojets (e.g., Kisabeth and Rostoker 1971; Rostoker and Kisabeth 1973; Kisabeth and Rostoker 1974; Wiens and Rostoker 1975; Rostoker and Hron 1975; Hughes and Rostoker 1977). Data from the Alaska chain especially were used to investigate the relationship between auroral electrojets and field-aligned currents (Yasuhara et al. 1975; Kamide and Akasofu 1976; Kamide et al. 1976; Kamide and Rostoker 1977).

Meridian chains of magnetometers are naturally inadequate if it is intended to study magnetic disturbances whose spatial variation in the east-west direction is as large as the meridional variation. In certain cases, this disadvantage may be circumvented by assuming that the observed local time behaviour reflects the unobserved dependency on east-west coordinates. However, to date little is known about short-lived magnetic disturbances which occur during a magnetic substorm and which are highly localized in both horizontal directions. For example, such disturbances may be related to auroral events such as the westward traveling surge, the midnight break-up region, and spirals (Kisabeth and Rostoker 1973; Untiedt et al. 1978). The near-midnight sector of the Harang discontinuity (Heppner 1972), i.e., the region near the eastward end of the afternoon-evening sector eastward electrojet, may also be considered as another important example within this context. A recent publication by Kawasaki and Rostoker (1979) should also be mentioned. These authors present first results derived from a magnetometer array which consists of an east-west and a north-south chain of 7 stations altogether. Their observations particularly include regions of large magnetic perturbations associated with eastward drifting auroral structures.

In this paper, we will report on a two-dimensional array of 32 (later 36) magnetometers which has been installed within and near the Scandinavian auroral zone for ground-based observations of near-earth electric currents throughout the International Magnetospheric Study (IMS). In July–September, 1974, a similar array was operated in Canada, at comparable geomagnetic latitudes, by Bannister and Gough (1977, 1978; see also Gough and Bannister 1978). However,

we believe that the Scandinavian Magnetometer Array is of special value because its supporting role in the IMS offers unique opportunities for cooperation and multi-method studies. Furthermore, it is important that this array is operated throughout the winter when simultaneous optical observations of the aurora within the same area are possible.

2. The Magnetometers and Their Arrangement

The instrument on which our array is based is of the Gough-Reitzel type (Gough and Reitzel 1967), with important modifications as partly described by Küppers and Post (1979). As in the case of classical magnetometers, the deflections of three wire-suspended magnets are optically recorded on 35 mm film. The exposure is intermittent. Its sequence has a 10-s period (except for our southernmost stations where it is 20 s), controlled by a quartz clock. To avoid aliasing effects, the magnets are partly surrounded by copper blocks which provide for strong electromagnetic damping of higher frequency oscillations. At a period of 100 s the amplitude attenuation is already negligible whereas phase distortion is of the order of 10 degrees (Küppers and Post 1979). The whole instrument, including camera, battery, and electronic control device, is housed within a 1.6-m-long airtight aluminium tube, which is buried in the ground. With the use of both a 10 s exposure cycle and Kodak RAR 2498 film, the magnetometer operates unattended for approximately 73 days. After that time, servicing is implemented, including (i) changing the rechargeable 7.5 Ah battery, (ii) changing the camera (with film), (iii) examination of the automatic calibration currents, (iv) measurement of the clock error, (v) readjustment of the quartz oscillator frequency (if necessary), (vi) resetting the clock (also only if necessary), and (vii) examination of the light traces of the three magnetic field components (position and mobility). On average, the observed clock errors are less than 5 s. Even if the assumption that the clock error is a linear function of time between two subsequent services may be rather simplistic, a time accuracy to within a few seconds can be assumed for all cases.

In general the scale values of the magnetometers are of the order of 40 nT mm^{-1} on film and are known within 0.5%. The optical resolution of the recorded traces is about 0.05 mm. The scale value and the optical resolution give a magnetic resolution of about 2 nT.

Every sixth hour (or twelfth for 20 s cycle instruments) during operation, calibration fields are applied automatically to the three components by means of small Helmholtz coils surrounding each magnet. For all three components, the range of the observable magnetic field variations is increased by adding a parallel auxiliary trace from a virtual second light source (obtained by mirror splitting). For typical instruments, the total range is about $-1{,}500$ nT to $+1{,}000$ nT for H, -500 nT to $+1{,}000$ nT for D, and -800 nT to $+1{,}200$ nT for Z, with reference to the undisturbed levels. Figure 1 illustrates a short section of film record which includes a calibration and the appearance of two auxiliary traces.

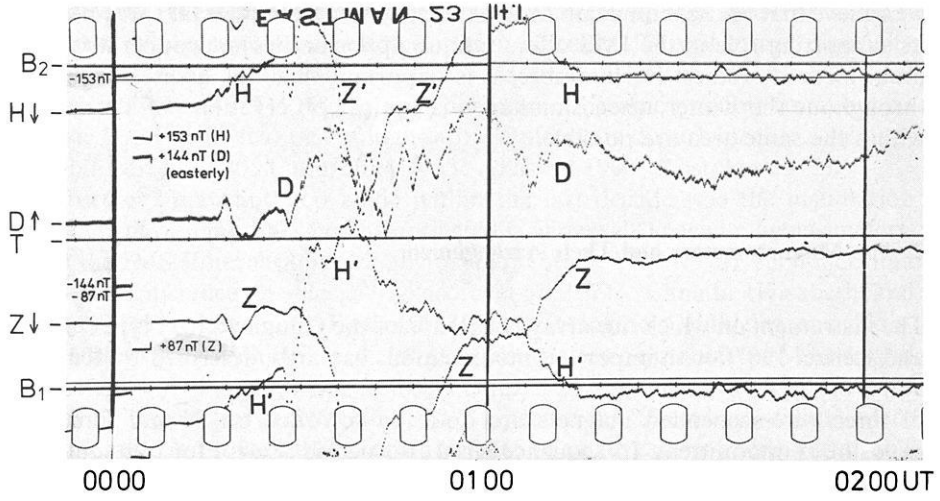


Fig. 1. Example of a record, from July 29, 1977, and station EVE (cf. Fig. 2). *H, D. Z* main traces; *H', Z'* auxiliary traces; B_1, B_2 base lines; *T* temperature trace. Note that the 0000 UT time mark is a little stronger (due to longer exposure) than other hour marks. Every sixth minute the recording points are also stronger than normal. Between 0000 UT and 0006 UT calibration fields (cf. indicated values) have been applied

At the north German magnetic observatory of Wingst the three recording magnets of each magnetometer were aligned to their correct orientations with reference to magnetic north and to vertical, and the calibration coil constant (in nT mA^{-1}) for each component of each instrument was determined. At the recording site, the instrument is levelled and rotated around its vertical axis until the *D* component light trace is at the correct position. In this orientation, the variations of the local magnetic north (*H*), east (*D*), and vertical (*Z*) components are recorded. In order to convert the measured local field variations to the geographically oriented system, the local declination has to be known. This quantity (cf. Table 1 below) was measured as often as possible at the location of each magnetometer by means of a compass theodolite. To avoid errors due to a large difference between the static magnetic field at the magnetometer and the field at the theodolite, stations were installed with preference to locations where the static magnetic field gradient was less than 1 nT m^{-1} (as indicated from observations with a torsion balance).

Burying the instruments in the ground protects them from the very low temperatures in winter time. During operation, the temperatures within our instruments have never fallen below $-5°$ C. This avoids corresponding malfunction of cameras, batteries, and electronics. Furthermore, it is important that due to the very effective shielding of the ground, daily temperature variations are not able to influence our measured data in any appreciable way (Küppers and Post 1979). In order to facilitate an analysis of daily or longer variations, the majority of the instruments contains a bimetallic strip temperature sensor, the variation of which is also recorded on the film (cf. Fig. 1).

Table 1. List of stations from the Scandinavian Magnetometer Array of the University of Münster

Symbol	Name	Country	Geographic		Rev. Corr. Geom.		x_{KI} (km)	y_{KI} (km)	γ_{KI} (deg)	D (deg)	Installation
			Lat.	Long.	Lat.	Long.					
MAL	Malöy	N	62.18	5.10	60.67	88.99	−350	−887	26.2	−6.5	76/9
HEL	Hellvik	N	58.52	5.77	56.75	87.44	−735	−1033	25.5	−3.9	76/9
KLI	Klim	DK	57.12	9.17	54.96	89.37	−961	−911	22.4	−6.0	76/10
NAM	Namsos	N	64.45	11.13	62.24	94.95	−239	−509	20.8	−2.9	76/9
FLO	Flötningen	S	61.88	12.23	59.55	94.13	−527	−557	19.8	−1.0	78/7
ARV	Arvika	S	59.60	12.60	57.19	93.18	−774	−623	19.4	−1.9	76/9
ESM	Esmared	S	56.74	13.22	54.16	92.35	−1089	−694	18.7	−1.5	78/7
FRE	Fredvang	N	68.08	13.17	65.67	99.20	108	−285	19.0	−0.5	76/9
GLO	Glomfjord	N	66.90	13.58	64.44	98.55	− 20	−311	18.6	−1.2	75/10
OKS	Okstindan	N	65.90	14.27	63.37	98.29	−136	−317	18.0	−0.4	76/9
RIS	Risede	S	64.50	15.13	61.89	97.95	−296	−326	17.2	−0.5	76/9
HAS	Hassela	S	62.07	16.50	59.35	97.45	−575	−338	15.9	0.5	76/9
AND	Andenes	N	69.30	16.02	66.62	102.38	202	−134	16.4	−0.1	76/9
EVE	Evenes	N	68.53	16.77	65.79	102.22	112	−129	15.7	0.4	76/9
RIJ	Ritsemjokk	S	67.70	17.50	64.90	102.06	15	−124	15.0	−1.1	76/9
KVI	Kvikkjokk	S	66.90	17.92	64.07	101.72	− 75	−130	14.6	4.0	76/9
SRV	Storavan	S	65.78	18.18	62.93	101.04	−198	−150	14.4	1.7	76/9
LYC	Lycksele	S	64.57	18.68	61.67	100.55	−335	−160	13.9	1.3	76/9
MIK	Mikkelvik	N	70.07	19.03	67.14	105.12	255	0	13.6	0.7	74/8
ROS	Rostadalen	N	68.97	19.67	65.99	104.64	130	− 4	13.0	2.7	74/8
KIR	Kiruna	S	67.83	20.42	64.80	104.27	0	− 2	12.3	0.0	74/12
NAT	Nattavaara	S	66.75	21.00	63.67	103.82	−122	− 3	11.8	3.0	74/8
PIT	Pitea	S	65.25	21.58	62.16	103.22	−291	− 10	11.3	5.1	74/8
HOP	Hööpaka	SF	63.01	22.56	59.86	102.60	−545	− 11	10.4	3.0	78/7
SOY	Söröya	N	70.60	22.22	67.39	107.94	287	129	10.6	4.4	75/10
MAT	Mattisdalen	N	69.85	22.92	66.62	107.77	200	139	10.0	4.5	75/10
MIE	Mieron	N	69.12	23.27	65.86	107.40	118	139	9.7	4.7	75/10
MUO	Muonio	SF	68.03	23.57	64.75	106.70	− 2	131	9.4	4.7	75/10
PEL	Pello	SF	66.85	24.73	63.47	106.59	−140	159	8.3	5.5	75/10
OUL	Oulu	SF	65.10	25.48	61.77	106.07	−337	166	7.7	5.1	75/10
JOK	Jokikylä	SF	63.77	26.13	60.39	105.73	−488	177	7.1	7.5	75/10
SAU	Sauvamäki	SF	62.30	26.65	58.82	105.32	−654	184	6.7	5.1	75/10
BER	Berlevag	N	70.85	29.13	67.16	113.10	282	384	4.2	8.2	76/9
VAD	Vadsö	N	70.10	29.65	66.42	112.75	197	397	3.7	7.1	76/9
SKO	Skogfoss	N	69.37	29.42	65.70	111.95	117	383	3.9	8.8	76/9
RKS	Roksä	SF	62.57	30.26	58.95	108.29	−640	372	3.4	6.9	78/7

Country: DK Denmark, N Norway, S Sweden, SF Finland

Revised corrected geomagnetic coordinates according to Gustafsson (1970). x_{KI}, y_{KI} coordinates in the Kiruna system (cf. Fig. 2 and text)

γ_{KI} westerly deviation of the local parallel to the positive x_{KI} axis from geographic north

D measured or adopted (from published data, magnetic maps etc.) value of magnetic easterly declination

The last column gives year and month of installation. Most stations are scheduled for operation until spring 1980

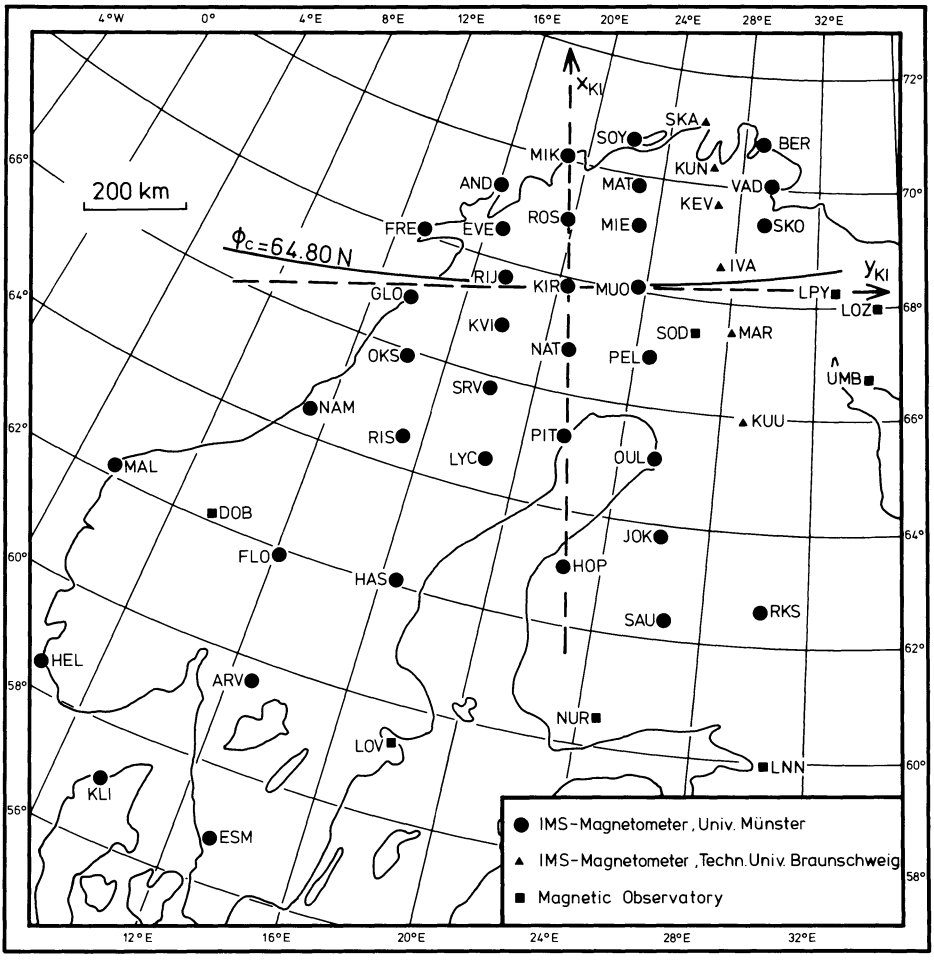


Fig. 2. Station map in geographic coordinates. Also indicated are the axes of the Kiruna system (cf. Sect. 3), and the line of constant revised corrected geomagnetic latitude $\Phi_c = 64.80$ N (after Gustafsson 1970)

Usually magnetic variations at high latitudes are studied with reference to a quiet time level. Because a large time interval of many days may exist between the magnetic event which is of interest (e.g., a specific substorm) and the closest quiet day, errors may be induced by magnetometer drifts. For most of our instruments, these drifts are caused mainly by annual temperature variations and are less than 0.1 nT/day (B. Inhester, personal communication). Occasionally, our Z components show strong anelastic behaviour. However, even in these cases the drifts are rarely larger than 1 nT/day and may easily be recognized and eliminated during data analysis by comparing the quiet time levels recorded by adjacent instruments.

The spatial distribution of our magnetometers is shown in Fig. 2, and their geographical coordinates, with other relevant information, are given in Table 1.

Most of the instruments are approximately situated along straight lines perpendicular to the lines of constant revised corrected geomagnetic latitude, as given by Gustafsson (1970). It was considered that such an arrangement would ease the analysis of the data since it could be anticipated that electrojets flowing approximately parallel to the auroral oval would predominate. In northern Scandinavia (down to LYC and OUL, cf. Fig. 2), which is situated under the statistical auroral oval around magnetic midnight, and where accordingly the most inhomogeneous magnetic field variations could be expected to occur, the spacing of our stations along the north-south lines is least and is 120 km on average. Such a separation between stations was chosen with regard to the fact that external magnetic fields with horizontal wavelengths of less than 200 km are appreciably depressed (according to the factor e^{-kh}, where $k=2\pi\lambda^{-1}$, λ is the wavelength, and $h\sim 100$ km) between the ionosphere and the ground. The results from the University of Alberta meridian chain (e.g., Kisabeth and Rostoker 1971; Rostoker and Kisabeth 1973) suggested a similar spacing. Because it was anticipated that in interesting cases east-west gradients might be as important as north-south gradients, the distance between adjacent north-south profiles was also chosen to be not much greater than 120 km. Note that this spacing is not latitude invariant because of the convergence of the profiles toward north. At the geomagnetic latitude of stations GLO-MUO (Fig. 2) this longitudinal spacing is on average 150 km.

In north-eastern Scandinavia our array is supplemented by a north-south chain of 6 digital three-component fluxgate magnetometers operated during the IMS by the Technical University of Braunschweig (Maurer and Theile 1978). The remainder of our instruments are distributed over the southern half of Scandinavia with a much wider separation between stations. Together with the existing permanent magnetic observatories, these southern stations are valuable for studying, e.g., apparent return currents of auroral electrojets, far-field effects of field-aligned currents, and electrojets flowing during phases of strong auroral oval expansion. To the north-east, it is important that the Polar Geophysical Institute at Apatity (USSR) is operating several magnetometer stations on the Kola peninsula (Fig. 2).

Our array ends abruptly at the northern coast of Scandinavia (revised corrected geomagnetic latitude 67.4 N) and thus the two Norwegian permanent magnetic observatories on Bear Island (BJN) and on Spitsbergen (NAL) are invaluable for our studies. BJN is approximately situated on our station line SAU-SOY, about 450 km to the north of SOY.

The array, which is scheduled to be in operation until spring 1980, was installed in steps, with the first north-south chain (MIK-PIT) recording from autumn 1974, and with the last stations (FLO, ESM, HOP, and RKS) added in July 1978 (cf. Table 1).

3. The Kiruna System of Coordinates and Magnetic Field Components

The main purpose of the present array is detection and analysis of local magnetic disturbance fields. Any analysis of the data by methods based on potential

theory (e.g., separation into internal and external parts, or field continuation towards the source, Mersmann et al. 1979) is greatly eased if Cartesian coordinates and corresponding field components may be used. Such a *local* system, which is applicable only to data from the Scandinavian region, will now be described and will be called the 'Kiruna system'.

When defining this system, our basic postulates have been that the spherical earth's surface of the Scandinavian region should be projected onto a tangential plane by conformal mapping, and that the magnetic potential at a projected point should be equal to the potential before projection. As the common point of the sphere and of the tangential plane, and as the origin of the new coordinate system, we have chosen the station Kiruna (KIR, cf. Fig. 2), i.e., a point approximately in the middle of our array. The horizontal axes of the Kiruna system have been defined to be related to a line of constant revised corrected geomagnetic latitude Φ_c (Gustafsson 1970) in approximately the manner in which the geographic north and east directions are related to circles of geographic latitude.

The transformation of an arbitrary point P given by geographic colatitude θ and longitude λ and of its northward and eastward magnetic disturbance field components X and Y to the Cartesian coordinates x_{KI} and y_{KI} and to the corresponding field components A and B, respectively, is undertaken via the following steps:

(a) Instead of the geographic pole, KIR is adopted to be the pole of a new spherical system. Accordingly, the spherical coordinates θ_{KI} and λ_{KI} will be attributed to P. The new magnetic horizontal components are X_{KI} (pointing towards 'north' of the new system, i.e., towards KIR) and Y_{KI} (towards 'east'). They are the result of a rotation around the vertical axis at P. The new meridian $\lambda_{KI}=0$ is defined in such a way that it is perpendicular to the line $\phi_c=64.80$ N which passes through KIR (Gustafsson 1970), and that it turns approximately towards *south* from KIR.

(b) P is now transformed to P', a point on the tangential plane with the desired Cartesian coordinates x_{KI} and y_{KI}, by means of a stereographic projection (cf. Fig. 3a and b), according to

$$x_{KI}=-2R_E \tan(\theta_{KI}/2) \cos \lambda_{KI}$$
$$y_{KI}=2R_E \tan(\theta_{KI}/2) \sin \lambda_{KI}$$

where $R_E=6{,}371$ km is the adopted value of the earth's radius.

Simultaneously, the horizontal magnetic components are transformed according to

$$X'_{KI}=\cos^2(\theta_{KI}/2)\, X_{KI}$$
$$Y'_{KI}=\cos^2(\theta_{KI}/2)\, Y_{KI}$$

where the X'_{KI} and Y'_{KI} field components lie within the tangential plane (see Fig. 3a), and the X'_{KI} component points towards KIR. The factor $\cos^2(\theta_{KI}/2)$ provides for constancy of the magnetic potential (see above).

(c) By a rotation within the tangential plane (Fig. 3b) at the point P', we obtain the final horizontal magnetic field components

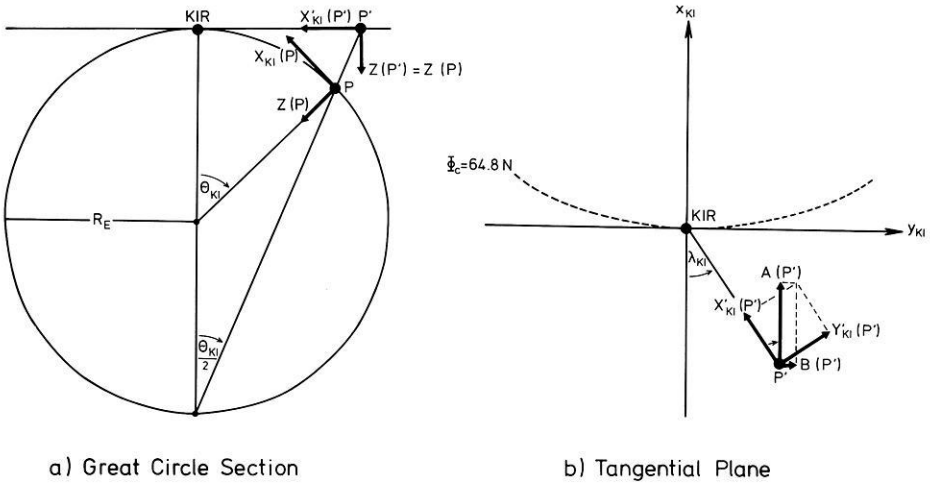


Fig. 3. Diagrams illustrating the derivation of the Kiruna system

$$A = X'_{KI} \cos \lambda_{KI} + Y'_{KI} \sin \lambda_{KI}$$
$$B = -X'_{KI} \sin \lambda_{KI} + Y'_{KI} \cos \lambda_{KI}$$

which are directed parallel to the x_{KI} and y_{KI} axes, respectively.

The vertical component Z remains unchanged throughout. Because we intend to apply this system only to data from Scandinavia, the factors representing spatial distortion or field reduction within the above equations may be assumed to be close to unity. For a station in southern Scandinavia, θ_{KI} may be 8°, for example. In this case, the distortion or reduction factors are, respectively, (for the first factor, cf. Fig. 3a)

$$2 \tan (\theta_{KI}/2)/\theta_{KI} = 1.0016, \quad \cos^2 (\theta_{KI}/2) = 0.9951,$$

if θ_{KI} has been transformed into radians. Hence, only the rotations involved within the different steps of the transformation are important for our purposes.

For the horizontal magnetic disturbance field, the sequence of the above rotations is described completely by the angle γ_{KI}, which gives the westward deviation of the local A component direction (parallel to the x_{KI} axis, cf. Fig. 2) from geographic north. Together with the x_{KI} and y_{KI} coordinates, this angle is given in Table 1 for our array stations, and in Table 2 for other magnetic stations within and near the Scandinavian region.

4. Internal Contributions From Induced Currents

It would require much consideration if the magnetic variations recorded in Scandinavia were appreciably influenced by the magnetic fields from electric currents induced in the earth's crust and upper mantle, especially if conductivity anomalies were present (Gough 1974; see also discussion at the occasion of

Table 2. List of other magnetic stations in the Scandinavian region. For explanations, see Table 1

Symbol	Name	Country	Geographic		Rev. Corr. Geom.		x_{KI} (km)	y_{KI} (km)	γ_{KI} (deg)	Type of Station
			Lat.	Long.	Lat.	Long.				
DOB	Dombas	N	62.07	9.12	60.07	91.89	−448	−701	22.6	PMO
LOV	Lovö	S	59.35	17.83	56.43	97.12	−888	−348	14.7	PMO
TRO	Tromsö	N	69.67	18.95	66.75	104.77	212	− 13	13.7	PMO
NUR	Nurmijärvi	SF	60.50	24.65	57.05	102.88	−839	51	8.5	PMO
NAL	Ny Alesund	N	78.92	11.93	75.89	114.65	1260	87	20.5	PMO
BJN	Bear Island	N	74.52	19.02	71.34	110.34	736	116	13.7	PMO
SOD	Sodankylä	SF	67.37	26.63	63.90	108.47	− 93	248	6.6	PMO
SKA	Skarsvag	N	71.12	25.83	67.61	111.04	324	268	7.3	TMS (TUB)
KUN	Kunes	N	70.35	26.52	66.84	110.79	236	282	6.6	TMS (TUB)
KEV	Kevo	SF	69.75	27.03	66.23	110.61	167	294	6.2	TMS (TUB)
IVA	Ivalo	SF	68.60	27.47	65.05	109.94	38	298	5.8	TMS (TUB)
MAR	Martti	SF	67.47	28.28	63.90	109.76	− 89	320	5.1	TMS (TUB)
KUU	Kuusamo	SF	65.92	29.05	62.37	109.24	−264	339	4.4	TMS (TUB)
LNN	Leningrad	USSR	59.95	30.70	56.22	107.44	−934	379	3.1	PMO
RYB	Rybachy	USSR	69.90	31.90	66.09	114.23	171	482	1.6	TMS (PGIA)
LPY	Loparskaya	USSR	68.25	33.08	64.37	113.87	− 13	525	0.6	PMO
LOZ	Lovozero	USSR	67.98	35.02	64.08	115.10	− 42	606	−1.1	PMO
UMB	Umba	USSR	66.70	34.50	62.88	113.82	−185	586	−0.6	TMS (PGIA)

Type of station: PMO permanent magnetic observatory. TMS temporary (IMS) magnetic station, from Technical University of Braunschweig (TUB) or Polar Geophysical Institute at Apatity (PGIA)

the planning of North-American IMS magnetometer stations by Camfield and Gough 1976 and Lanzerotti and Sugiura 1976). Such influences would hinder a rather direct interpretation of observed magnetic fields in terms of ionospheric and near-earth magnetospheric currents, and would necessitate a field separation into internal and external parts for every study. Because of the large spacing (100–150 km) between stations, such a separation of the observed magnetic variations would not give reliable results if conductivity anomalies of a smaller scale were of importance. Fortunately, it appears from initial analyses of our data that internal contributions to the observed magnetic disturbance fields are small, or at least rather homogeneous, except for higher frequency variations near the coast. To a first approximation, the internal field may be neglected for the horizontal magnetic components. Preliminary arguments will be given in this section. A thorough analysis including an investigation of the electrical conductivity structure under Scandinavia will need appreciable effort, including magnetotelluric measurements, and is beyond the scope of this paper.

The first technique employed by us to obtain information on the importance of internal contributions to the observed magnetic fields was a two-dimensional field separation (Kertz 1954; Siebert and Kertz 1957; the formalism may also be found in Mersmann et al. 1979). Such a separation is only possible if, at the earth's surface, a field is observed which is a function of the vertical coordinate z and one horizontal coordinate (e.g. $\tilde{x}$) only. The corresponding magnetic components may be Z and $\tilde{A}$. From $\partial \tilde{A}/\partial \tilde{y}=0$, together with Maxwell's first

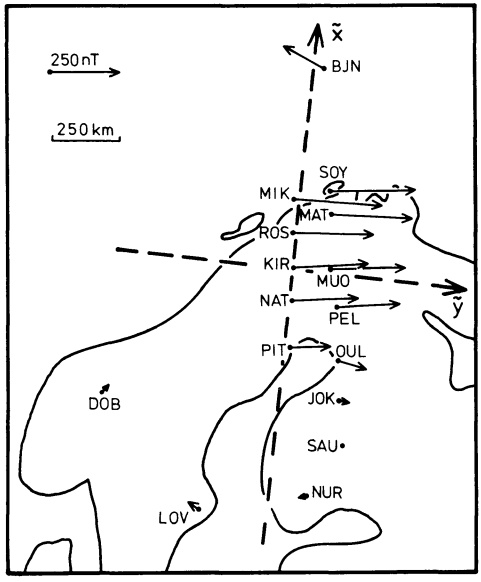


Fig. 4. Equivalent overhead current vectors (at height zero, in nT) on April 7, 1976, 1336 UT. The rotated system $\tilde{x}$, $\tilde{y}$ (the corresponding magnetic components are $\tilde{A}$, $\tilde{B}$) mentioned in the text is also shown

equation and the absence of any current flowing immediately above the earth's surface, it follows that $\partial\tilde{B}/\partial\tilde{x}=0$, i.e., $\tilde{B}=$constant along an $\tilde{x}$ profile, where $\tilde{y}$ and $\tilde{B}$ denote the remaining Cartesian coordinate or component, respectively. It should be noted that the formalism given by Kertz (1954) and Siebert and Kertz (1957) renders no physically significant separation of any quasihomogeneous part of $\tilde{A}$ or Z, since this part is formally bisected upon separation.

A local, approximately two-dimensional disturbance field may be found at certain stages of an individual substorm. A typical example is illustrated in Fig. 4 which shows the distribution of equivalent overhead current vectors (horizontal magnetic disturbance vectors turned 90 degrees clockwise) at 1336 UT on April 7, 1976, when two magnetometer chains, later forming a part of our array, were already operating. The disturbance field components have been defined as deviations from the 12–15 UT field recorded on March 22, 1976, which was the very quiet day closest to the day under consideration. The digitized magnetic data from each station were low-pass filtered, with a cut-off frequency of 4 mHz, in order to avoid errors introduced by non-two-dimensional disturbance fields in the pulsation frequency band (see below). The pattern described by Fig. 4 changes rather slowly within some tens of minutes. It shows an eastward electrojet above northern Scandinavia, but deviations from two-dimensionality are indicated to the south (NUR-OUL) and to the north (BJN). The $\tilde{x}$ axis, which is delineated in Fig. 4, was determined by minimizing the average value of $(\partial\tilde{B}/\partial\tilde{x})^2$ along the two chains of stations.

For these two chains, the observed $\tilde{x}/\tilde{A}$ and Z magnetic field components are given in Fig. 5, together with their internal parts as calculated by two-dimensional field separation. Although the applied interpolation and extrapolation of the observed field values may be somewhat unrealistic, especially to

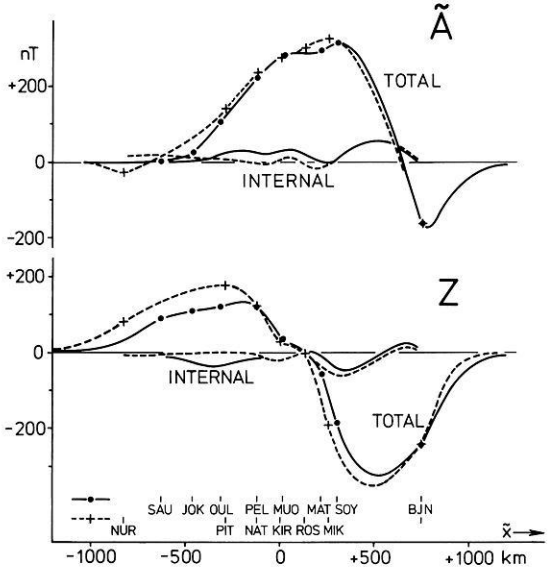


Fig. 5. Low-pass filtered magnetic disturbance components $\tilde{A}$ and Z along two approximately parallel north-south station chains (cf. Fig. 4), and the corresponding internal parts after separation, on April 7, 1976, 1336 UT. For $\tilde{A}$, see caption to Fig. 4; Z vertical component (positive downward). Observed values indicated (*crosses*: station chain NUR-MIK-BJN; *dots:* station chain SAU-SOY-BJN; cf. bottom of figure)

the far north where data were available only from one station (BJN), and although the two chains give slightly different results, it is apparent that there are no important internal disturbance field contributions present at lower frequencies within this region of Scandinavia. Note that another form of extrapolation, which would correspond to a broader westward electrojet to the north, would yield a nearly constant internal part of the order of not more than a few tens of nanoteslas, for both $\tilde{A}$ and Z, over the area of interest. The result from the analysis of another field disturbance recorded by the same two chains of stations, which supports the present conclusion, may be found in a recent paper by Mersmann et al. (1979).

Preliminary analyses of data, recorded by the whole array, utilizing the more usual Geomagnetic Depth Sounding (GDS) techniques (for a description of some of these methods see, for example, Lilley 1975), additionally showed that low frequency magnetic variations over the whole of northern Scandinavia do not indicate the presence of strong anomalies of internal conductivity structure. Also inspection of equivalent current vector diagrams derived from various magnetic events reveals no recognizable distortion of the vector patterns due to internal contributions. Hence, we are confident that at frequencies below about 2 mHz the measured horizontal magnetic fields in northern Scandinavia are not grossly perturbed by internal currents.

Unlike this slowly varying part of magnetic disturbances, the Z component of higher frequency fluctuations exhibits large internal contributions, especially

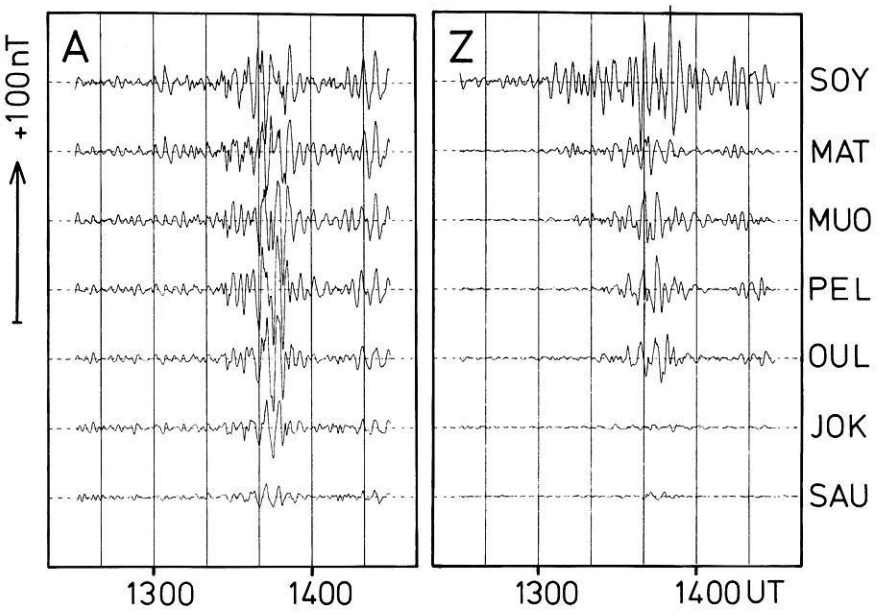


Fig. 6. High-pass filtered (≥ 4 mHz) magnetic variations on April 7, 1976, along the eastern chain of the two station chains shown in Fig. 4. *A* denotes the component along the x_{KI}-axis (cf. Fig. 2) of the Kiruna system; *Z* vertical component

near the coast. This feature is demonstrated by Fig. 6, which gives for the previously mentioned substorm of April 7, 1976, the high-pass filtered (cut-off frequency 4 mHz) *A* and *Z* variations for the chain of stations SAU-SOY (cf. Fig. 2 or 4). The *B* component is not presented here because qualitatively it behaves similarly to the *A* component shown. It is apparent that the rapid *Z* fluctuations are amplified appreciably at the coastal station of SOY. The presence of a strong coast effect within the *Z* component in northern Scandinavia has been recognized for some time by workers who have studied geomagnetic pulsations within this region (O. Hillebrand and E. Steveling, Göttingen, personal communication).

In addition, Fig. 6 shows the remarkable feature that the amplitudes of the *Z* variations are appreciably attenuated to the south of OUL. Because no corresponding drastic effect is seen in the *A* and *B* components, and because we have found a similar attenuation of *Z* high frequency activity within the same region in another case studied, we conclude that at some depth (10–100 km) within the crust or upper mantle the electrical conductivity below southern Finland is larger than that to the north (Schmucker 1973).

During the break-up phase of an auroral substorm (Akasofu 1968) near magnetic midnight at high latitudes large amplitude magnetic disturbance fields seem to be 'switched on' within a few minutes (e.g., Untiedt et al. 1978). It might be expected that such disturbance fields with a large high-frequency content exhibit anomalous amplification of the *Z* component at coastal stations, as geomagnetic pulsations do. Figure 7 illustrates that this indeed is observed.

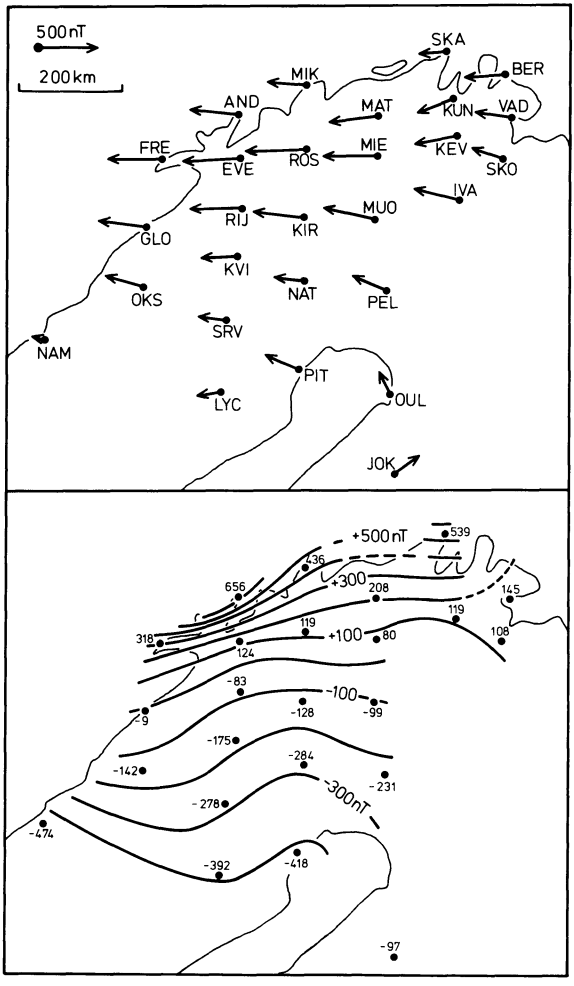


Fig. 7. Differential magnetic disturbances (values at 2147 UT minus values at 2140 UT) on December 2, 1977, from an interval of time which included a strong magnetic onset. *Top:* Equivalent current vectors (cf. Fig. 4). *Bottom: Z* values and isolines. Due to partial malfunction of instruments there are no *Z* values for some stations

This figure is based on the magnetic observations during an intense substorm occurring on the evening of December 2, 1977, when the auroral oval was largely expanded. At 2141 UT (~0010 MLT, according to Whalen 1970) an intense auroral break-up was observed above OUL (H.J. Opgenoorth and R. Pellinen, personal communication). Before that time a westward equivalent current flow corresponding to 200–300 nT was concentrated above southern Finland. Probably in connection with the break-up and the subsequent poleward auroral expansion (Akasofu 1968) which was also observed, a strong (up to 500 nT in the horizontal components) *additional* magnetic disturbance field was 'switched on' between 2140 and 2147 UT for which the equivalent overhead

current vectors are given in Fig. 7 (top). The pattern describes a westward electrojet current flow concentrated above northern Scandinavia and some current loop in the south, possibly related to the auroral break-up (Untiedt et al. 1978). The corresponding distribution of the additional Z disturbances (Fig. 7, bottom) shows a concentration of Z isolines especially near the coastal stations AND and FRE. This concentration, together with the direction of the isolines, may be called anomalous in the sense that it does not correspond to the mentioned relatively uniform westward equivalent current flow within this region (Fig. 7, top). We feel that the anomalously large Z disturbances, which we have also found in another case, especially at AND and FRE, may be due to the fact that the deep ocean border (the 2,000 m depth isoline, say) comes very close to the Norwegian coast only within this region. Otherwise northern Scandinavia is surrounded by shallow ocean water with depths of around 300 m. This may explain why we see a clear coast effect at all coastal stations only at high frequencies.

Although we should like to emphasize that at low frequencies the internal part of magnetic variations is negligible to a first approximation, we do not state that Scandinavia is free from noticeable conductivity anomalies. Studying more data from the magnetometer array, we have, for example, met growing evidence that the phases of magnetic variations are anomalous within the region of OUL, and that there appear to be anomalous high-frequency (~ 5 mHz) Z responses observable in the region of OKS-RIS-SRV (cf. Fig. 2).

5. Magnetic Substorm Development Over Scandinavia on October 7, 1976

In order to demonstrate both the capability and the bounds of the magnetometer array in revealing the small-scale spatial structure of geomagnetic disturbances, this section presents a phenomenological description of several types of disturbance fields recorded during an isolated substorm which occurred over Scandinavia on the evening of October 7, 1976. Within the limits of this paper, we do not intend to present any physical interpretation of the observations. Some physical aspects of this substorm have already been discussed by Baumjohann et al. (1978).

The standard magnetogram from the observatory Kiruna (KIR) (Fig. 8) shows a mainly northward magnetic disturbance field until about 1848 UT. At the beginning, this field varies slowly but after 1834 UT it exhibits an impulsive amplification before a similarly short negative excursion. After 1900 UT another, but longer lasting, negative bay in the X component occurs. More magnetograms for this event may be found in (Baumjohann et al. 1978).

For presentation of our observational results we have chosen the well-known equivalent overhead current presentation at height zero (i.e., horizontal disturbance vectors rotated clockwise by 90 degrees). For brevity, we shall use the word 'current' instead of 'equivalent overhead current at height zero (in nanoteslas)' in the following description. However, we are well aware of the fact that the currents presented may be very different from the true spatial distribu-

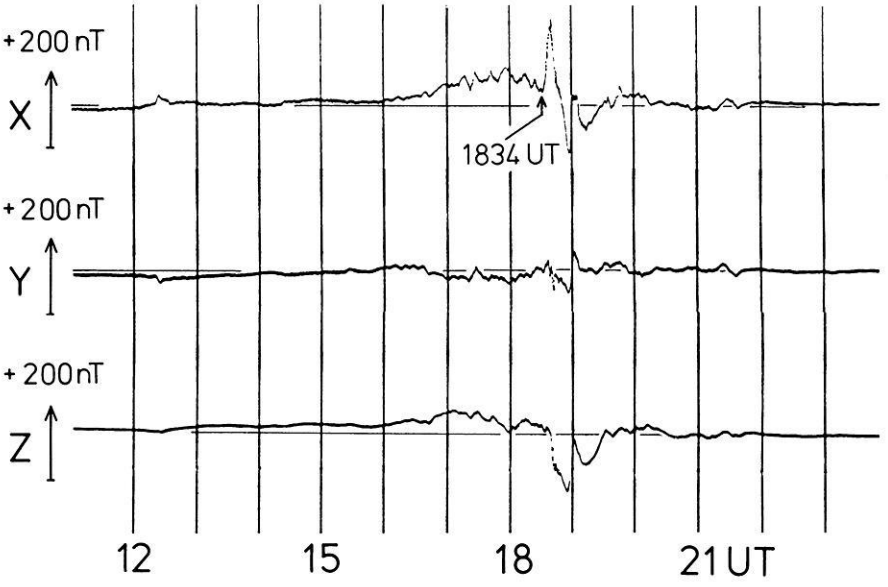


Fig. 8. Standard magnetogram from the geomagnetic observatory at Kiruna for October 7, 1976, 11–24 UT

tion of electric currents because of the appreciable height of the ionospheric main current layer (about 100 km), the possible presence of net field-aligned currents and balanced ionospheric-magnetospheric currents (e.g., Baumjohann et al. 1979) – the latter generating toroidal magnetic fields that are unobservable on the ground –, and the presence of currents induced within the earth.

During the early stages of operating the array, more data was lost than later on, due to initial difficulties in running the instruments. Hence for this event results are lacking from a few of our stations, for example the important coastal stations of FRE, AND, and BER (cf. Fig. 2). On the other hand, it is very important that we were able to include data from BJN, SOD, LOV, NUR, SKA, KUN, KEV, LPY, and LOZ, i.e., from stations that are run by other institutions (see Acknowledgements).

In order to describe the spatial evolution of the substorm over the Scandinavian area, Fig. 9a and b give a sequence of current distributions, each of which characterizes a certain stage of the development of this event up to 1900 UT. The disturbance field components have been defined and measured as deviations from the 7–8 UT quiet time levels on the same day. At KIR magnetic midnight occurred at about 2135 UT (as calculated from Whalen 1970).

Early within the substorm, until approximately 1636 UT, the current flow over Scandinavia was almost uniformly directed changing rather steadily from south-eastward (Fig. 9a, 1604 UT) to eastward, with intensity increasing toward the northeast (as defined in the Kiruna system). Simultaneously, BJN (cf. Fig. 9a, top of diagram given for 1604 UT) showed vector directions quite variable with time, eastward at the beginning and mostly northwestward later.

Afterwards, for more than one hour, the flow over northernmost Scandinavia was almost entirely eastward directed and typical of the eastward electrojet,

but northwestward at BJN. The weaker currents to the south varied several times between southeastward and northeastward (Fig. 9a, 1646 and 1726 UT).

Between 1748 and 1800 UT the northernmost current vectors turned northward (Fig. 9a, 1756 UT), approaching the northwestern current direction at BJN. This branching-off of the northern part of the eastward electrojet later weakened a little (Fig. 9a, 1808 UT), until between 1820 and 1832 UT the currents at our northern stations became rather irregular (Fig. 9a, 1828 UT).

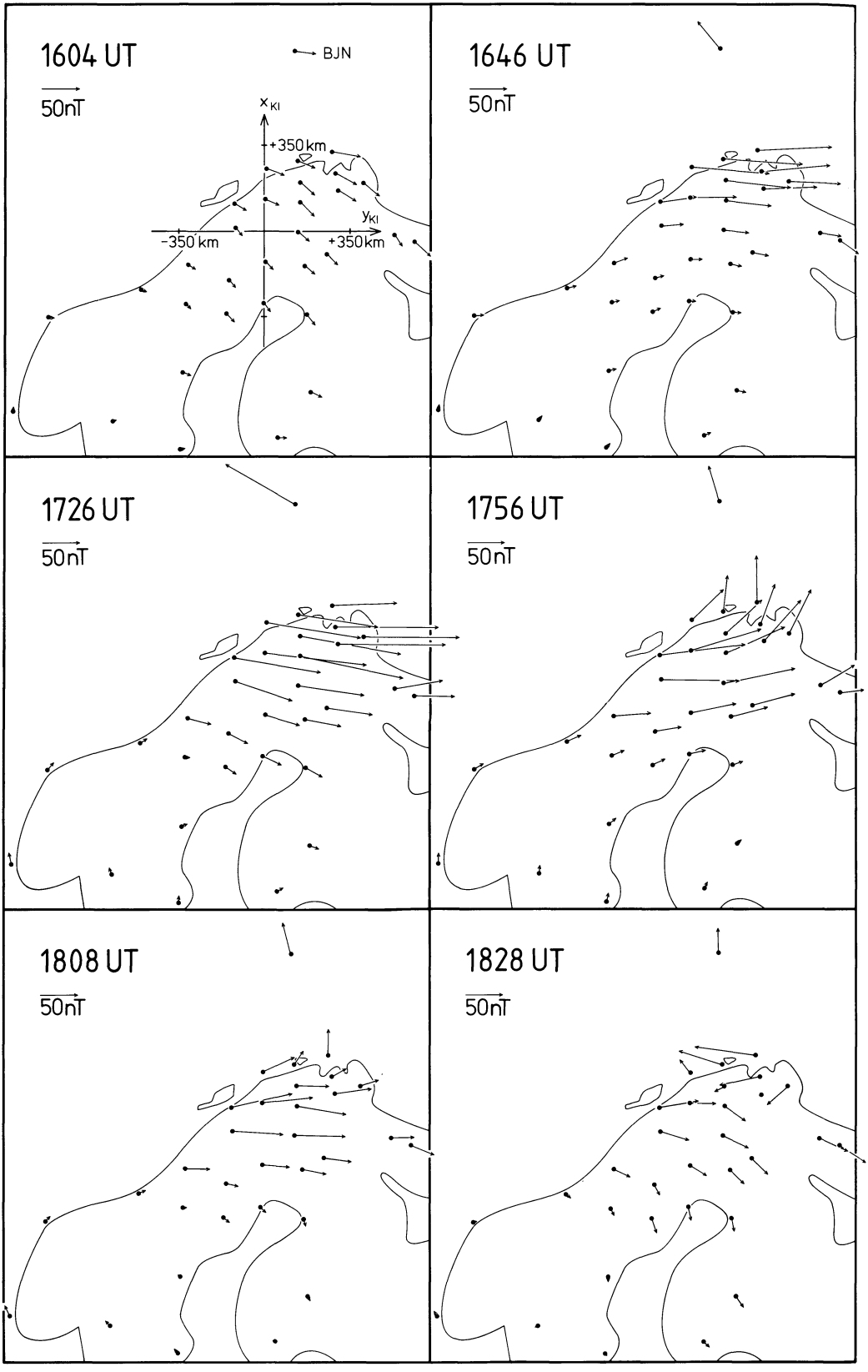
Starting at about 1834 UT (cf. Fig. 8) all the currents intensified appreciably, but their directions remained southeastward over most of Scandinavia and northwestward at BJN. Only within the region of our northern stations did dramatic changes occur. At first, a very local current vortex, somewhat elongated in east-west direction, appeared (Fig. 9b, 1836 UT), probably indicating a concentrated upward flow of field-aligned current (Untiedt et al. 1978). Later a pattern showing a spatial transition from eastward via northward to strong northwestward currents developed within a few minutes (Fig. 9b, 1837 and 1840 UT) which then appeared to travel westward (Fig. 9b, 1848 and 1854 UT) until both BJN and our northernmost stations indicated currents directed to the west (Fig. 9b, 1900 UT). At 1900 UT a remarkable U-like distortion of current was visible just to the south of the westward electrojet. Also within this time interval (Fig. 9b, 1854 UT), there was a more regional, but otherwise similar transition from eastward to northwestward currents visible over the middle of Scandinavia. We believe that our observations between about 1838 and 1854 UT were caused by the passage of the Harang discontinuity (Harang 1946; Heppner 1972; Maynard 1974; Kamide 1978; Nielsen and Greenwald 1979), i.e., the transition region between the eastward and westward electrojets.

6. Conclusions and Summary

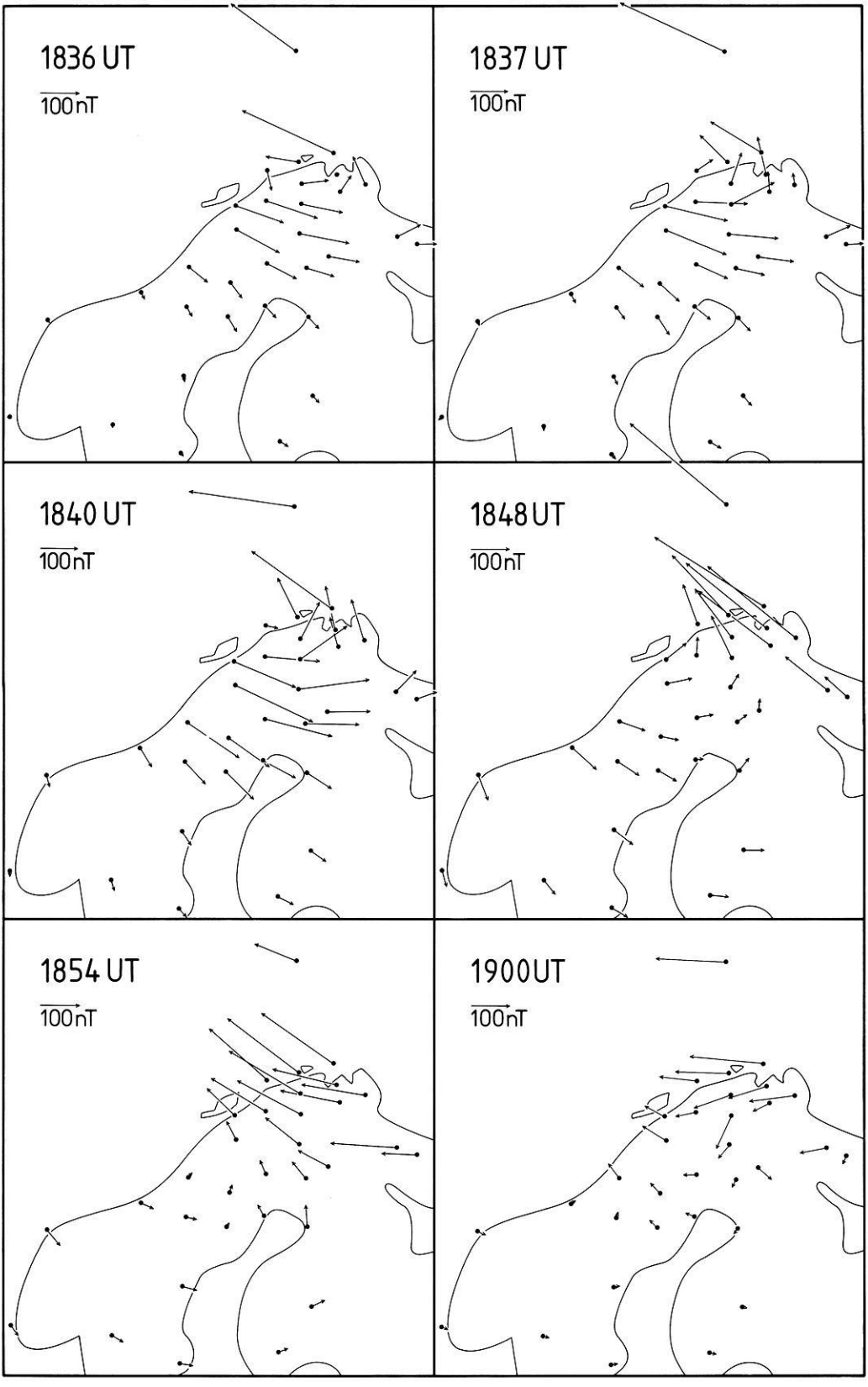
The phenomenological results presented in the preceding section (see also Baumjohann et al. 1978) clearly demonstrate that at many time instances within the course of a substorm the regional magnetic situation within the auroral zone may be adequately observed only by means of a densely spaced two-dimensional magnetometer array. This is particularly true for the O-type (Fig. 9b, 1836 UT), U-type (Fig. 9b, 1840 and 1900 UT) and X-type (Fig. 9a, 1828 UT) configurations of the equivalent current flow which have been shown. In this respect, and considering that observations by other methods have been greatly intensified within the same region over the same time, the magnetic data being acquired during the IMS by the Scandinavian Magnetometer Array facilitates new and detailed studies.

However, two limitations of the value of our magnetometer array are obvious. The optical recording on film requires time-consuming and, because of manifold crossing of traces in the case of strong disturbances, often difficult digitizing of the data. Accordingly, our array is particularly appropriate for detailed single event studies, but not well-suited for statistical investigations.

Fig. 9a and b. Equivalent current vectors over the Scandinavian region at different instances of time during the course of the magnetic substorm which occurred on October 7, 1976 (cf. Fig. 8). Note that vector scales are different between the two parts of the figure


1604 UT
BJN
50nT
x_{KI}
+350 km
y_{KI}
-350 km
+350 km
1646 UT
50nT
1726 UT
50nT
1756 UT
50nT
1808 UT
50nT
1828 UT
50nT

a


1836 UT
100nT
1837 UT
100nT
1840 UT
100nT
1848 UT
100nT
1854 UT
100nT
1900UT
100nT

b

Furthermore, the lack of stations at magnetic latitudes greater than 68 N restricts, under not-too-disturbed conditions (i.e., a moderately expanded auroral oval), studies of magnetic disturbances to the southern part of the night-time auroral oval and to subauroral variations at local times other than near magnetic midnight.

With respect to the internal magnetic disturbance field contributions in the Scandinavian area it is concluded that these are negligible at lower frequencies (≤ 1 mHz) and are mainly negligible also in the horizontal components at higher frequencies (~ 5 mHz). The vertical component, Z, often is most indicative of horizontal conductivity variations. At higher frequencies, the behaviour of this component implies the existence of a crust-mantle conductivity structure which is very different beneath southern Finland as compared to northern Scandinavia, with OUL defining approximately the demarcation line between the two regimes. Also indicated, by analyses of data from the total array, is the presence of a conductivity anomaly in the region of OKS-RIS-SRV. Perhaps due to the fact that the water surrounding Scandinavia is shallow, except near the stations FRE and AND which are very close to the continental edge, there is hardly any recognizable coast effect at low frequencies. At higher frequencies the Z variations are greatly amplified at all coastal stations, but this does not appear to lead to a corresponding major perturbation of the horizontal magnetic disturbance vectors.

Acknowledgements. The Scandinavian Magnetometer Array of the University of Münster is part of the European IMS program and has been conceived under the guidance of the Committee on Coordination of Observations with GEOS (CCOG, chairman Dr. W. Stoffregen, Uppsala). Professor Rolf Boström, the CCOG reporter for magnetometer stations, very actively participated in planning and preparing the array. Finally, the array has been installed and operated in cooperation with the following institutions and persons: Aarhus University (Laboratory for Geophysics, Professor S. Saxov), Department of Plasma Physics of the Royal Institute of Technology at Stockholm [Professor R. Boström, (now at Uppsala), Dr. U. Fahleson], Finnish Meteorological Institute at Helsinki (Dr. C. Sucksdorff, Dr. R. Pellinen), Geophysical Institute of the University of Bergen (Dr. E. Gjoen), Geophysical Observatory Sodankylä of the Finnish Academy of Science and Letters (Dr. E. Kataja, Dr. A. Ranta), Kiruna Geophysical Institute (Dr. G. Gustafsson), University of Oulu (Professor J. Oksman), University at Tromsö (previously Auroral Observatory, Mr. St. Berger), Uppsala Ionospheric Observatory (Professor R. Boström). To all of them our deep and sincere thanks are due for permanent support in many respects and for their lasting interest in the project. We also should like to thank the local authorities and persons at the station locations who gave their permission for installing the magnetometers on their ground, and who supported us very kindly in many and often difficult situations throughout the field work. In this respect, we particularly wish to mention Mr. Teuvo Rinkinen and his family at Sauvamäki, and the rocket range administration at Andenes (Andöya).

Installing the many magnetometers and operating them over several years would not have been possible without a vigorous and substantial participation of other members of our institute and particularly of many of our students in the field work. For their very active, responsible and in most cases manifold help in this respect we have to thank H.J. Opgenoorth, W.D. Pelster, J. Heller, B. Inhester, K.-H. Glaßmeier, R. Helm, H. Hopf, U. Mersmann, M. Herzog, M. Keil, R. König, H. Stuhldreyer, O. Hennecke, H. Sulzbacher, H. Dornseif, K. Brüning, and H. Knese.

Almost half of the instruments were constructed especially for this project; the others had to be altered for our purpose. All this has been done by W. Bartsch, H. Dornseif, R. Giesbert, R. Helm, O. Hennecke, H. Knese, and W. Wilting. Difficult instrumental problems were solved particularly by R. Helm and W. Wilting.

We are indebted also to the Deutsches Hydrographisches Institut at Hamburg for the permission to calibrate and test our instruments at its magnetic observatory Wingst.

Financially, the project has been made possible by several grants from the Deutsche Forschungsgemeinschaft, which are gratefully acknowledged.

In this paper, on several occasions we have been able to present magnetic data from additional stations. We are much obliged to Mr. St. Berger (Tromsö), Dr. F. Eleman (Stockholm), Dr. E. Gjoen (Bergen), Dr. G. Gustafsson (Kiruna), Dr. E. Kataja (Sodankylä), Dr. G.A. Loginov (Apatity), Mr. H. Maurer (Braunschweig), and Dr. C. Sucksdorff (Helsinki) for placing the corresponding magnetograms at our disposal.

We thank R. Boström and another unknown referee for valuable critical comments on the original version of the manuscript.

References

Akasofu, S.-I.: Polar and magnetospheric substorms. Dordrecht: Reidel 1968

Akasofu, S.-I., Wilson, C.R., Snyder, L., Perreault, P.D.: Results from a meridian chain of observatories in the Alaskan sector (I). Planet. Space Sci. **19**, 477–482, 1971

Bannister, J.R., Gough, D.I.: Development of a polar magnetic substorm: A two-dimensional magnetometer array study. Geophys. J.R. Astron. Soc. **51**, 75–90, 1977

Bannister, J.R., Gough, D.I.: A study of two polar magnetic substorms with a two-dimensional magnetometer array. Geophys. J.R. Astron. Soc. **53**, 1–26, 1978

Baumjohann, W., Greenwald, R.A., Küppers, F.: Joint magnetometer array and radar backscatter observations of auroral currents in northern Scandinavia. J. Geophys. **44**, 373–383, 1978

Baumjohann, W., Untiedt, J., Greenwald, R.A.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 1. Three-dimensional current flows associated with a substorm-intensified eastward electrojet. J. Geophys. Res., submitted 1979

Birkeland, K.: The Norwegian aurora polaris expedition 1902–1903, Vol. 1: On the cause of magnetic storms and the origin of terrestrial magnetism, Sect. 1. Christiania: Aschehoug 1908

Birkeland, K.: The Norwegian aurora polaris expedition 1902–1903, Vol. 1: On the cause of magnetic storms and the origin of terrestrial magnetism, Sect. 2. Christiania: Aschehoug 1913

Camfield, P.A., Gough, D.I.: Comment on use of anomalous stations for IMS. EOS Trans. Am. Geophys. Union **57**, 850, 1976

Chapman, S., Bartels, J.: Geomagnetism, Vol. I. Geomagnetic and related phenomena. Oxford: Clarendon Press 1940

Chen, A.J., Rostoker, G.: Auroral-polar currents during periods of moderate magnetospheric activity. Planet. Space Sci. **22**, 1101–1115, 1974

Fukushima, N.: Polar magnetic storms and geomagnetic bays. J. Fac. Sci. Univ. Tokyo **8**, 293–412, 1953

Gough, D.I.: Effects of induction in the earth upon measurements of external time-varying magnetic fields. EOS Trans. Am. Geophys. Union **55**, 595–599, 1974

Gough, D.I., Bannister, J.R.: A polar magnetic substorm observed in the evening sector with a two-dimensional magnetometer array. Geophys. J.R. Astron. Soc. **55**, 435–450, 1978

Gough, D.I., Reitzel, J.S.: A portable three component magnetic variometer. J. Geomagn. Geoelectr. **19**, 203–215, 1967

Gustafsson, G.: A revised corrected geomagnetic coordinate system. Ark. Geofys. **5**, 595–617, 1970

Harang, L.: The mean field of disturbance of polar geomagnetic storms. Terr. Magn. Atmos. Electr. **51**, 353–380, 1946

Heppner, J.P.: The Harang discontinuity in auroral belt ionospheric currents. Geophys. Publ. **29**, 105–120, 1972

Hughes, T.J., Rostoker, G.: Current flow in the magnetosphere and ionosphere during periods of moderate activity. J. Geophys. Res. **82**, 2271–2282, 1977

Kamide, Y.: On current continuity at the Harang discontinuity. Planet. Space Sci. **26**, 237–244, 1978

Kamide, Y., Akasofu, S.-I.: The auroral electrojet and field-aligned current. Planet. Space Sci. **24**, 203–213, 1976

Kamide, Y., Akasofu, S.-I., Rostoker, G.: Field-aligned currents and the auroral electrojet in the morning sector. J. Geophys. Res. **81**, 6141–6147, 1976
Kamide, Y., Rostoker, G.: The spatial relationship of field-aligned currents and auroral electrojets to the distribution of nightside auroras. J. Geophys. Res. **82**, 5589–5608, 1977
Kawasaki, K., Rostoker, G.: Perturbation magnetic fields and current systems associated with eastward drifting auroral structures. J. Geophys. Res. **84**, 1464–1480, 1979
Kertz, W.: Modelle für erdmagnetisch induzierte elektrische Ströme im Untergrund. Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl. 2A: 101–110, 1954
Kisabeth, J.L., Rostoker, G.: Development of the polar electrojet during polar magnetic substorms. J. Geophys. Res. **76**, 6815–6828, 1971
Kisabeth, J.L., Rostoker, G.: Current flow in auroral loops and surges inferred from ground-based magnetic observations. J. Geophys. Res. **78**, 5573–5584, 1973
Kisabeth, J.L., Rostoker, G.: The expansive phase of magnetospheric substorms. 1. Development of the auroral electrojets and auroral arc configuration during a substorm. J. Geophys. Res. **79**, 972–984, 1974
Küppers, F., Post, H.: A second generation Gough-Reitzel magnetometer. J. Geomagn. Geoelectr., accepted 1979
Lanzerotti, L.J., Sugiura, M.: Reply. EOS Trans. Am. Geophys. Union **57**, 850, 1976
Lilley, F.E.M.: Magnetometer array studies: A review of the interpretation of observed fields. Phys. Earth Planet. Inter. **10**, 231–240, 1975
Loginov, G.A., Vasilev, E.P., Grafe, A.: Some results of the investigation of magnetic variations of the auroral electrojets concluded from observations of the geomagnetic meridian project (GMP). Gerlands Beitr. Geophys. **87**, 249–262, 1978
Maurer, H., Theile, B.: Parameters of the auroral electrojet from magnetic variations along a meridian. J. Geophys. **44**, 415–426, 1978
Maynard, N.C.: Electric field measurements across the Harang discontinuity. J. Geophys. Res. **79**, 4620–4631, 1974
Mersmann, U., Baumjohann, W., Küppers, F., Lange, K.: Analysis of an eastward electrojet by means of upward continuation of ground-based magnetometer data. J. Geophys. **45**, 281–298, 1979
Nielsen, E., Greenwald, R.A.: Electron flow and visual aurora at the Harang discontinuity. J. Geophys. Res. **84**, 4189–4200, 1979
Rostoker, G.: Polar magnetic substorms. Rev. Geophys. Space Phys. **10**, 157–211, 1972
Rostoker, G., Hron, M.: The eastward electrojet in the dawn sector. Planet. Space Sci. **23**, 1377–1389, 1975
Rostoker, G., Kisabeth, J.L.: Response of the polar electrojets in the evening sector to polar magnetic substorms. J. Geophys. Res. **78**, 5559–5571, 1973
Schmucker, U.: Regional induction studies: A review of methods and results. Phys. Earth Planet. Inter. **7**, 365–378, 1973
Siebert, M., Kertz, W.: Zur Zerlegung eines lokalen erdmagnetischen Feldes in äußeren und inneren Anteil. Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl. 2A: 87–112, 1957
Untiedt, J., Pellinen, R., Küppers, F., Opgenoorth, H.J., Pelster, W.D., Baumjohann, W., Ranta, H., Kangas, J., Czechowsky, P., Heikkila, W.J.: Observations of the initial development of an auroral and magnetic substorm at magnetic midnight. J. Geophys. **45**, 41–65, 1978
Whalen, J.A.: Auroral oval plotter and nomograph for determining corrected geomagnetic local time, latitude, and longitude for high latitudes in the northern hemisphere. Environmental Res. Papers, No. 327, Air Force Cambridge Res. Lab., Bedford, Mass. 1970
Wiens, G.R., Rostoker, G.: Characteristics of the development of the westward electrojet during the expansive phase of magnetospheric substorms. J. Geophys. Res. **80**, 2109–2128, 1975
Wilhjelm, J., Friis-Christensen, E.: Magnetometer chain in Greenland. Geophys. Papers, R-48, Danish Meteorol. Inst., Copenhagen 1976
Wilhjelm, J., Friis-Christensen, E., Potemra, T.A.: The relationship between ionospheric and field-aligned currents in the dayside cusp. J. Geophys. Res. **83**, 5586–5594, 1978
Yasuhara, F., Kamide, Y., Akasofu, S.-I.: Field-aligned and ionospheric currents. Planet. Space Sci. **23**, 1355–1368, 1975

Received June 27, 1979; Revised Version November 22, 1979

J. Geophys. 46, 451–454, 1979

Journal of Geophysics

Short Communication

Estimation of the Dispersion of Compressional Waves in Rocks From Ultrasonic to Seismic Frequencies*

R. Ramananantoandro

Physique et Mécanique des Matériaux Terrestres, Institut de Physique du Globe, Université de Paris 6. 4, place Jussieu, F-75230 Paris, France

Key words: Ultrasonic wave velocity - Dispersion - Anelasticity.

In the investigation of the composition of the crust and the mantle of the earth, velocities measured in the laboratory are compared to seismic wave velocities. Because of dispersion due to anelasticity, question has been raised as to the validity of such comparision (Liu et al., 1976). The laboratory velocities are indeed obtained at ultrasonic frequencies of a few megacycles per second while the seismic velocities at frequencies of a few cycles per second. To date no laboratory velocity measurements on the same sample and at these two extreme frequencies have ever been made to determine the amount of dispersion. In this note, we have made an estimation of the dispersion of compressional waves in a very fine grain limestone. The amount of dispersion was calculated from the internal friction data.

It has been shown that, in a linearly attenuating body, the velocity at a given frequency is uniquely determined from the knowledge of the internal friction Q^{-1} of the body over all frequencies, and vice versa (see e.g. Futterman, 1962). The velocity dispersion relation can be written (Kanamori and Anderson, 1977):

$$\frac{1}{v(f_i)} = \frac{1}{v(\infty)} \left(1 + \frac{1}{2\pi} P \int_{-\infty}^{\infty} \frac{Q^{-1}(f_i)}{f - f_i} df\right) \tag{1}$$

where f is frequency, $v(f_i)$ the velocity at frequency f_i, $v(\infty)$ the velocity at infinite frequency and P stands for Cauchy principal value of the integral. The applicability of this relation is limited by the lack of knowledge of Q^{-1} over all frequencies. Q^{-1} has been measured on rocks at frequencies of a few hertz to a few megahertz, but generally over a narrow frequency band for a particular sample.

For our calculation, we have taken from the literature a rock for which detailed Q^{-1} data are available over the widest frequency band. This is a very

* Contribution no. 338, IPG, University of Paris 6

0340-062X/79/0046/0451/$01.00

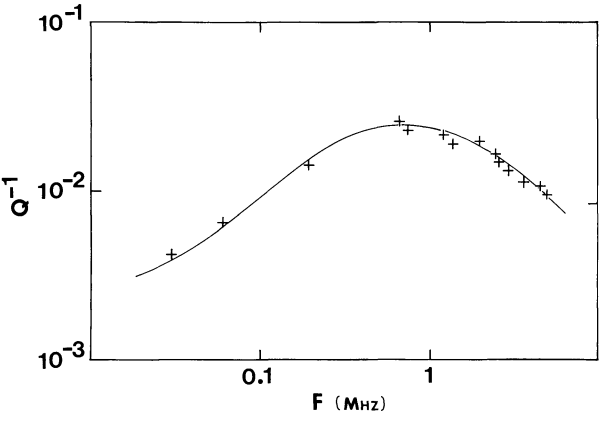


Fig. 1. Internal friction Q^{-1} versus frequency for a Solenhofen limestone (after Mason, 1971)

fine grain Solenhofen limestone, whose internal friction for longitudinal vibrations was measured at ambient conditions from 30 kHz to 5 MHz (Mason, 1971). Given the small grain size of the rock (average grain size 9 μm), absorption due to scattering was considered negligible to a frequency of 5 MHz (Mason, 1971). According to Mason (1971), we have fitted the following equation to the experimental Q^{-1} data (Fig. 1):

$$Q^{-1}(f)=\frac{2.54\cdot 10^{-3}+5.57\cdot 10^{-8}f}{1+1.31\cdot 10^{-12}f^2} \tag{2}$$

Equation (2) is used for a rough estimation of Q^{-1} outside the frequency band of the actual measurements. At low frequencies (1 Hz to 10 kHz), the estimated Q^{-1} ($2.5\cdot 10^{-3}$) falls within the values of Q^{-1} measured on various limestones, as given by Bradley and Fort (1966). These values range from $1\cdot 10^{-3}$ to $5\cdot 10^{-3}$. The high values of Q^{-1} correspond to rocks with high microcrack density. No Q^{-1} data are available below 1 Hz for limestone. Published data on oxides (see Jackson and Anderson, 1970) indicate that, at very low frequencies (below 10^{-5} Hz), Q^{-1} may be large. The effect of large absorption at low frequencies will be discussed below. At high frequencies, absorption due to scattering is dominant but is not considered here, as it does not introduce dispersion.

In order to satisfy the requirement that Q^{-1} be an odd function of frequency (see Futterman, 1962), equation (2) is truncated at $f=0$, and for $f<0$, Q^{-1} is taken to be equal to:

$$Q^{-1}(f)=-\frac{2.54\cdot 10^{-3}-5.57\cdot 10^{-8}f}{1+1.31\cdot 10^{-12}f^2} \tag{3}$$

Using Eqs. (1), (2), and (3), the percentage difference between the velocities at the frequencies f_1 and f_2 is given by:

$$d=\frac{B_2-B_1}{1+B_1}\cdot 10^2 \tag{4}$$

where B_1 and B_2 are the values of

$$\frac{1}{2\pi} P \int_{-\infty}^{\infty} \frac{Q^{-1}(f_i)}{f-f_i} df \tag{5}$$

at f_i equal to f_1 and f_2 respectively. For $f_1=1$ MHz and $f_2=1$ Hz for instance, d is found to be 2.5.

If at frequencies below 1 Hz, the actual Q^{-1} is larger than the estimated one, as mentioned above, the value of d should be smaller than 2.5. Moreover, at high pressure, Q^{-1} decreases due to grain-boundary microcrack closure (Birch and Bancroft, 1938; Walsh, 1966). Birch and Bancroft data (1938) for torsional vibrations in Solenhofen limestone show a decrease in Q^{-1} by a factor of about 2 from 0.2 to 4 kbar, at a frequency of the order of a few kHz. Such a decrease in Q^{-1} would further reduce the value of d. Thus the 1 MHz velocity would differ from the 1 Hz velocity by at most 2.5%. It is noteworthy that, in the present limestone, the frequency interval, 1 Hz – 1 MHz, is outside the frequency domain of the unrelaxed elastic state.

Knowledge of Q^{-1} for other rock types is not as detailed as that of the present limestone. However, values of Q^{-1} measured at ambient conditions on a few mafic and ultramic rocks, at frequencies between 1 Hz and 10^5 Hz, also fall in the 10^{-3} range (Bradley and Fort, 1966). Measurements by Birch and Bancroft (1938) on nine different rock types at a frequency of a few kHz, at room temperature and under a pressure of 4 kbar, show a variation in Q^{-1} by a factor of about three from one rock type to another. One of the highest value of Q^{-1} they have reported is that of a Solenhofen limestone. At megacycle frequencies, Mason (1971) found comparable Q^{-1} for longitudinal vibrations in slate and limestone. His granite shows higher Q^{-1}, but this is presumably due to large microcrack density. On the basis of the above data, the internal friction and therefore the velocity dispersion for varied rock types should not differ much from those for the Solenhofen limestone considered here.

In this note, we have made an estimation of the dispersion of the compressional wave velocity in rocks from 1 Hz to 1 MHz. The dispersion was calculated from laboratory Q^{-1} data collected at ambient conditions. To date, no high temperature laboratory Q^{-1} data over a wide frequency band is available. The effect of temperature, however, is to decrease the viscoelastic relaxation time. This results in a shift of the high frequency limit of the relaxed elastic domain towards higher frequencies and a decrease in the elastic wave velocity dispersion in the frequency band of interest here. On the basis of the laboratory Q^{-1} data for the Solenhofen limestone studied here, the velocities at 1 Hz and 1 MHz is estimated to differ by less than 2.5%. Thus velocities measured at 1 MHz in laboratories are directly comparable to seismic wave velocities measured at a frequency of 1 Hz.

Acknowledgement. The author is grateful to Professor C.J. Allègre, Drs. A. Cisternas, and J.P. Poirier for discussions and to two anonymous reviewers for critical comments.

References

Birch, F., Bancroft, D.: The effects of pressure on the rigidity of rocks. J. Geol. **46**, 113–141, 1938

Bradley, J.J., Fort, A.N. Jr.: Internal friction in rocks. In: Handbook of Physical Constants, Geol. Soc. Am. Mem. **97**, S.P. Clarck, Jr., ed.: pp. 175–193. New York: The Geological Society of America Inc. 1966

Futterman, W.I.: Dispersive body wave. J. Geophys. Res. **67**, 5279–5291, 1962

Jackson, D.D., Anderson, D.L.: Physical mechanisms of seismic-wave attenuation. Rev. Geophys. Space Phys. **8**, 1–63, 1970

Kanamori, H., Anderson, D.L.: Importance of physical dispersion in surface wave and free oscillation problems: Review. Rev. Geophys. Space Phys. **15**, 105–112, 1977

Liu, H.P., Anderson, D.L., Kanamori, H.: Velocity dispersion due to anelasticity: Implication for seismology and mantle composition. Geophys. J. R. Astron. Soc. **47**, 41–58, 1976

Mason, W.P.: Internal friction at low frequencies due to dislocations: Applications to metals and rock mechanics. In: Physical Acoustics: Principles and Methods, Vol. 8, W.P. Mason and R.N. Thurston, eds.: pp. 347–351. New York: Academic Press 1971

Walsh, J.B.: Seismic wave attenuation in rock due to friction. J. Geophys. Res. **71**, 2591–2599, 1966

Received January 22, 1979; Revised Version August 22, 1979