Online flood forecasting in fast responding catchments on the basis of a synthesis of artificial neural networks and process models

Dissertation

In partial fulfilment of the requirements for the degree of
Doctor rerum naturalium

Submitted to the Faculty of Forestry, Geo- and Hydro-
Sciences of Dresden University of Technology

Dipl. Hydrol. Johannes Cullmann

Reviewer:

Prof. Dr. G. H. Schmitz, University of Technology, Dresden
Prof. Dr. H. B. Horlacher, University of Technology, Dresden
Prof. Dr. M. Disse, University of the German Armed Forces, München

Dresden, November 2006

Abstract

A detailed and comprehensive description of the state of the art in the field of flood forecasting opens this work. Advantages and shortcomings of currently available methods are identified and discussed. Amongst others, one important aspect considers the most exigent weak point of today's forecasting systems: The representation of all the fundamentally different event specific patterns of flood formation with one single set of model parameters. The study exemplarily proposes an alternative for overcoming this restriction by taking into account the different process characteristics of flood events via a dynamic parameterisation strategy. Other fundamental shortcomings in current approaches especially restrict the potential for real time flash flood forecasting, namely the considerable computational requirements together with the rather cumbersome operation of reliable physically based hydrologic models. The new PAI-OFF methodology (Process Modelling and Artificial Intelligence for Online Flood Forecasting) considers these problems and offers a way out of the general dilemma. It combines the reliability and predictive power of physically based, hydrologic models with the operational advantages of artificial intelligence. These operational advantages feature extremely low computation times, absolute robustness and straightforward operation. Such qualities easily allow for predicting flash floods in small catchments taking into account precipitation forecasts, whilst extremely basic computational requirements open the way for online Monte Carlo analysis of the forecast uncertainty. The study encompasses a detailed analysis of hydrological modeling and a problem specific artificial intelligence approach in the form of artificial neural networks, which build the PAI-OFF methodology. Herein, the synthesis of process modelling and artificial neural networks is achieved by a special training procedure. It optimizes the network according to the patterns of possible catchment reaction to rainstorms. This information is provided by means of a physically based catchment model, thus freeing the artificial neural network from its constriction to the range of observed data – the classical reason for unsatisfactory predictive power of net-based approaches. Instead, the PAI-OFF-net learns to portray the dominant process controls of flood formation in the considered catchment, allowing for a reliable predictive performance. The work ends with an exemplary forecasting of the 2002 flood in a 1700 km² East German watershed.

Zusammenfassung

Eine umfassende Analyse der bestehenden Hochwasservorhersagesysteme beleuchtet Vorteile und Schwachpunkte heutzutage angewandter Ansätze. Dabei fällt besonders auf, dass in aktuellen Hochwasservorhersagemodellen nur ein einziger Parametersatz zur Beschreibung der vielfältigen Muster der Hochwasserentstehung verwendet wird. Die Arbeit begegnet diesem strukturellen Schwachpunkt mit einem alternativen Ansatz der Modellparametrisierung. Die damit mögliche Verwendung multipler Parametersätze in einem Modell erlaubt die Integration grundlegend verschiedenen Muster der Hochwasserentstehung in einem Modell. Weitere Schwachpunkte der aktuellen Hochwasservorhersage erschweren derzeitig den operationellen Echtzeiteinsatz, besonders die langen Rechenzeiten der detaillierten prozessbeschreibenden Modelle und deren komplizierte Anwendung. Die neue PAI-OFF Methodik (Process Modelling and Artificial Intelligence for Online Flood Forecasting) bietet einen Ausweg aus dieser generellen Problematik. PAI-OFF vereinigt die Zuverlässigkeit und Prognosefähigkeit detaillierter, physikalisch begründeter Prozessmodellierung mit den Vorteilen künstlicher neuronaler Netze. Letztere zeichnen sich insbesondere durch einfache Handhabung und enorm kurze Rechenzeiten in der Anwendung aus. Somit entsteht ein schnelles, robustes und trotzdem zuverlässiges Modell, welches die Online-Betrachtung von Vorhersageunsicherheiten mittels Monte-Carlo-Simulationen erlaubt. Die Synthese der physikalisch fundierten Modelle mit künstlichen neuronalen Netzen wird durch eine spezielle Trainingsstrategie möglich. Dabei werden die Netze nicht mehr – wie bisher üblich – anhand historischer Daten trainiert. Das PAI-OFF-Netz lernt stattdessen die allen möglichen Hochwasserentstehungsmustern zugrunde liegende Prozessdynamik, welche mit den physikalisch fundierten Modellen beschrieben wird. Somit entsteht ein zuverlässig vorhersagefähiges Modell für den operativen Einsatz, welches beispielhaft zur Vorhersage des Augusthochwassers 2002 am Pegel Kriebstein im Osterzgebirge angewendet wird.

TABLE OF CONTENTS

1 Flash flood forecasting: A challenging task

Flash floods represent one of the most common and dangerous natural hazards. However, the menace arising from flash floods worldwide is often not clearly addressed in the media and lacks public awareness. The true dimension of this jeopardy becomes clear from the 30-year average of mortality caused by natural hazards in the United States: More than 120 people are killed by flood events each year. The National Weather Service (2004) states that the majority of the flood related casualties in the United States are the result of flash floods. This is roughly double the death toll of the second most dangerous natural disaster: Hurricanes.

Flash floods are characteristic for small to medium sized catchments. Usually, they are a consequence of severe rainstorms. Regarding the total volume, flash floods are often much smaller than inundations. Nonetheless, due to the immense flow velocities and steep gradients, flash floods pose the most serious threat to human life.

The definition of flash flood is rather soft. The term “flash” reflects a rapid response to the causative event, with rising water levels in the drainage network reaching their peak within minutes to a few hours of the onset of the event. A threshold of approximately 6 hours often is used to distinguish a flash flood from a slow-rising flood (Mogil *et al*; 1978; Georgakakos, 1986; Gruntfest and Huber, 1991; Polger *et al*, 1994; National Weather Service, 2004b). The internationally agreed definitions of flash floods leave an ample scope for interpretation. Symptomatically, the American Meteorological Society (2000) defines: A flash flood is a flood that rises and falls quite rapidly with little or no advance warning, usually the result of intense rainfall over a relatively small area. In this study, flash floods cover events which reach their peak flow less than 24 hours after the onset.

1.1 Restrictions and benefits of flash flood forecasting

Mountainous catchment areas - with their steep slopes and short flow paths, high non linearity and pronounced dynamics - restrict the performance of current flood forecasting models and, thus, only allow for short warning periods. Generally, whenever the minimum required forecast horizon necessary for the implementation of a flood management intervention is shorter than the hydrological response time of the considered catchment, hydraulic routing models can be employed to predict discharge from upstream gauging stations. In this case, the prediction benefits from the accurate portrayal of flood wave propagation arising from the solution of the Saint-Venant equations. This is a very reliable way of forecasting, but unfortunately, when dealing with small- and medium-sized watersheds, the lead-time provided by the hydraulic models is often insufficient to allow for the implementation of protective measures. Moreover, most small watersheds lack upstream gauges, ruling out the above mentioned forecasting approach. Extending the lead time to a period, which allows for an effective reaction to the forecast requires taking into consideration a quantitative precipitation forecast and a detailed and physically based description of the important rainfall runoff processes. Therefore rainfall-runoff models are the means of choice. If the forecasting starts on the basis of measured rainfall, the gain in lead-time is a function of various processes, e.g. interception, surface water storage, soil storage, surface water travel times and flood wave propagation in river channels. When the lead-time is still insufficient - this is generally the case in small and steep catchments - radar-based now-casting and/or quantitative precipitation forecast allow for prolonging the lead-time significantly. Here, the drawback is the inclusion of the precipitation forecast uncertainty. Online evaluation of this uncertainty requires an extremely fast forecasting model.

In the effort of anticipating flash flood related damage to human lives and economic goods, constructional flood retention is the most frequently applied measure. It is a very effective and save way to reduce flood impact on human society if a timely flood forecasting can be used to optimally control artificial flood retention structures. Constructional flood retention requires major investment if rare flood events are to be held back by dams and reservoirs.

Nonetheless, there is no absolute protection from constructive measures if extremely rare events occur.

Especially in cases of extreme flash flood events, more efficient flood forecasting systems are of primary importance. Prolonging the warning periods enables the affected people to safeguard their belongings as well as their lives in case of a devastating event. One of the most evident advantages of flood forecasting systems is their low cost. Besides, they are environmentally friendly because they do not interfere with the natural dynamics of a watershed environment. A valuable and gratuitous side-effect of forecasting and communicating the forecast is the strengthening of awareness for the nature of flash floods throughout the whole society. While the overall system for reducing casualties caused by storms, including not only forecasts and warnings but also public preparedness, has improved steadily since the 1950s and continues to be developed, the comparable flash flood warning system has not experienced the same amount of progress (Doswell *et al* 1996). This statement well portrays the situation around the globe and clearly shows the necessity to improve our understanding of the natural processes governing a flood event in order to push forward the development of more efficient flood forecasting systems. The most crucial drawback of actually available flood warning systems is their short lead-time. In order to reduce the risk for human life, prolongation of the warning period is urgently needed.

The following section of this work comprises a general presentation of the current developments in the field of flood modelling from the various classical approaches to tentative alternatives.

1.2 Flash flood forecasting: Current practice and research efforts

The two focal points of recent scientific effort to improve flood forecasting are:

- ➔ Classical modelling approaches, which rely on routing, rainfall runoff modelling and probabilistic methods in the context of flood forecasting.
- ➔ Artificial neural networks. They are being investigated as a fast and reliable means of flood forecasting.

This section comprises the development of flood prediction techniques from the basics to highly sophisticated models.

Classical modelling approaches:

The development and adaptation of classical concepts in order to enhance flood prediction has been a field of major research activities. Approaches based on routing alone are normally not well-suited for flash flood forecasting. Nevertheless, in order to gain a complete view of the available methods, this class of models is briefly treated in the following.

Approaches based on improved routing:

Barbero *et al* (2001) mainly consider the wave propagation in streams for their flood forecasting system. They employ hydrodynamic numerical modelling together with a simplistic rainfall-runoff model in a rather large catchment, without giving exact details on the computational requirements. In the view of the possibly considerable CPU execution times when dealing with complex rivers and their interaction with tributaries, the numerical experience necessary for operating the system together with the risk of failing convergence seems to represent a hurdle for a routine real time operation. Moreover, being aware of the limited extrapolation performance as regards simplistic rainfall runoff modelling, this process most probably needs more attention, especially for rapidly reacting smaller to medium sized catchments.

Tate and Cauwenberghs (2005) describe a flood forecasting system which has been lately implemented in the Belgian Demer basin. It is built upon a

combination of a simplistic, bucket type rainfall-runoff module and a kinetic description of the flood wave propagation. This system is an example for an approach which mainly relies upon an optimal updating scheme. The modules themselves are likely too simplistic to allow for a reliable extrapolation into the range of extreme floods.

As the approaches based on flood routing are not suited to extend the lead-time in fast responding catchments the focus is turned to methods relying on detailed rainfall-runoff modelling in order to improve the forecast performance.

Approaches based on rainfall-runoff models:

Ludwig *et al* (2006) present a tool for operational flood forecasting. This system is used for flood forecasting in Baden-Württemberg, a federal state of Germany. It is based on the LARSIM model and is well suited for predicting the expected value of flood peaks in catchments where updating with reliable upstream gauge information is possible.

Meetschen and Simmer (2006) propose to use radar data for operational flood forecasting. Their now-casting based system yield promising results. It is not suited for being used in combination with rainfall forecasts yet. This restricts the use of this approach in fast responding, small catchments.

Liu and Todini (2002) consider a quantitative precipitation forecast in their TOPKAPI flood forecast model. It is one of the best-suited models for flood prediction available today. Yet, with a typical TOPKAPI application, it is impossible to correctly portray the highly variable process dynamics over the whole range of possible flood patterns. This is mainly due to the restriction to only one parameter set per model application. This weakness of portraying dynamic processes represents the classical shortcoming of state of the art rainfall-runoff models in hydrology and is likewise affecting all of the approaches described in the following.

Luce *et al* (2006) present a classical approach for online flood forecasting, using the LARSIM model. Their system is based on updating internal state variables and performs excellently for downstream gauges of large rivers. As many other researchers, they use a single set of parameters to describe the whole runoff spectrum, trying to integrate moderate and extreme flood events. This, in

combination with the rough approach of LARSIM, inevitably leads to shortcomings in rare event forecasting. As the system performance is depending on updating, it is not suited for headwater catchments and watersheds which are characterized by short travel times.

Rabuffetti and Barbero (2005) use a conceptual model to forecast floods in a mesoscale Alpine catchment. The model structure does not allow for detailed results from small headwaters. They justify their lumped approach with too long simulation times of high-resolution models. This statement emanates from Michaud and Sorooshian (1994) and Beven (1989). Todini (1996) reports similar drawbacks of detailed physically based models.

These publications lead to a similar conclusion. The computer power currently available is not sufficient to operate a detailed, distributed and physically based online flood forecasting system in environments which do not allow for updating procedures with measured flow or rainfall data. This fact is even more compelling if the evaluation of forecast uncertainty is desired. In order to overcome this general problem, alternative ways have been emphasised lately. One of the proposed answers to the problem of limited computer power is the use of probabilistic approaches.

Probabilistic approaches:

Researchers have endeavoured to simplify and evaluate flood forecasting in order to satisfy needs of watershed management operators. Within this frame, Krzysztofowicz (1999, 2002), Krzysztofowicz and Kelly (2000) and Kelly and Krzysztofowicz (2000) propose probabilistic evaluation of river stage forecasting in headwaters. They show that Bayesian analysis of uncertainties is a possible way for estimating precipitation and hydrologic model uncertainty in the context of flood forecasting. Their contributions are applicable to any kind of forecasting system, but they do not aim at improving the forecast in terms of speed, better process representation or more suitable parameterisation (reliability).

In order to facilitate operational decision making in the case of flood defence, Martina *et al* (2006) propose to circumvent online flood forecasting models with the help of a threshold based critical rainfall rate. The presented approach is

easy to use and fast. Despite the innovative approach, their forecast considerably looses predictive power if the forecast lead-time exceeds 6 hours and their false alarm rate surpasses 50 % for a forecast horizon of 12 hours. This might be due to the employed TOPKAPI model and the historical time series used to derive the threshold values. However, the forecast performance is not satisfying the needs of a distinct enhancement of currently available flood forecasting systems.

Forecasting approaches based on artificial neural networks:

Trying to get around the inconveniences of highly sophisticated numerical approaches (simulation duration, data requirements), a considerable amount of research has been invested for adapting the theory of artificial neural networks (ANN) – also referred to as neural nets - to flood modelling and forecasting during the last 10 years. This is mainly motivated by the principal advantage of neural nets: Once they are trained they are extremely easy to use and outperform classical models by far in terms of simulation speed. General aspects concerning artificial neural networks and their role in hydrology are concisely reviewed in (ASCE 1 and ASCE 2, 2000). Apart from this basic work, a vast number of detailed publications describe the advances in the field of applying neural nets to hydrological modelling:

Kothyari and Garde (1991) developed a regression model for predicting the annual runoff volume from the annual rainfall, basin area, average temperature, and vegetation cover. Sajikumar and Thandaveswara (1996) demonstrated the supremacy of the multilayer feed forward neural network (MLFN) over the regression model in terms of predictive power for the same data.

Halff *et al* (1993) applied a three layer net to portray hydrographs recorded by the U. S. Geological survey (USGS) at Bellvue, Washington. They used observed rainfall hyetographs as inputs. Their net consisted of five nodes in the hidden layer. A total of five storms were considered. On a rotation basis, data from four storms were used for training, while data from the fifth storm were used for testing the network performance.

Hjelmfelt and Wang (1993) developed a neural network based on the unit hydrograph theory. Using linear superposition, a composite runoff hydrograph for a watershed was developed by the appropriate summation of a unit

hydrograph ordinate and the corresponding runoff excesses. The numbers of neurons in the input and hidden layers were kept constant. Rainfall and runoff data from 24 large storm events were chosen from Goodwater creek Watershed (12.2 km²) in central Missouri to train and test their neural network with promising results.

Zhu *et al* (1994) predicted the flood hydrograph in Butter Creek, New York. Online predictions with neural nets outside the range of training data lead to poor results. With increasing forecast lead-time, neural network performance deteriorated. The nets were able to perform better than the autoregressive moving average models (ARMA) compared in this study.

Bonafe *et al* (1994) assessed the performance of a neural network in forecasting daily mean flow of the upper Tiber river, Italy. The previous day discharge, daily precipitation, daily mean temperature, total rainfall of the previous five days and mean temperature of the previous ten days were selected for net input. They concluded that the multi layer nets were able to yield much better performance than ARMA models.

Smith and Eli (1995) applied a back propagation neural network model to predict peak discharge and peak time for a hypothetical watershed. Linear or non-linear reservoir models generate data sets for training and validation. By representing the watershed as a grid of cells, it was possible for the authors to incorporate the spatial and temporal distribution of rainfall into the neural net model. They trained the neural network to map a time series rainfall patterns. Discharge series are modelled using a discrete Fourier series fit of the rainfall hydrograph.

These first approaches towards using artificial neural networks in hydrology do not directly yield practical benefit for flood forecasting, nevertheless the results pointed out the promising potential of the neural network approach at the time of publication. Thus more effort was made in order to harness the neural networks for hydrological modelling and forecasting purposes. Among the numerous attempts the following are the most interesting:

Hsu *et al* (1995) lead the way to single step predictions of stream flow employing a 3-layer network. Hsu *et al* (1997) further develop the potential of

recurrent neural approaches in the context of hydrological modelling. Castellano–Mendez (2004) compared a forecast algorithm on the basis of the ARMA concept (Box-Jenkins 1976) with a multi-layer net based forecast strategy. For daily single step runoff predictions he found the non-linear multi-layer net performing better than the linear ARMA approach.

Dawson and Wilby (1998) used a three-layer back propagation network to determine runoff from the catchments of the rivers Ambers and Mole. The two catchments are of nearly equal size (about 140 km²). Observed flow data and mean historical rainfall data serve as inputs in their study. Their results show that nets perform similar to an existing model on less input information.

Tokar and Johnson (1999) reported that an artificial neural net model yields higher training and testing accuracy when compared with regression and simple conceptual models. Their aim was to predict daily discharge for the Little Patuxent River, Maryland. Daily precipitation, temperature, and snowmelt equivalent served as inputs in their study. It was found that the selection of training data had a great impact on the accuracy of the prediction. The authors trained and tested their nets with wet, dry and average-year data respectively as well as combination of these in order to illustrate the impact of the training series on network performance. The net that was trained on wet and dry year data had the highest prediction accuracy. The length of the training data had a much smaller impact on network performance than types of training data.

Zealand *et al* (1999) described the potential of neural nets for the short term forecasting of stream flow. Their work explored the capabilities of artificial neural networks and compared their performance to conventional approaches used to forecast stream flow. Nets were examined for sensitivity with respect to the type of input data as well as the number and the size of the hidden layers. The test of their net approach was realized for a part of the Winnipeg River in northwest Ontario, Canada. Forecasting was conducted on a catchment area of approximately 20,000 km², using quarter weekly time intervals. A very close fit was obtained during the training phase and the developed neural network model consistently outperformed a conventional model during the verification (testing) phase for all of the four forecast lead-times considered.

Tokar and Markus (2000) compared neural network models with traditional conceptual models for the purpose of predicting watershed runoff as a function of rainfall, snow water equivalent and temperature. They applied neural network technique to model watershed runoff in three basins with different climatic and physiographic characteristics. In the Fraser River watershed monthly stream flow was modelled with artificial neural nets and compared to a conceptual water balance model (Watbal). The net technique was used to model the daily rainfall-runoff process and was compared to the Sacramento soil moisture accounting model in the Raccoon River watershed. The daily rainfall-runoff process was modelled in the Little Patuxent River basin with a neural network. Here the training and testing results were compared to those the simple conceptual rainfall-runoff model (SCRR). Their initial results indicate that artificial neural nets are powerful tools for modelling the precipitation-runoff process for various time scales, topography, and climate patterns.

In an attempt to overcome the limited extrapolation capacity of neural networks, Imrie *et al* (2000) use different activation functions for the output layer of their multi layer net. They show that a cubic polynomial function performs better than linear and sigmoid functions for their validation data set. However, they restrict their statement to one attribute of one event (maximal flow within the validation dataset). This seems questionable if the models purpose is focused on prediction instead of analysis of defined sections of hydrographs. Furthermore, the use of different output layer transfer functions does not impact on the quality of the training. A neural networks performance is only as good as its training database. If there are not enough sampling points for the net the various transformations in the output layer can help improving single applications, but do not cure the syndrome of nets trained on limited or unrepresentative data.

Hu *et al* (2001) have tried to enhance their multi.-layer net based river stage modelling approach by training various nets for different parts of the observed flow. Therefore they divided the flow spectrum in low, medium and high flows. This approach takes advantage of the closer distance of sampling points each of the nets is trained on. This consequently leads to better training and generalisation performance of the three networks. The advantage of this approach - driven by observed values - has to be paid for by the specialisation of each of the three nets. They make use of very restricted data only, thus training

of a restricted part of the flow dynamics hinders the application potential for the forecast over the whole range of flows. It remains unclear if the modelling approach can be transferred to the much more complex task of rainfall runoff prediction on the highly dynamic hourly time step.

Earlier, Zhang and Govindaraju (2000) had presented a similar approach for rainfall runoff prediction on the monthly time step. They had trained various so-called expert networks, each being used to portray a certain range of stream flows. Then a "gating network" represents the output of the expert networks depending on the actual state of the process. This methodology works well for monthly predictions, where highly dynamic flow components do not play a decisive role. It uses very simplistic input information and it might therefore be difficult to predict flash floods on the basis of this approach.

Hettiarachchi *et al* (2005) try to improve the forecast ability of their multi-layer net by incorporating an estimated maximum flood (EMF) in the training data set. This leads consequently to better validation results if the EMF hydrograph is contained in the validation data set. However, this evaluation strategy has little significance because it focuses on just one sampling point (the EMF), ignoring the range of all possible flows between the observed and the EMF. Their net evaluates five input vectors only: Two antecedent flows, the actual rainfall rate and two antecedent rainfall features. It is questionable weather a network provided with such scarce information is able to reproduce the flow dynamics over a wide range of possible realisations.

Cigizoglu (2003) used ARMA models to generate synthetic series. These data are incorporated into the training database of neural networks to increase the predictive ability. The method was applied to the monthly mean river flow data of a station in Turkey with good results. The extension of input and output data sets in the training stage improves the accuracy of forecasting based on artificial neural nets. Introducing periodicity components in the input layer also increases the forecasting accuracy of neural networks.

Laio *et al* (2003) compared two different non-linear models, non-linear prediction and artificial neural networks for multivariate forecasting of the water stages of river Tanaro, Italy. Very good results are reported for both methods.

However, non-linear prediction performs slightly better for short lead times (1-6 h) while the situation is reverse for longer forecast horizons.

Dawson and Wilby (2001) provide a good review regarding neural networks and their application with respect to rainfall runoff modelling. Sajikumar and Thandaveswara (1999) and also Shamseldin (1997) tried to predict flood waves using neural nets which are trained exclusively by observed rainfall runoff data. Notwithstanding the fact that their results were more than satisfactory as far as the phenomena to be predicted remained in the range of the observations, their approach cannot be used for general flood forecasting. This emanates from the poor extrapolations capacities of nets which are tellingly characterized by the statement of Minns and Hall (1996): *"artificial neural networks are a prisoner of their training data"*. As observations practically never cover the full range of possible flood peaks, such a completely empirical approach is doomed to failure as regards reliable flood forecasting.

Comparative research has been carried out on various net types in view of their respective abilities to predict discharge time series (Hsu *et al* 2002). Bruen and Yang (2005) used functional networks which were recently introduced as an alternative for real time flood forecasting. They applied two types of functional network models, separable and associative networks to forecast discharge for different lead-times. They compared their results with a conventional artificial neural network model, an ARMA model and a simple baseline model in three catchments. Results show that functional networks are flexible and comparable in performance to artificial neural networks. Their results were obtained with only the most simple structure of functional networks and they recommend that the use of complex forms of functional networks might further improve the forecast.

Foka (1999) uses polynomial neural networks (PoNN) for discharge modelling. His attempt focuses on relatively small numbers of input vectors (up to a few time steps back into the past only) as well as high-degree polynomials (up to 7.th degree). This makes the adaptation of the polynomial nets easier, but weakens the ability to generalize, which in turn spoils the predictive power of the approach.

Further attempts aim at improving the potential forecast performance of neural networks by combining different net types with methods of time series analysis or fuzzy logic approaches (Rajurkar and Nayak *et al*, 2004). Kang and Kang (2006) compare the performance of several different multi-layer feed forward neural networks with the Grey-model. However, the nets in their study are exclusively trained with historical data, rendering the extrapolation beyond observed events impossible. Further, the forecast horizon is restricted to 6 hours. All together, this makes their study a valuable contribution towards the understanding of watershed specific neural network architecture, but their work does not contain new developments towards the operational use of neural networks in flood forecasting.

Unfortunately current strategies are all built upon purely empirical approaches and, thus, feature the same shortcomings – i.e. they cannot reliably predict a rare extreme flood event if it is not part of the training data. Schmitz *et al* (2005) as well as Cullmann *et al* (2006) aim to offer a way out of this dilemma. They comprehensively enlarge the training database by exploiting catchment specific topographic and soil hydraulic properties with the help of detailed hydrological catchment modelling. Following these lines they achieve first promising results regarding flood forecasting in a small catchment. However, their multi-layer net approach did not turn out to work fully stable under all conditions – an effect which was observed to aggravate with growing catchment area.

Summary:

Analysing the research effort so far leads to the awareness that all of the aforementioned classical approaches are confronted with two general obstacles limiting model performance in the context of flood forecasting. The first hurdle is the “singular parameter set problem”. It reflects the failure to adequately portray the different patterns of flood formation with one singular model parameter set. This shortcoming prevents actual systems from accurately describing the whole range of natural processes governing the flood formation (infiltration excess flow, soil water movement etc.). The second reason that impedes more reliable forecasts is referred to as uncertainty. Predictive uncertainty has been the focus of a vast number of studies during the last decade (Freer, 1996; Beven, 1992; Beven and Freer, 2001; Christiaens, 2002; Prudhomme, 2003; Butts *et al* 2004; Beven 2005). There are two main sources

of uncertainty in real-time flood forecasting: The input data (measured or predicted), and the hydrological models (Krzysztofowicz 1999). The hydrological model uncertainty originates from the "degree of ability" of the different models transforming the input data into the flood forecast. The different types of input data uncertainty are:

1. The measurement uncertainty, which is mainly related to spatial distribution and dynamics of rainfall fields;

2. The meteorological forecasting uncertainty, arising from the possibility of extending the forecasting horizon beyond the response time of the physical system by means of quantitative precipitation forecast.

Recapitulating this review leads to evident consequences:

Simple bucket type models are fast, but not suitable for extrapolating the forecast into the range of extreme event patterns due to the lack of fidelity to the governing hydrological processes.

Physically based distributed models well portray the dynamics of flood formation, even for extreme events. However, one single set of the many parameters of these complex models is not enough to describe all naturally occurring flood patterns. Furthermore, the process information inherent to these models is inversely proportional to their application speed, i.e. better models are slower, best models are the slowest.

Models based on artificial neural networks are fast and reliable. Nevertheless, they are not suited for extrapolating beyond the range of their training data. This makes forecasting all potential extreme event patterns impossible if the training database exclusively consists of observed data.

The way out of this cul-de-sac is a combination of the advantages of neural networks (speed, easy application) with the advantages of physically based, detailed hydrological models (predictive ability).

1.3 Overcoming the obstacles of state of the art flash flood forecasting

In response to the deficiencies emerging from the analysis of existing flood forecasting models and as an answer to the shortcomings illustrated above, the main objective of this study is the development of a new approach for online flood forecasting especially suited for fast reacting catchments. The proposed PAI-OFF (Process modelling and artificial intelligence in the context of flood forecasting) methodology builds upon catchment specific hydrological and meteorological data, physically based catchment modelling, and artificial intelligence. Thus, PAI-OFF offers an alternative to the limited extrapolation capacity of a flood forecasting system based upon either simplistic hydrologic approaches or on neural networks trained solely on the basis of historical data. These types of models make the prediction of very rare - or even unobserved - flood events unreliable (e.g. Minns and Hall, 1996). The main contribution of this work consists of:

Setting up a detailed distributed hydrological catchment model that guarantees the portrayal of the rainfall-runoff process by a parameterisation strategy, which introduces additional knowledge about event specific characteristics into the calibration procedure. This allows for taking into account the different dominant processes governing the whole range of naturally occurring patterns of flood formation. This is a clear advantage in comparison to the classical way of parameterising catchment models, where the focus is set on determining one best (or equifinal) set of parameters. This latter strategy always implies the averaging of model process representation in that one set of calibrated parameters describes the best average fit of the different concerned calibration events (implying the loss of process sharpness for the calibrated parameters).

Developing an artificial neural network which is capable of adequately representing all important rainfall-runoff processes in the flash flood context.

Synthesising the representation of the physical processes inherent to the catchment model with the artificial neural network on the basis of a comprehensive database which includes all the catchment specific response patterns to rainstorm events depending on watershed internal preconditions. The result is a custom-made artificial neural network forecasting system. This model

is extremely fast, incorporates various parameterisations and thus opens the way for online uncertainty estimation on the basis of ensemble forecasts.

The presented work takes full advantage of artificial intelligence in the form of neural networks without loosing the physical meaning of the processes dominating flash floods. Therefore, the following sections 2 and 3 encompass the basics of flash flood modelling, considering process models and neural network based strategies respectively. Section 4 is a detailed description of the PAI-OFF methodology.

2 Foundations of hydrological modelling in the context of online flash flood forecasting

Process-oriented models are assumed to portray system dynamics even beyond the range of calibration data. Due to the fact that the required input data are not always completely available and model structures are only a crude description of the natural processes, model parameters need to be calibrated. The parameterisation of hydrological models has been the subject of enormous scientific effort throughout the last decades (Gupta *et al* 1994; Gupta and Dawdy 1995; Post and Jakeman 1996) The results of these studies contribute a valuable share to our general understanding of parameterisation of both conceptual and process models and substantiate the awareness that model parameterisation is the key to setting up a reliable hydrological model.

Considering this, a selected hydrological model is studied in order to improve the principles of successful parameterisation in the context of flash flood forecasting in this chapter. It is essential to carefully choose the model in order to incorporate as much process knowledge as possible. This statement is based on the awareness that more process information makes operating the model - even under conditions that have not yet been observed - more reliable. Therefore the vast number of existing catchment models is already restricted to physically based and distributed approaches. Lumped models and purely conceptual approaches do not fulfil the requirements of fidelity to the natural processes and therefore do not yield enough extrapolation capacity to portray extreme events. In the following, three distributed rainfall-runoff models are briefly described. One of them is selected and evaluated with a basic sensitivity analysis in a first step. On the basis of the insights from this study, a more advanced approach of parameter identification is employed to gain further information about meaningful model parameter ranges and eventual relations of optimal parameter sets to internal state variables such as the catchment pre-event wetness index. In a next step, an automatic calibration approach is used to optimise model parameters for 36 separate events. These parameter sets are then statistically analysed for a potential link to a priori knowledge concerning the event pre-conditions and characteristics. The conclusions from these exercises lead to a

new parameterisation strategy. It allows for taking into account different classes of flood events. Thus the set-up of hydrological models can be improved by skilfully including additional, event class specific information into the calibration procedure.

Model description and selection:

The **TOPMODEL** approach (Beven and Kirkby, 1979; Beven *et al*, 1995) allows for distributed modelling of hydrological processes, but needs heavy calibration of parameters which mostly lack physical interpretability. The integration of topographic information within the model structure allows for the simulation of the spatial distribution of soil water content and groundwater levels. The variable source area concept leads to a dynamic distinction between fast saturation excess overland flow (Cappus, 1960; Dunne and Black, 1970) and slow subsurface flow. The model has been used in numerous applications during the last years (Beven, 1997). However, as emphasized by Beven (1997), several simplifications and assumptions restrict the physical basis of TOPMODEL and require that the model is used with care, especially in the context of different spatial scales or time steps. These shortcomings are documented and described in detail in Liu and Todini (2001). The TOPMODEL approach has therefore been excluded for further use within this study.

The **TOPKAPI** (TOPographic Kinematic APproximation and Integration) model (Todini, 1996; Ciarapica and Todini, 1998; Ciarapica and Todini, 2002; Liu and Todini, 2001) is a conceptual rainfall-runoff model integrating physically meaningful parameters. It was developed as a result of the critical analysis of TOPMODEL and the Stanford Watershed Model (Crawford and Linsley, 1966) with the aim of:

- Exploiting the potential of distributed models based upon physically meaningful parameters (Abbott *et al* 1986; Beven, 1989);

- Incorporating the possibility of obtaining a lumped version of the same model, by integration of the processes over increasing size domains, avoiding recalibration during the application of the model at increasing spatial scale - from hill slope to catchment - (Todini, 1995).

The model is based upon the assumption that the horizontal soil internal flow, overland flow and channel flow can be approximated by means of a kinetic wave model. Two versions of the TOPKAPI model are available, namely the distributed version and the lumped version. In the distributed TOPKAPI model, the point assumption is integrated up to a finite pixel dimension, thus transforming the original differential equation into a non-linear reservoir equation based upon physically meaningful parameters (Liu and Todini, 2001). The TOPKAPI model has been tested extensively in the preliminary phase of this dissertation. Unfortunately the documentation of the model is very poor, the property of calibration parameters is not described and the source code of the model is not available. The model input and output interfaces are unclear without further documentation. Even a close contact to the developers of this model did not clarify all doubts; this finally led to the exclusion of the model from the catchment hydrologic studies which are part of this work.

WaSiM-ETH/6.4 (in the further referred to as WaSiM) (Schulla 1997) is a distributed, grid-based model. It has been successfully used in various studies on mountain hydrology in the Swiss Alps (Gurtz 1999; Gurtz 2003). The main components of the model are:

- Interpolation of meteorological input data (Schulla and Jasper1999)
- Topography based radiation correction
- Interception storage and evaporation
- Evapotranspiration (Monteith 1965)
- Snow accumulation and snowmelt (Braun 1985, Anderson 1973)
- Glacier melt (Hock 1999, Badoux 1999)
- Infiltration and surface runoff formation (Green and Ampt 1911)
- Soil water storage, percolation, interflow formation
- Soil moisture loss by vegetation activity
- Groundwater recharge
- Runoff formation, concentration and discharge routing

The soil module of WaSiM determines the infiltration of water and the surface runoff formation according to Green and Ampt (1911) using the two-step model approach (Peschke 1987). The vertical flow of water in the unsaturated zone (percolation) is calculated by means of the discrete 1D-Richards equation.

Interflow is generated in different soil layers depending on the suction, the drainable water content, the hydraulic conductivity and gradient as well as the flow density.

The WaSiM model has been chosen for the model parameter analysis in this study because its physically based unsaturated zone module maintains the characteristic physics of dynamic rainfall-runoff processes even for unobserved events. This is especially important for correctly portraying the pre-event catchment conditions. WaSiM is therefore – amongst the available models - one of the best suited for extrapolation into the range of extreme flood events.

2.1 Basic sensitivity analysis for model parameterisation

A sensitivity analysis is carried out in order to deepen the understanding of model sensitivity to input parameter changes described in Schulla (1997). A univariate analysis according to Benaman (2002) evaluates the sensitivity with respect to a set of reference parameters. The focus is set on identifying and interpreting the most sensitive parameters affecting runoff formation. Therefore the study concentrates on the soil water module of WaSiM.

2.1.1 Study area and data

The study is carried out in the watershed of the Schwarze Pockau River at Zöblitz gauging station. The catchment covers the northern slopes of the Ore Mountains in Eastern Germany from about 1000 MSL down to 450 MSL. Roughly 40 % of the watershed is covered by forest whilst the agriculturally used area accounts for about 30 %, the rest being fallow. The basin hydrology is near natural state; there are no major human impacts to the flow dynamics. The Bük 200 is the basis for soil classification and the Van Genuchten parameters are derived according to the AG Boden (2005). The land use data is taken from Corine (2000). The meteorological data used in the study consists of 1 km grids of precipitation, temperature, wind speed, air humidity and global radiation. The data is interpolated from meteorological stations using external drift kriging for precipitation and temperature and ordinary kriging for the other input data.

The analysis is carried out for a summer flood that occurred in 1983. This event has a return period of about 20 years. In the analysis 201 hourly time steps are evaluated starting from the 3rd of August 1983, 2:00 am.

2.1.2 Parameters, methodology and results of the analysis

In a first step the model is checked for a realistic representation of the various model generated flow components for the reference parameters, i.e. it is assured that interflow (generated from the soil module) and direct runoff (as a result of the Green and Ampt model) both exhibit realistic values for the reference model. In figure 1 these components reflect the characteristics of the considered 1983 flood event. After a very dry summer period a heavy rainfall hits the watershed. The reference model well portrays the first reaction of the watershed in the form of direct runoff. This can be interpreted with the first hydrophobic reaction of the very dry soil surface. After some time, the soil surface is wet and precipitation starts to infiltrate. At this stage the runoff is dominated by interflow. The base flow is constantly low. This is also typical for periods at the end of a long and dry phase in the test watershed. All in all, the model well represents the dynamics of the event and all the model inherent flow components are correctly representing the test hydrograph. WaSiM is now used for stepwise univariate sensitivity analysis with respect to the reference model.

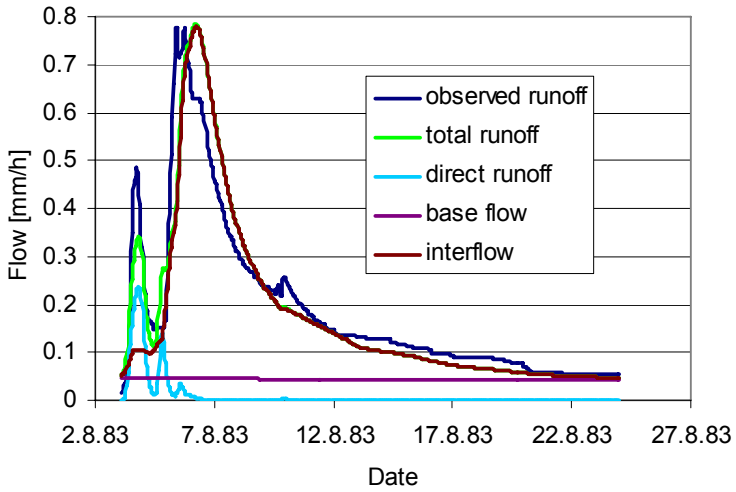


figure 1: Flow components of the reference model

The model parameters considered for the analysis are listed in table 1. This table also contains the reference model parameter set and the range of the parameter variation as well as the step for the uniform distribution of the test parameters.

table 1: Parameters used in the sensitivity analysis

Parameter	Reference	Range	Step	Model impact
dr [-]	4.2	1-20	0.2	Scaling of interflow
Ki [h]	4	2-50	0.5	Storage of interflow
Kd[h]	15	1-33	0.5	Storage of direct runoff
krec[-]	0.1	0.05-0.99	0.01	Soil hydraulic conductivity gradient
thick [m]	0.6	0.15-1.5	0.05	Numerical soil parameter

WaSiM transforms rainfall into runoff according to the scheme shown in figure 2. Here, three exemplary soil water compartments receive infiltration from the Green and Ampt approach. This module is also used to determine the direct runoff *Qd* in the model. *Qd* is then routed via a flow-time grid and finally projected cell-wise to the watershed outlet by means of a simple bucket type function (equation 1). The recession coefficient of this function is *Kd*.

$$Qd = Qd_{i-1} \cdot e^{-\Delta t / Kd} \qquad \textbf{(equation 1)}$$

Where *Qd* is the direct runoff and *Qd(i-1)* is the runoff in the preceding time step Δt. The soil water movement through the layers is modelled by means of the discrete form of the Richards-equation:

$$\frac{\Delta \Theta}{\Delta \mathrm{t}} = \frac{\Delta q}{\Delta z} = q_{in} - q_{out} \qquad \textbf{(equation 2)}$$

Here $\Delta \Theta$ denotes the change in soil water content, Δt defines the time step, Δq is the change in specific flux. The fluxes q_{in} and q_{out} characterize the influx and efflux from the specific soil layer respectively. Finally, Δz defines the thickness of the soil layer. This last parameter is equal to the parameter *thick* in the analysis. The Van-Genuchten parameters used for the solution of the Richards equation are used as fixed values in this study. This allows for focusing the

analysis of model sensitivity to the more conceptual parameters listed in Table 1. Each soil layer produces interflow (Q_{ifl}) according to (equation 3), which is cell-wise scaled with *dr*.

$$Q_{ifl} = k_s(\Theta_m) \cdot \Delta z \cdot dr \cdot \tan\beta \qquad \textbf{(equation 3)}$$

where: *ks* = hydraulic conductivity
Θ_m = Water content of the specific layer
dr = scalar
β = local slope

The interflow is again projected to the watershed outlet by means of the flow-time grid and a second bucket type function. Herein, *Ki* represents the recession coefficient in analogy to (equation 1, *Kd*). The parameter *krec* ranges between 0 and 1. It describes the gradient of the soil hydraulic conductivity within each of the soil layers to account for integral conductivities of partly saturated layers (three in figure 2), i.e. the saturated soil hydraulic conductivity is multiplied with *krec* to reduce the saturated conductivity to an "effective" soil layer conductivity. The soil layers are topping the compartment from which base flow is generated by means of a simple, empirical approach. This approach is not affecting the modeling in the context of flood forecasting, therefore it is not further mentioned in the context of this study. More details about WaSiM are documented in Schulla (1997).

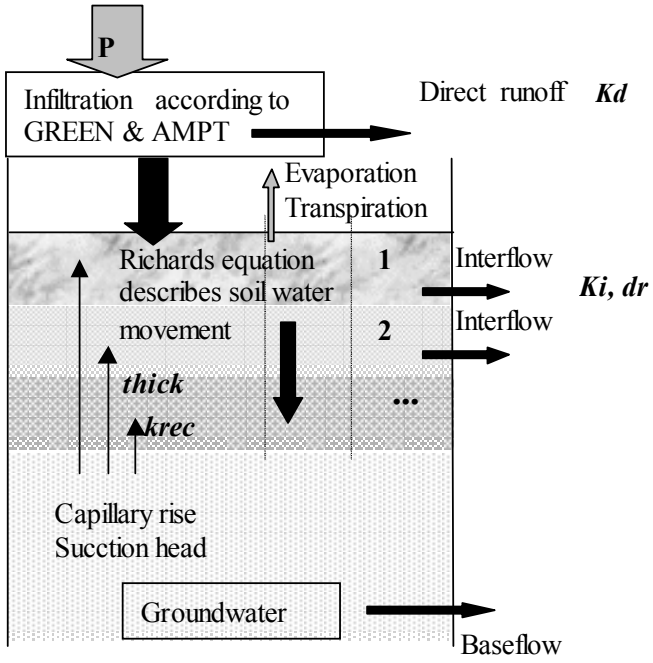


figure 2: Scheme of the WaSiM soil module with location of impact of model parameters (bold)

The range of the test parameter variation (table 1) is set according to values published by Schulla (1997). In the cases of missing reference (*Ki, Kd*) realistic maximum values for the test catchment were chosen on the basis of analysing the recession phase of direct runoff and interflow components of observed events. The parameter *thick* is a numerical parameter and has to be chosen according to model numerical stability criteria (equation 2). It is used in this context for comparative purposes only.

Benaman (2002) explains that if the analysis outputs vary significantly for a specific parameter, then the output is sensitive to the specification of the input distribution of the test parameter, and hence this parameter should be defined with care. As a relatively simple deterministic measure, a sensitivity coefficient can be used to describe the magnitude of change in an output variable V per unit change in the magnitude of an input parameter value P from its base value P_0. The model sensitivity to a variation of the considered parameter is proportional to the value of the index, that means the closer the index to 0, the smaller is the sensitivity. Let Si be the sensitivity index for an output variable V with respect to

a change ΔP in the value of the input variable P from its base value P_0. Noting that the value of the output $V(P)$ is a function of P, the sensitivity index is defined as:

$$Si = \frac{V(P_0 + \Delta P) - V(P_0 - \Delta P)}{2\Delta P} \qquad \textbf{(equation 4)}$$

The sensitivity index is calculated for each of the test parameters considering the peak flow values. This reflects the importance of correctly portraying the peak flow rates in flash flood forecasting applications. In addition to Benaman (2002), a second sensitivity measure describes the mean absolute sensitivity (MV). This index is calculated once for the whole event [MV(201)] and once for the 101 first time steps [MV(101)]. This latter index describes the model only for the rising flood wave and the peak flow area. The two mean sensitivity measures are expressed as:

$$MV(201) = \frac{1}{201}\sum_{1}^{201}\frac{[V(P_0 + \Delta P) - V(P_0 - \Delta P)]^2}{V(P)} \qquad \textbf{(equation 5)}$$

$$MV(101) = \frac{1}{101}\sum_{1}^{101}\frac{[V(P_0 + \Delta P) - V(P_0 - \Delta P)]^2}{V(P)} \qquad \textbf{(equation 6)}$$

The results of the sensitivity analysis are shown in table 2 and figures 3 through 7. From table 2 it becomes clear that the calculation of various sensitivity measures may lead to contradictory results. Specifically, *thick* is the most sensitive parameter for all sensitivity measures, whilst *krec* is the least sensitive parameter if the mean absolute sensitivity measure over the whole period is considered. This contrasts the fact that *krec* is more sensitive than *Ki* and *Kd* for the peak value sensitivity index. Generally, the mean absolute sensitivity measure for the first 101 time steps is smaller than the mean absolute measure for the whole period. The exception to this rule is *krec*. This contradiction can be explained with differing regions of impact of the two sensitivity measures. The mean sensitivity index for the first 101 time steps gives more weight to the rising limb of the event. The recession phase is only accounted for by the mean sensitivity index covering the whole period As *krec* has impact on exactly this recession phase alone, it can only be evaluated properly with the index covering the whole period.

table 2: Sensitivity criteria for the test parameters

Parameter	*Peak sensitivity index*	*Mean absolute sensitivity*	*Mean absolute sensitivity*
	Si	*MV(201)*	*MV(101)*
dr [-]	0.023	0.01	0.02
Ki [h]	-0.006	0.0036	0.0049
Kd[h]	0.0009	0.0042	0.083
krec[-]	0.003	0.0016	0.001
thick [m]	-0.070	0.355	0.699

More valuable information is inherent to the graphs themselves (figure 3,4,5,6,7) and should be evaluated to complete the information obtained by means of the sensitivity measures.

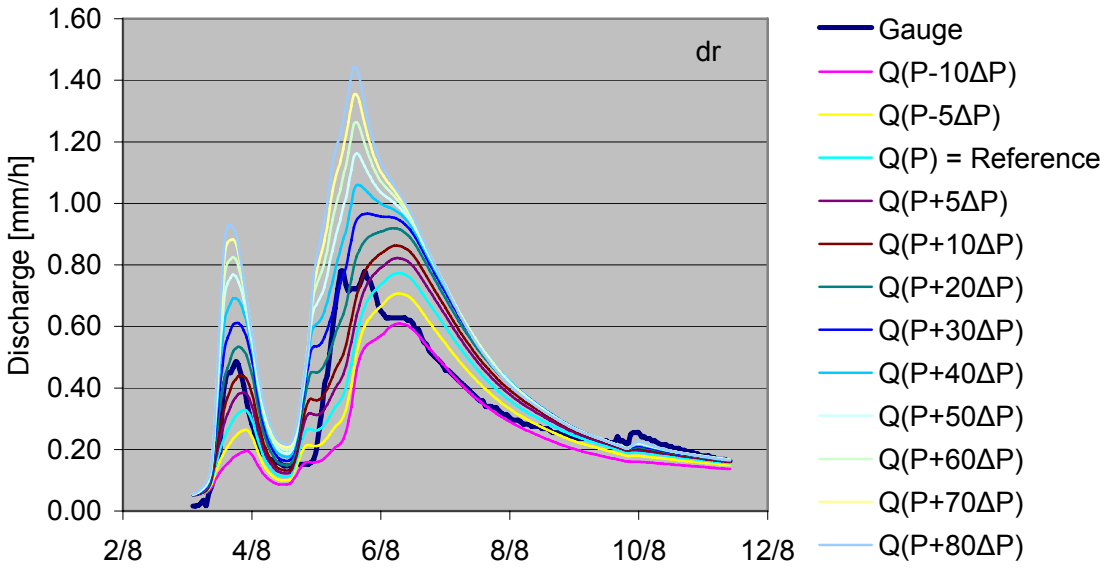


figure 3: Model sensitivity with respect to the parameter dr

From figure 3 we learn, that *dr* is a most important parameter for model calibration. It influences the hydrograph strongly over the whole test period.

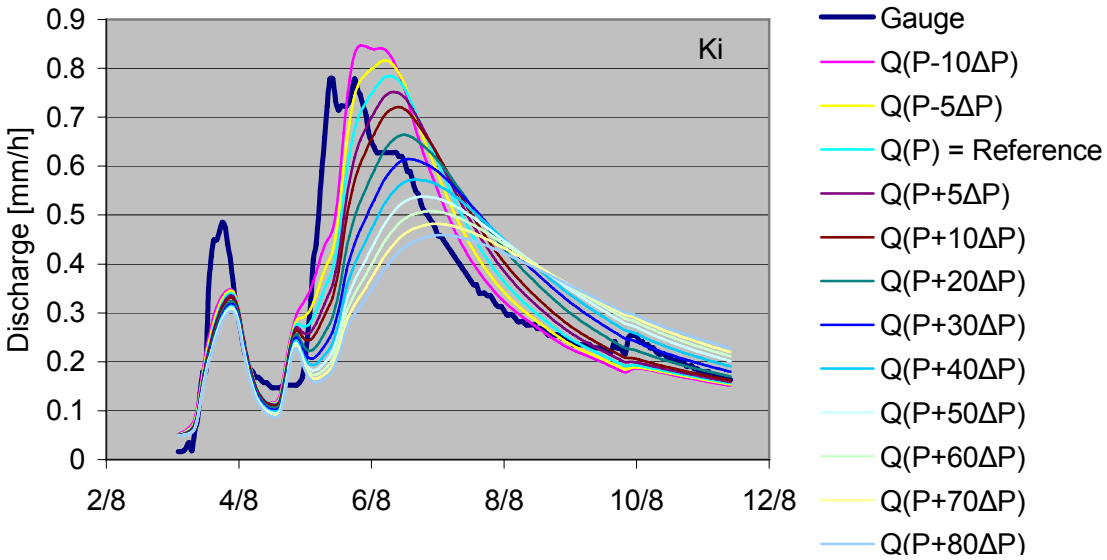


figure 4: Model sensitivity with respect to the parameter Ki

Ki and *Kd* are meaningful where the corresponding flow components dominate the hydrograph, e.g. the first little peak is very sensitive to *Kd* (figure 5). This is also expressed by the high value of the mean sensitivity index for the first 101 time steps (table 2). The main peak is more sensitive to *Ki* (figure 4), although this does not become clear at first sight from the sensitivity measures alone. This reveals another weak point of the simple sensitivity measures. They are subjected to changes of process dynamics throughout the considered test phase. This becomes clear if figure 1 and figure 5 are jointly considered. For the test event, direct runoff is only dominating the first peak of the hydrograph. If the sensitivity indices are calculated for periods without direct runoff, the numerator of the equations becomes 0. This is obvious humbug and might mislead the interpretation of a simple sensitivity analysis.

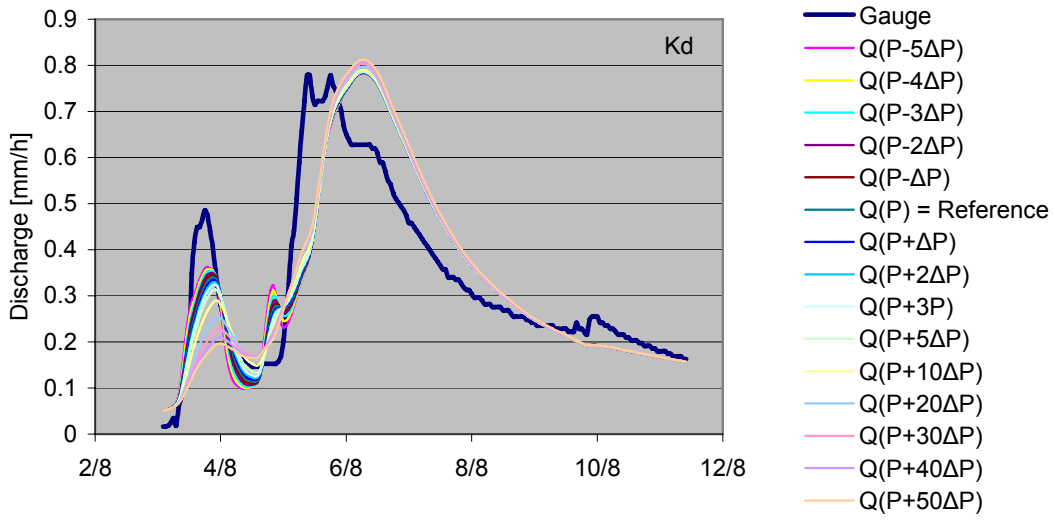


figure 5: Model sensitivity with respect to the parameter Kd

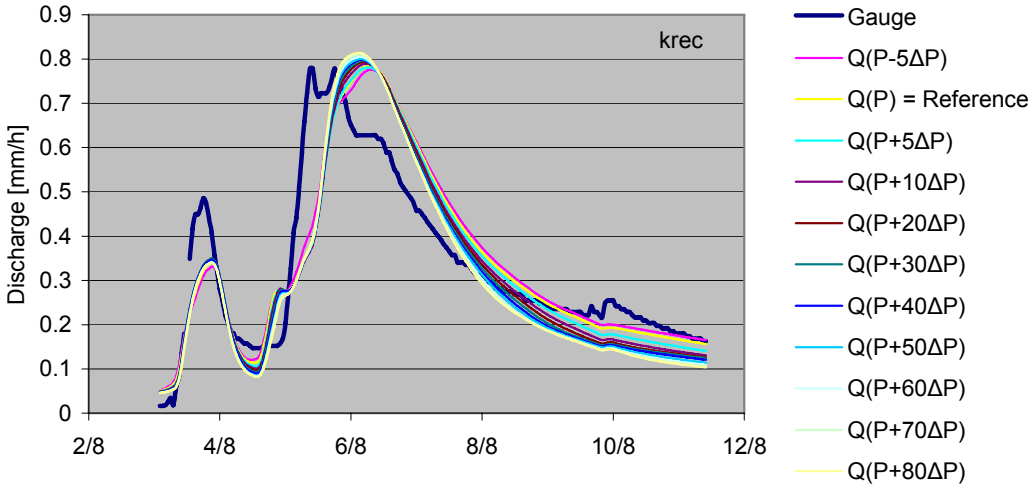


figure 6: Model sensitivity with respect to the parameter krec

It therefore becomes especially tricky to evaluate simple sensitivity measures if various events or long time series are used in the analysis. The *krec*-graph shows the least sensitivity for the parameter variation. This holds for the whole range of the test period. The graph is shown in figure 6 and reveals that the most sensitive interval of this parameter is the second half of the test period where the recession phase is portrayed. As the special emphasis in flood forecasting is on the rising limb of the event and the main peak, the parameter is not of crucial importance for modelling in the context of flood warning. The graph shown in figure 7 reveals that the parameter *thick* is extremely sensitive, but as it

represents the segmentation of the soil for numerical purposes (equation 2) in the course of the solution of the Richards-equation it cannot be calibrated freely without putting at risk the model stability. Here an inexperienced user might get caught in a trap if only the sensitivity index is evaluated without taking into account the real meaning of the parameter.

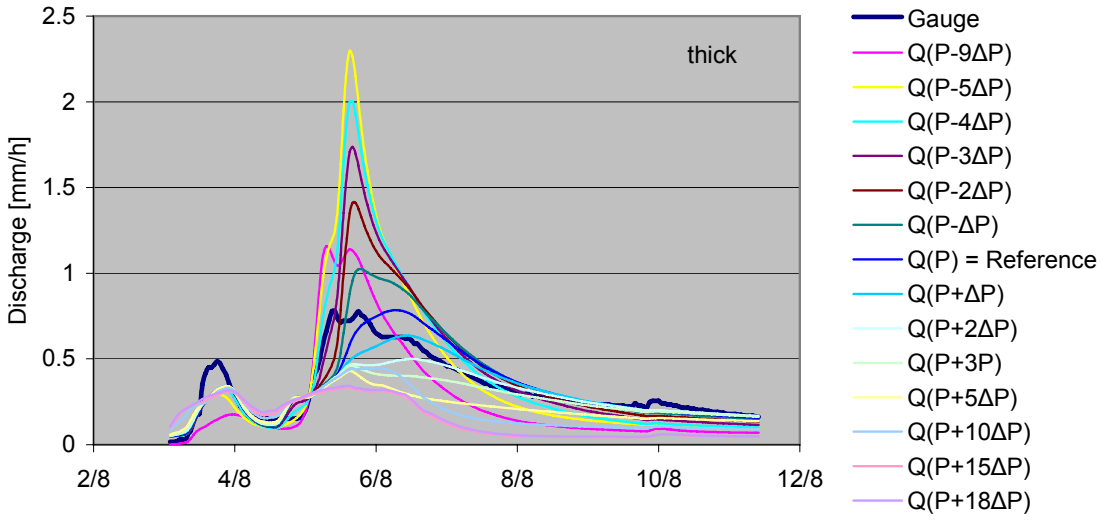


figure 7: Model sensitivity with respect to the parameter thick

2.1.3 Summarized results of the basic sensitivity analysis

The simple sensitivity analysis leads to general results for the test parameters: The parameter *krec* is not of first choice if the characteristics of rising flood events is the main task of a modelling study. The model is very sensitive to *dr*, *Ki* and *Kd* where the dominant processes are portrayed by interflow and direct runoff respectively.

A fundamental conclusion drawn from the study is:

➔ The simple sensitivity analysis only allows for roughly understanding the model sensitivity. Calculating different measures may lead to babylonic results, especially if the dominant processes change throughout the test period. The basic sensitivity analysis must always be judged with care and should be completed by a detailed check of the respective sensitivity graphs.

Taking into account this awareness, the results of the simple sensitivity analysis for just one single event seem inadequate to serve as the basis of setting up a flood forecast system which is expected to reliably portray all kinds of flood formation patterns.

As a consequence, DYNIA (dynamic identifiability analysis) is applied to WaSiM results in the following section. This approach allows for further investigating on the model parameter behaviour for different hydrologic conditions.

2.2 Model parameter identifiability in the light of dominant hydrological processes

For continuous simulations - as required in the forecast context - the optimum parameter set may change in time concurrently to variations in boundary conditions and process characteristics (Wriedt and Rode 2006). The event-based or subset-specific variance of best model parameter sets may result from uncertainty of input data, observation data, and equifinality of the system (Beven and Binley 1992). However, these parameter changes may also result from systematic changes of system behaviour, revealing model structural weakness caused by inadequate process representation or the failure of a single parameter set to fully portray rainfall-runoff dynamics for all types of flood events within a certain model structure. In this case, systematic relations of parameter optima and state variables exist. To approach this problem, it is necessary to shed light on the transient characteristics of best parameter sets and to link systematic changes to corresponding criteria (state variables). Determining the appropriate model parameters in relation to these transient dominant process controls has apparently not been sufficiently addressed by the research community and seems to be a major hurdle in the way towards a more successful online flash flood forecasting approach.

The DYNIA (Dynamic Identifiability Analysis) method presented by Wagener *et al* (2003) is a proposal to ease the identification of suitable parameter sets in the context of transient hydrological processes. It allows to trace temporal evolution of optimum model parameters and to relate parameter changes to state variables and process characteristics.

Thanks to these features, DYNIA is the perfect means for further investigating the model parameter sensitivity in the light of the hydrological process.

2.2.1 Study area and DYNIA methodology

The Rietholzbach catchment is analysed in this section, the watershed has been chosen in order to reduce the perturbing impact of data heterogeneity. It is a well-observed test catchment of the ETH Zürich, which drains a 3.18 km² hilly pre-alpine watershed. The average precipitation is 1600 mm per year, generating a mean annual runoff of 1046 mm. It is located in North-Eastern Switzerland, in the centre of the Thur basin with elevations ranging from 681 MSL to 938 MSL. The land use mainly consists of pasture (73%), the rest is covered by forest (23%) including some few settlement areas. The various soil types reach from gley soils to more permeable brown soils and regosols with relatively large soil water storage capacities. The catchment is equipped with a meteorological station, continuous time domain reflectometry (TDR) soil moisture measurements, a lysimeter, and a well-defined runoff profile at Rietholzbach gauging station. Data sets for the meteorological input parameters (temperature, humidity, wind, global radiation and precipitation) as well as flow data at the catchment outlet are available for the period 1981 through 1999. A calibrated WASIM version published in Schulla 1997; Schulla and Jasper 2001; Gurtz *et al* 2003 serves as a reference for simultaneously evaluating up to three parameters affecting the runoff formation of WaSiM.

Wagener *et al* (2003) developed the DYNIA algorithm (figure 8) as an extension of the regionalized sensitivity analysis (Spear and Hornberger 1980; Hornberger and Spear 1981). The analysis is based on a Monte-Carlo-Simulation (Buslenko and Schreider 1964) of parameters of the rainfall-runoff model.

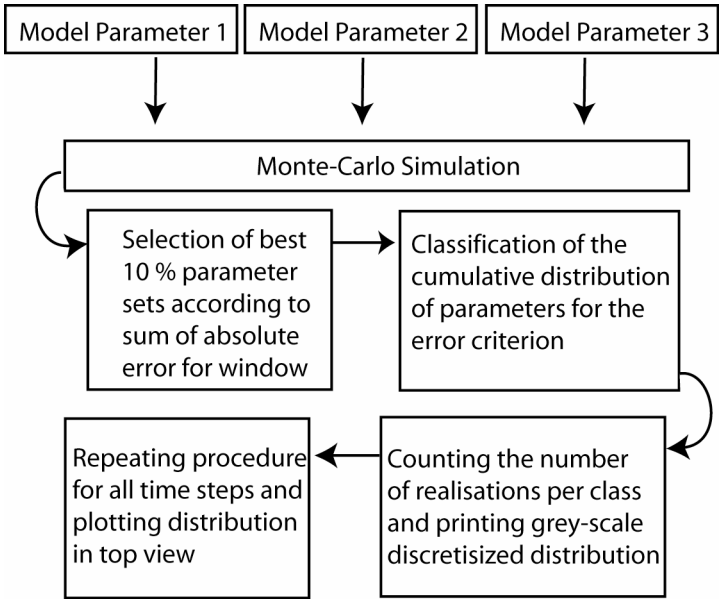


figure 8: Simplified flow chart of DYNIA steps for a 3 parameter analysis.

The considered parameters are either drawn by pure chance from within a specified parameter range, or by means of the Latin Hybercube method (Iman and Conover 1982). This latter approach is a fine way to economize on simulation time by assuring that the whole parameter range is sampled, whilst parameter realisations which are too similar are avoided (they do not yield additional information). Only the best 10% of simulations, according to an appropriate support measure (here: sum of absolute errors, SAE), are included in the analysis (figure 8). This pre-selection eases the interpretation of the considerable number of Monte-Carlo-Simulation results without loosing the important information on parameter identifiability. The basic idea is to calculate the probability distribution of the model parameters of the 10 % best simulations for each model time step within a specified time frame (moving window). The results are visualized in a 2D plot of parameter values vs. time, where the parameter probability density is shaded in a grey scale (the darker, the better is the identifiability).

In this study, additionally to the conventional DYNIA method, a modified approach is used to ease the visual assessment of state dependent changes of optimum model parameters ranges. In this modification proposed by Wriedt and Rode (2006), the data are re-ordered against an observed state variable, instead of using the original time series. This allows for running the DYNIA algorithm

for parameter combinations ordered by discharge or any other characteristic feature of interest.

2.2.2 Data, parameters and results

The DYNIA analysis requires a continuous simulation. The number of Monte-Carlo-Simulations and the high temporal resolution (1h) result in a considerable computation time of WaSiM, restricting the total simulation period to one calendar year. The hydrological catchment behaviour is evaluated from 01.04.1994 – 31.10.1994. This restriction to the summer period allows for analysing the runoff formation processes without the interference of snow storage and snow melt processes. WaSiM portrays these processes separately in the snow-module. The Monte-Carlo-Simulation is carried out using the set of the three most sensitive model parameters that have been identified in the basic sensitivity analysis (section 2.1). The considered parameters are controlling the direct runoff (*Kd*) and the interflow formation (*Ki*, *dr*). 1000 parameter sets are sampled from the parameter space (table 3) using a Latin Hypercube approach. The list of parameters and the associated parameter range is given in table 3. A window size of 101 hours is applied in the DYNIA analysis. Different objective functions (Sum of absolute errors – SAE, Nash-Sutcliffe efficiency – NSE and Index of Agreement – IoA) serve for generally characterizing the model accuracy.

table 3: Parameters included in the Monte-Carlo-Simulation

Parameter	Description	Calibrated Value	Range
Kd [h]	Recession constant for direct runoff	4	1–20
Ki [h]	Recession constant for interflow	8	1–20
dr [-]	"Drainage density", controls interflow	6	1–20

The parameter sets (table 4) used in the Monte-Carlo-Simulation show varying ability to portray the observed hydrograph (figure 9). The range between best and worst simulation is 0.18-0.80 for the Nash-Sutcliffe-efficiency If only the best 10% simulations are considered, the Nash-Sutcliffe ranges between 0.75-0.80 for worst and best simulations respectively. The dotty plots (figure 10)

suggest that the reference parameters (*Ki* = 4, *Kd* = 8, *dr* = 6) quite well describe the catchment hydrological response to rainfall. Most interesting is that the sum of the absolute error (figure 10) is related to the parameter *dr*. The lower margin of the dotty plot for *dr* shows a distinct behaviour over the parameter space. It therefore seems reasonable to not only confirm the results of the simple sensitivity analysis in that *dr* is a sensitive parameter, but it even seems possible to explain the parameter in the context of the hydrological process. For the parameters *Ki* and *Kd,* the shape of the scatter clouds (figure 10) do not suggest a clear functional relation between the sum of absolute errors and the parameter values, the overall identifiability is low. This also falls in with the insights of the simple sensitivity analysis. During the course of the Monte-Carlo-Simulations, there are presumably times without direct runoff. So here the parameter *Kd* fails to identifiably describe a portion of the discharge.

table 4: Error criteria for the best and worst parameter sets

Error criteria	Best set	Worst set
Sum of absolute errors [m³/s]	*175*	*445*
Index of Agreement [-]	*0.94*	*0.79*
Nash-Sutcliffe-efficiency [-]	*0.80*	*0.18*

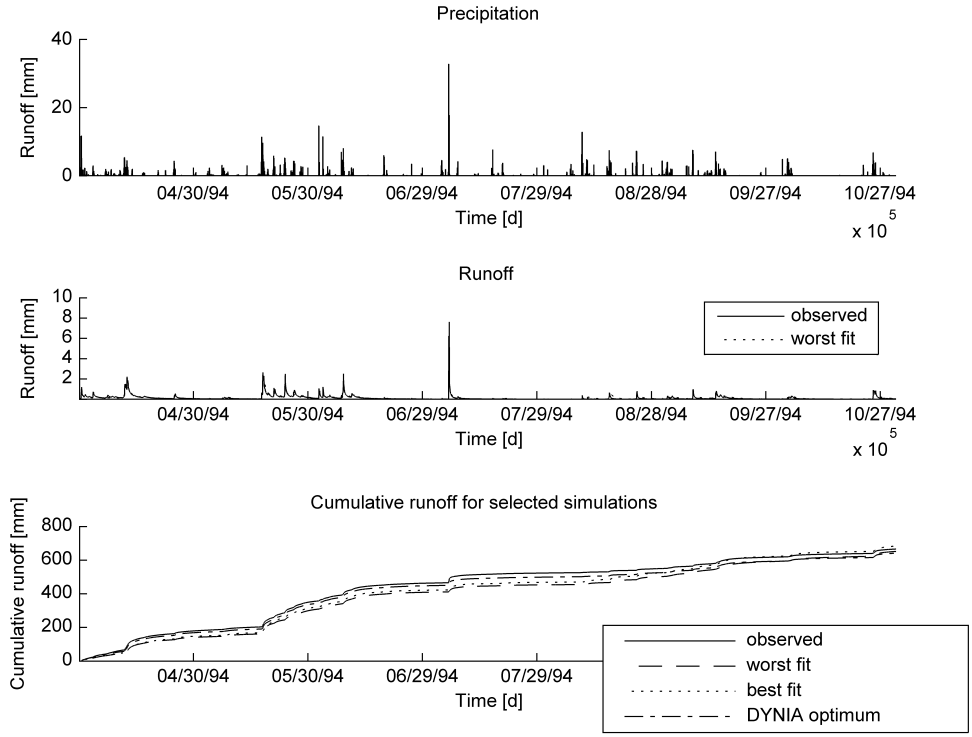


figure 9: Precipitation and runoff time series, cumulated runoff for observed data, best and worst simulation of Monte-Carlo-Simulation 1 and DYNIA optimum

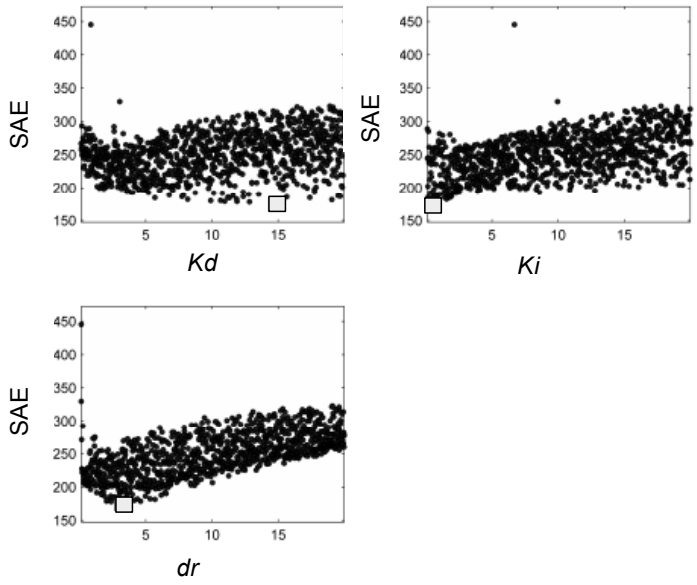


figure 10: Dotty plots of the parameter space Kd, Ki and dr for the sum of absolute errors (best parameters marked with grey dots)

The classical DYNIA method for the test period, plotted for the parameter *dr,* leads to the result shown in figure 11. Areas of high parameter identifiability (characterized by darker grey shades) alternate with periods where the parameter is almost not identifiable. It is difficult to capture the relation between the parameter and the discharge from this classical depiction of the DYNIA result.

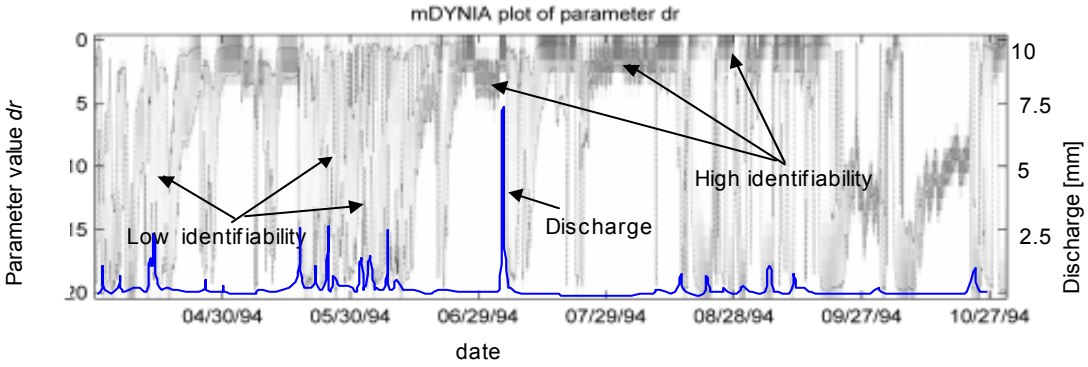


figure 11: The evolution of identifiability of dr along the test hydrograph

In figure 12 the DYNIA analysis for the parameter *dr* is reordered according to the discharge volume, i.e. the abscissa of the graph is sorted by discharge. This figure easily reveals a strong relation of *dr* and the observed discharge. The lowest discharges are considered first and the graph evolutes towards the highest discharge values for higher time steps to the right end side. This is visualized by means of the blue discharge function in figure 12. The low flow periods to the left of the graph in figure 12 generally coincide with *dr* parameter values below 5 (the dark grey shades concentrate near the upper end of the ordinate). This period is interrupted by various sections (time steps 500 to 2000 hours), where *dr* is best identifiable at values between 10 and 20. These fluctuations in parameter identifiability are attributed to reordering the results of the Monte-Carlo-Simulations. The parameter *dr* is highly identifiable already at the beginning of flood events, where the flow is still negligible, thus provoking the high parameter values even for low discharges. For runoff values of 1mm (time steps > 3500), the lower confidence limit increases from 0 to 3, while the upper *dr* limit is found at 15 (time steps 4000-4800 hours). For runoff data > 2mm, *dr* confidence limits cover the entire parameter range; the parameter is not clearly identifiable. In general terms: The optimum parameter range shifts towards

higher values with increasing discharge, but this coincides with increasing uncertainty for the highest discharges.

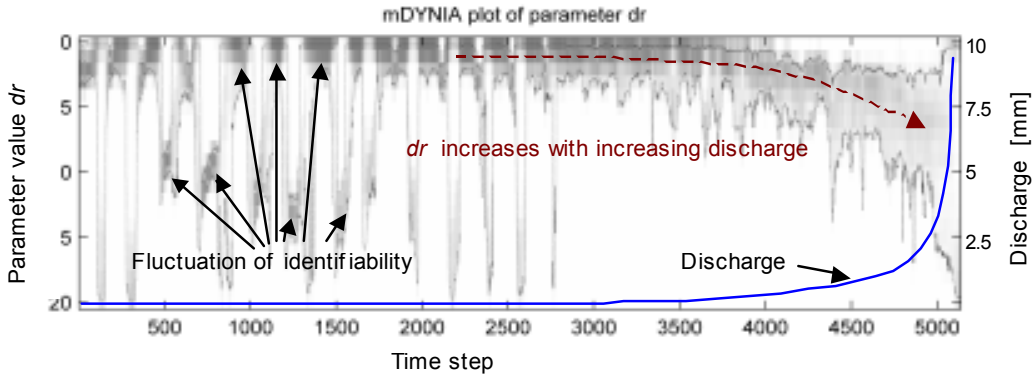


figure 12: dr parameter plot for time series reordered by observed discharge.

2.2.3 Summary of the DYNIA analysis:

The DYNIA analysis does not reveal any clear relations for *Ki* and *Kd*. As *Kd* controls fast surface runoff, increased identifiability is limited to single runoff peaks. When low flow situations prevail, Kd does not affect runoff, high uncertainty dominates the parameter values. Ki is characterized by a similar behaviour (figure 10). Neither for *Ki* nor *Kd*, system state variables correlate significantly, when applying reordered data series.

The observed relation of *dr* and discharge can be interpreted from a process-oriented view. As stated before, observed discharge can be taken as an integrative measure of the catchment wetness state. Soil moisture conditions considerably influence connectivity of subsurface flow pathways and the conductivity of the soil matrix: The hydraulic conductivity of matrix flow is directly related to soil water content and increases considerably with soil moisture above field capacity; preferential flow paths and highly conductive areas are linked with increasing soil moisture. This results in the fact that more lateral subsurface flow occurs in wet soils.

The general conclusion from this investigation of model parameter behaviour is: As each and every model is bound to the underlying assumptions and concepts, model parameter fidelity to the natural processes varies greatly over the course

of the hydrograph. This insight is based on the examination of figure 12. The *dr* parameter value is clearly depending on the observed discharge. Online flood forecasting suffers from a serious deficiency if just one set of model parameters describes the whole range of possible flood events. Variable parameter sets should therefore form the basis of improved flood forecasting. Despite the awareness that it is difficult to link best model parameters to a priori criteria, a further attempt evaluates the optimised parameters resulting from an automatic calibration approach statistically. Along these lines, more evidence for model parameterisation in the flood forecasting context is sought.

2.3 Relating model parameters to a priori knowledge of event pre conditions and characteristics

This section concentrates on linking optimal model parameters to a priori knowledge about the basin pre-event condition and/or the event characteristics for different patterns of flood formation. Therefore, event-based calibration leads to a number of optimal parameter sets, which are statistically analysed in the context of criteria describing the aforementioned characteristics. If best model parameters and criteria derived from a priori knowledge exhibit strong statistical links, this knowledge can be used for enhancing model parameterisation. In this study the Rietholzbach catchment (described in the preceding section) serves as test catchment. The reasons for this are: A reliable meteorological database is available. The considerable number of flood events and the good quality of the observed flow data ease the evaluation of parameter sensitivity.

2.3.1 Data and automatic calibration strategy

For the calibration study, 36 rainstorm events are chosen from the available database covering the years 1981 through 1999 (table 25 in appendix 1). Eight selected criteria describe the watershed pre-event condition and the characteristics of each rainstorm: The parameter *Duration* refers to the absolute length of the rainfall event. *Peak* denotes the peak rainfall rate, whilst *Volume* is the absolute volume of the rainstorm. *Form* is a combined parameter where the peak rainfall rate is divided by the absolute volume. *TI* is equivalent to the time

to rainfall peak divided by the peak intensity, this parameters describes the dynamics during the onset of the rainfall event. *TL* is a parameter describing the watershed retention and the rainfall characteristics by the time difference in rainfall and runoff peaks. This parameter cannot be calculated from a priori knowledge alone, but it serves for visually ensuring the consistency of the other parameter/criteria combinations. It is mentioned here in order to give a complete view of the considered criteria. *PF* is an exponential function describing the 14-day pre-event rainfall. *MPI* denotes the mean precipitation intensity for the corresponding event. Model calibration in this section is based on PEST (Parameter ESTimation). This package allows for automatic non-linear parameter estimation (Skahill and Doherty 2006).

Two approaches for parameter estimation are implemented in PEST; both methods try to minimize an objective function that is represented by least squares. The standard method uses the Gauss-Marquardt-Levenberg (GML) algorithm. This method is fast and stable as it switches between the steepest gradient search and the Gauß-Newton approach, depending on a scalar of the identity matrix (details in Mohamed and Walsh 1986). The drawback of this method is that it might "get stuck" in local minima, depending on the surface of the error and the start values for the optimisation run. The second method available in PEST is the global SCEUA search algorithm (DUAN *et al* 1992). This algorithm is capable of finding the global minimum of the objective function. It requires substantially more effort in terms of CPU time.

The pest set-up for automatic calibration is event based. WaSiM is calibrated for each of the 36 events independently, allowing at least four months for model warm-up (Cullmann *et al* 2006). This ensures that initial WaSiM conditions do not interfere with the evaluation of the calibrated parameters.

In order to check if PEST is reliably converging to the global minimum the shuffled complex evolution method (SCEUA) is tested against Gauss-Marquardt-Levenberg with various start values. This test is performed with data for event No. 4 (table 25 in appendix 1), which is chosen because its characteristic criteria are near to the mean criteria of all events. It is therefore considered a representative event well suited for a performance test of the PEST algorithms.

The results of this preliminary test are shown in table 5: They confirm the stable optimisation capability of both SCEUA and GML methods for the parameters considered with the start values used and the parameter ranges covered in this study. The difference in optimised parameters is marginal and leads to further employ GML as the standard method (GML is significantly faster than SCEUA). The optimised parameters for the test event listed in table 5 are close to the calibrated parameters used in section 2.2 (Kd = 4, Ki = 8, dr = 6). This confirms the representative characteristics of the test event used for checking out the convergence of the PEST optimisation. The error surface shown in figure 13 supports this statement. For a wide parameter range it shows a relatively smooth behaviour. Within the "near optimal" range, which is zoomed on the left side of figure 13, the gradient is also quite smooth and thus allows for using the GML method.

table 5: Test of convergence for the optimisation strategy. The variant adopted for the further testing is bold.

Method	Start Value			Parameter Range			Optimised Parameter		
	Kd	*Ki*	*dr*	*Kd*	*Ki*	*dr*	*Kd*	*Ki*	*dr*
	[h]	[h]	[-]	[h]	[h]	[-]	[h]	[h]	[-]
SCEUA	*	*	*	1-40	1-40	1-40	3.84	9.73	6.11
SCEUA	*	*	*	1-10^7	1-10^7	1-10^7	3.84	9.73	6.11
GML	**2**	**8**	**4**	**1-10^7**	**1-10^7**	**1-10^7**	**3.89**	**9.72**	**6.11**
GML	2	6	4	1-60	1-40	1-80	3.89	9.72	6.11
GML	35.9	27.3	29.9	1-40	1-40	1-40	3.89	9.72	6.11

*70 random points in parameter space

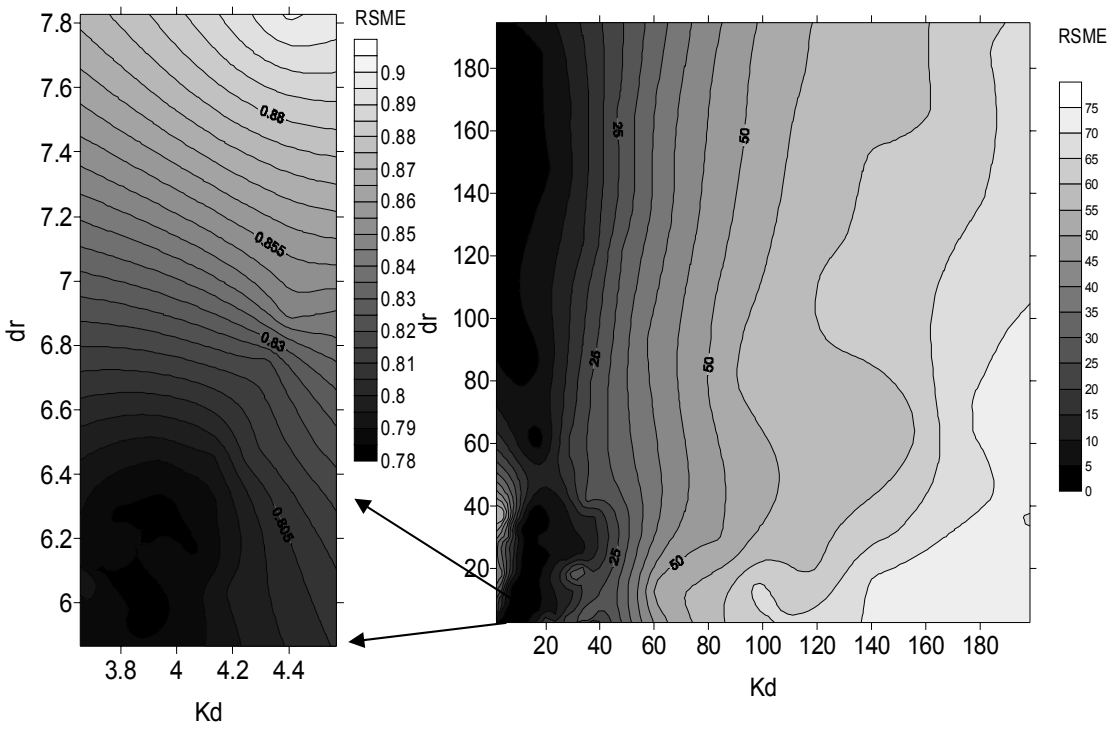


figure 13: Error surface of the parameters Kd and dr using SCE automatic parameter estimation

2.3.2 Automatic calibration and statistical analysis of the resulting parameters

The parameter estimation presented here is characterized by large upper boundaries (the variant used in this study is shaded in table 5). Nevertheless the optimal parameters for the 36 events listed in table 6 concentrate in the range of 1-100. Most optimisations yield small errors in the peak flow rooted mean square error (RSME) shown in the table. Some realisations for the parameter *dr* exhibit values of 200 and more, this often coincides with one or two of the other parameters drifting towards the lower boundary. In such cases of "extreme" parameter combinations, erratic input/flow data and/or model structural deficiencies lead to questionable realisations of the investigated parameters. This is also expressed by the extreme RMSE of these events. Therefore the parameter combinations of events 9, 12 and 22 are excluded from the further analysis of the results.

table 6: Results of automatic parameter estimation for the considered events.

Event	Year	dr	Ki	Kd	RMSE [%]	Event	Year	dr	Ki	Kd	RMSE [%]
1	1981	3.00	3.60	52.40	0.14	19	1994	9.00	20.50	1.60	0.01
2	1982	1.00	1.00	2.50	15.87	20	1994	8.50	4.90	1.40	0.10
3	1982	72.80	1.40	177.00	1.49	21	1994	5.70	5.80	1.40	0.16
4	1984	6.10	9.70	3.90	0.00	22	1994	225.90	1.00	2.20	16.26
5	1984	8.00	20.50	2.10	0.01	23	1994	82.30	4.80	35.40	0.63
6	1985	2.50	2.10	6.70	0.01	24	1994	1.00	5.20	1.00	0.00
7	1985	35.40	17.70	1.00	0.00	25	1994	18.30	1.00	1.00	0.07
8	1986	5.70	14.00	4.30	0.17	26	1994	15.30	1.00	1.50	0.00
9	1986	105.20	1.00	3.50	14.32	27	1994	20.00	3.30	4.40	0.00
10	1987	22.80	3.40	165.70	0.22	28	1994	5.00	10.00	2.10	0.00
11	1987	5.00	10.00	2.00	0.16	29	1994	2.30	7.90	2.70	0.00
12	1989	235.00	1.00	1.00	4.51	30	1994	1.90	2.80	4.20	0.00
13	1990	4.90	2.70	6.30	0.09	31	1994	5.70	10.40	4.40	0.00
14	1991	5.90	15.80	4.20	0.13	32	1994	21.00	7.50	8.50	0.00
15	1991	7.30	14.60	6.00	0.25	33	1995	5.10	11.20	1.00	0.04
16	1993	14.10	1.00	8.00	0.36	34	1995	6.10	7.90	2.10	0.00
17	1993	1.00	6.90	3.70	0.09	35	1996	5.00	42.80	2.10	0.01
18	1993	1.50	14.40	7.60	0.01	36	1999	12.50	3.70	1.00	2.26

The results shown in table 6 are analysed for eventual dependencies of optimal parameter sets on seven of the eight chosen characteristic criteria listed in table 25 (appendix 1). *TL* is excluded from the analysis because it is unsuitable for a predictive description of event characteristics (it is only known after the event has already taken place). This criterion had been used for a visual check in order to detect if any information regarding the rainfall-runoff transformation is not accounted for by the other seven criteria which can be calculated from precipitation forecast alone. It proved impossible to detect a satisfying regression or functional relationship between the parameters in table 6 and the criteria, thus quantitative statements about the setting of model parameters

according to a priori knowledge seems difficult. This partly confirms the findings of the DYNIA analysis described in section 2.2., here, no rule could be found for describing the parameters *Ki* and *Kd* by means of the actual discharge. To nevertheless judge weather the optimised parameters yield qualitative information that can be deduced from the characteristic criteria, the latter are divided into three equal classes each, namely: Low, medium and high.

ANOVA (analysis of variance) is first selected to investigate possible relations of criteria classes and optimised parameters. In preliminary examinations, a check of homogeneity of variances was applied to all the parameters and criteria. This test was negative for all parameters. Due to this general rejection of the assumption of homogeneity of variances, ANOVA cannot be applied. Instead, the Mann-Whitney-U test is used to interpret the results of the optimised parameter *dr* and the eventual dependencies on classes of characteristic criteria. The non parametric Man-Whitney-U test has less statistical power than ANOVA, but is accepts inhomogeneous variances. With this test, it is possible to compare two criteria classes, concerning their mean and median values for *dr*. It measures how much the average rank of one class differs from the average rank of another class. The parameter *dr* is selected for this analysis because visual control of box plots for the classes indicated a possible relation between the three criteria classes and the parameters only for *dr*. *Ki* and *Kd*, just like in the DYNIA analysis, seem unsuitable for being linked to a priori knowledge, even in the simplified form of classified criteria. To avoid the effect of possible parameter interaction within the best sets, the optimisation is repeated for the parameter *dr* alone. Fixed values of $Ki = 10$ and $Kd = 4$ are assumed. (optimised parameters for the typical event in table 5). The results of this second parameter estimation are shown in table 7. The RSME is calculated for the peak flow value. The results of this optimisation resemble the results for three free parameters listed in table 6 - with the exception of event 25. *Ki* and *Kd* are forced to the lower boundary in the three parameter optimisation. Lacking this possibility with fixed *Ki* and *Kd* in the second run, the RMSE decreases notably. Some of the error criteria are lower in table 7 than for the corresponding event in table 6. Where this difference is significant, one parameter is forced to the lower boundary or one parameter is set to a very high value. This effect might reflect the existence of local minima for events which cannot be easily modelled with a meaningful parameter range. Small differences

in RMSE can be attributed to rounding effects or early stopping of the optimisation algorithm.

The optimal parameter means and medians for the classified criteria are shown in table 8. The parameter means do not reveal general trends in relation to the classified criteria. However, when observing the medians, highest *dr* values coincide with the highest criteria classes. Only for *TI* this is not the case. That is a consequence of the peak flow intensity in the denominator of *TI*. It therefore behaves inversely proportional to the other criteria and what seems to be a deviant behaviour at first sight turns out to confirm the general trend.

table 7: Results of the optimisation of dr with fixed Ki and Kd

Event.	Year	dr	RMSE	Event.	Year	dr	RMSE	Event.	Year	dr	RMSE
1	1981	3.43	0.04	13	1990	10.58	0.07	25	1994	17.5	18.40
2	1982	1	15.73	14	1991	7.7	0.09	26	1994	33.81	0.00
3	1982	39.46	1.23	15	1991	11.58	0.17	27	1994	19.97	0.00
4	1984	6.18	0.00	16	1993	16	0.19	28	1994	4.61	0.00
5	1984	21.84	0.01	17	1993	1.41	0.24	29	1994	1.58	0.00
6	1985	6	0.00	18	1993	8.46	0.00	30	1994	2.53	0.00
7	1985	24.2	0.01	19	1994	5.95	0.02	31	1994	8.05	0.00
8	1986	9.92	0.12	20	1994	2.53	0.18	32	1994	32.03	0.00
9	1986	82.93	13.54	21	1994	2.06	0.22	33	1995	1	0.93
10	1987	13.57	0.09	22	1994	33.62	7.36	34	1995	3.76	0.03
11	1987	16.61	0.28	23	1994	26.15	0.00	35	1996	3.32	0.04
12	1989	249.32	4.55	24	1994	1	0.21	36	1999	5.65	2.69

table 8: Mean and median of dr values of the grouped characteristic criteria

dr		Duration [h]	Peak [mm/h]	Volume [mm]	Form [h]	TI [mm]	TL [h]	PF [mm]
Mean	Low	12	16.4	13.4	13.4	13.6	6.7	14.5
	Medium	14.6	5.7	15.3	15.3	6.6	3	8.1
	High	11.9	14.7	8.4	8.5	17	6.6	14.4
Median	Low	5.7	6.7	5.1	5	8.5	5.9	5.9
	Medium	5.6	5	7.3	7.3	5.7	5.4	5.7
	High	6.1	9.3	7.4	7.4	6	5.9	5.8

table 9: Results of the Mann-Whitney-U test

Criterion	Classes	σ [-]	Criterion	Classes	σ [-]
Duration	low/med.	0.85	TI.	low/med.	0.38
	med./high	0.67		med./high	0.45
Peak.	low/med.	0.15	PF	low/med.	0.84
	med./high	**0.13**		med./high	0.81
Volume	low/med.	0.2	MPI	low/med.	1.0
	med./high	0.34		med./high	0.49
Form.	low/med.	0.19			
	med./high	0.34			

The results of the Man-Whitney-U test are summarized in table 9. It shows the statistical power σ for low/medium and medium/high classes. Generally, it holds that for growing σ, the statistical power decreases. It is of minor interest for parameter characterisation if the low and high classes show any statistically robust behaviour, therefore they are not included in the table. From this test we learn, that the peak rainfall rate shows the most powerful statistical link to *dr*. If we assume a 15 % confidence level ($\sigma < 0{,}15$), the difference between the medium and high classes are statistically significant. With a median of $dr = 9.3$ the high peak rainfall class exhibits the highest overall value for *dr*, i.e. the

higher the peak rainfall, the larger is the optimal *dr*. This result must be seen cautiously in the light of an extremely low *dr* value for the middle container of the peak classification. This probably reflects the non-linear behaviour of model errors with respect to the process representation inherent in the modules of WaSiM. For a more clear statement further research is needed. But even – if we would assume that the middle container is not representative and replace it by the lower container - the difference between the low and high containers is still significant at a statistical level, confirming the general conclusions made here. The differences between low and medium classes become significant for *Peak, Volume* and *Form* on the 20 % confidence level. *Form* is a parameter evaluating *Peak* and *Volume* jointly. From this second result of the Man-Whitney-U test another general conclusion yields: The more voluminous the rainstorm event is, the larger is the optimal parameter *dr*.

2.3.3 Summarized findings of the statistical analysis of optimised parameters

A considerable span of best parameter sets for the event based calibration is observed for the different classes of pre-event conditions, characterized by eight selected criteria. It was impossible to find quantitative functional relations linking the parameters *dr*, *Ki* and *Kd* to the chosen criteria by means of the calibration study. For the investigated catchment it is therefore impossible to derive a priori best parameters with the aid of functions taking into account the chosen criteria. On the other hand, a more detailed study focusing on *dr* revealed that there is a distinct qualitative relation between the event characteristics and the best parameter. For the case of peak rainfall intensity, this relation is strong and can be statistically proven. The positive relation between *dr* and the classified criteria describing the volume of the rainstorm is also notable (table 24 in appendix 1). Generally, applying statistics to calibrated parameters can lead to an improved understanding of the degree of dependency of model parameter sets to criteria derived from widely available information. This gives a means of assessing model weaknesses regarding the sensitivity to characteristic inputs -e.g. rainstorms of different intensities, but equal volumes- not accounted for by singular sets of parameters.

2.4 Consequences of model parameter investigation

The results of the basic analysis (section 2.1) are:

- ➔ *Ki* and *Kd* are sensitive where the respective processes of interflow and direct runoff dominate the hydrograph;
- ➔ *krec* and *thick* are not well suited for model calibration in the flash flood forecasting context;
- ➔ *dr* is a very sensitive parameter well suited for calibration purposes.

The more detailed DYNIA sensitivity analysis (section 2.2) reveals:

- ➔ The parameters *Ki* and *Kd* are difficult to identify over the course of the hydrograph. It seems hard to link them to information about the pre-event condition or the observed discharge;
- ➔ In spite the considerable uncertainty characterizing *dr* for the highest discharges, it is generally a well identifiable parameter and positively related to the observed discharge.

The automatic parameter estimation and statistical analysis (section 2.3) results in the statements:

- ➔ The optimal value of *dr* is linked to the peak rainfall rate and the volume of the concerned rainfall event;
- ➔ This relation only becomes clear for classified criteria describing event characteristics and watershed pre-event characteristics.

DYNIA and the automatic calibration are based on the same catchment with the same data. For the 1994 DYNIA period, 13 events have been calibrated separately in section 2.3. It is therefore possible to directly compare the results of DYNIA with the parameters resulting from the automatic calibration. The graph in figure 14 reveals that in some few cases the highest DYNIA parameter identifiability differs significantly from automatically calibrated parameters for single events.

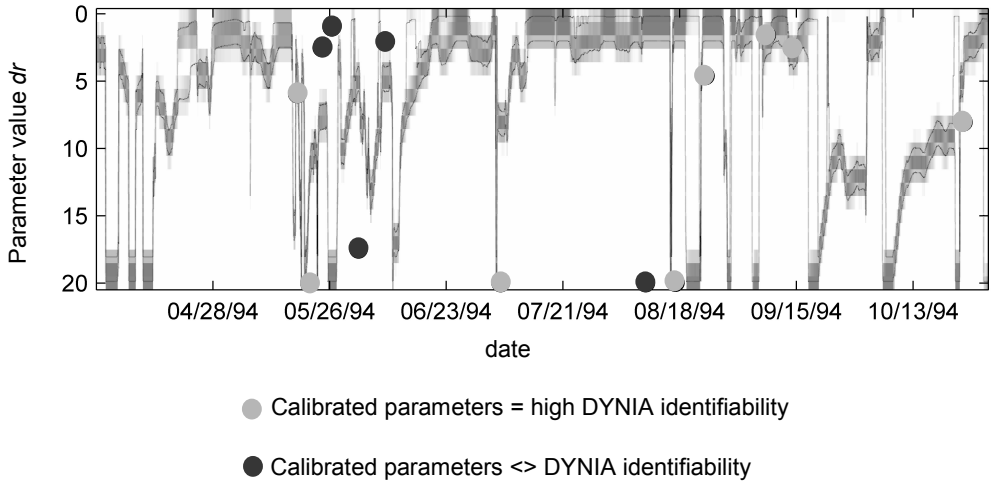


figure 14: DYNIA parameter identifiability compared to the optimal dr values from automatic calibration

A possible explanation for this is the DYNIA window size, which might level out peak errors in parameter identifiability. However, from the figure it also becomes clear that in most cases high DYNIA parameter identifiability coincides with the results of event based optimisation. The results are generally coherent and confirm the qualitative dependency of model parameters on the specific hydrological characteristics of the modelled event/period.

As a consequence of this awareness a new parameterisation strategy for online flood forecasting models is sought. To this end, it is favourable to introduce qualitative information regarding the dominant hydrological process into the calibration procedure of models. Here it is important to avoid merely adding more events for calibration purposes: This is but a quantitative improvement of the calibration. Such a step always has to be paid for by a loss in predictive performance for rare cases (the extreme flood events). Instead, the use of multiple parameter sets for different classes of flood formation characteristics is the way forward. In the next section, such an approach is exemplarily developed.

2.5 A new approach for model parameterisation: Considering the specific relevant characteristics of different flood types

The preceding chapters lead to the awareness, that state of the art hydrological models cannot satisfactorily portray the whole spectrum of naturally occurring runoff formation patterns with a singular set of parameters. Nevertheless, the singular parameter set represents the “classical” way for modelling rainfall-runoff processes in the context of flood forecasting today. In this chapter, a new approach for model parameterisation offers the possibility to introduce additional information about basically different flood types into the model set-up. WaSiM is used for an exemplary investigation, comparing the performance of the singular parameter set modelling approach with the results of the new strategy. The study is carried out in the catchment of the Schwarze Pockau River at Zöblitz gauging station. The area and data used here are described in section 2.1. First, a singular event is calibrated and then validated using periods not considered in the calibration process. This is put side by side to the new approach: dynamic parameterisation. In this approach, the basin specific flood patterns are divided into two classes on the basis of the magnitude of the flood peak. This classification reflects the experience (sections 2.2 and 2.3). that best model parameters depend on volume and intensity of the event and vary for different classes of flood formation. For each class, one event is used for calibration. Next, the model is executed for all validation data with both parameterisations. The results of this “double” model application are then superposed by means of a sigmoid function to form one single hydrograph.

Section 5.2 of this study confirms this approach on broader basis. In that section the same strategy is applied to another catchment (10 times larger) with comparable results for 13 events in two classes.

WaSiM is used for single events with a model warming period of 4 months according to Cullmann *et al* 2006. This ensures that the model internal state initial conditions do not interfere with the evaluation of the calibration and validation performance. The events used in this section are shown in table 10.

table 10: Flood events at Zöblitz gauging station

Event Nr.	Date	Peak Flow [m³/s]
1	20.06. - 01.08.1958	34.04
2	02.05. – 23.05.1978	23.65
3	18.07 – 08.08.1980	30.46
4	29.07 – 26.08.1983	27.59
5	27.05 – 01.07.1995	24.00
6	29.06 – 27.07.1996	15.05
7	07.03 – 04.04.1998	22.58

2.5.1 The singular parameter set approach

The 1983 flood event is chosen for calibration. It is a typical flood event triggered by an advective rainfall field. The manual calibration procedure concentrates on the processes describing the flood formation. The corresponding model parameters are *dr, Ki and Kd.* Their impact on the WaSiM runoff formation module is shown in figure 2. The calibration procedure leads to the result shown in figure 15. The model well portrays the observed hydrograph. The first peak is underestimated, but the model well describes the main peak (the most important feature for a flood forecasting system) and recession phase of the event. The calibrated parameters are: $Kd = 9$, $Ki = 16$, $dr = 37$. The model is now validated with the events not used in the calibration (events 1-3 and 5-7 in table 10). The validation performance for the peak values is visualized in figure 16. The model slightly under-predicts most of the events. 1995 is an exception from this rule. Here the observed discharge is smaller than the modelled. The largest errors occur for the 1980 and 1998 events with a relative peak error of 6.6 % and 4.6 % respectively. The fact that the overall relative validation error is only 2,8 % leads to the awareness that at least the 1980 event – relative peak error is more than 200% compared to the average - is not well portrayed by the model. A way of improving the model performance for the 1980 event without losing the good fit for the rest of the events is therefore sought after in the next section.

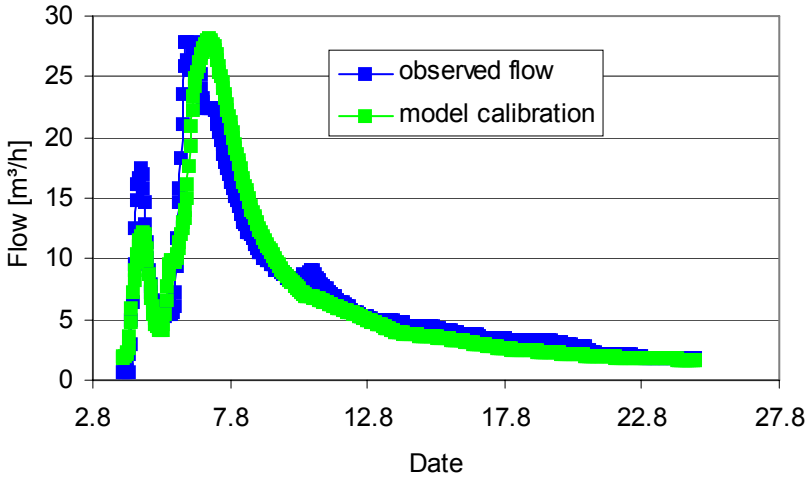


figure 15: Manual calibration for the 1983 flood event at Zöblitz gauging station

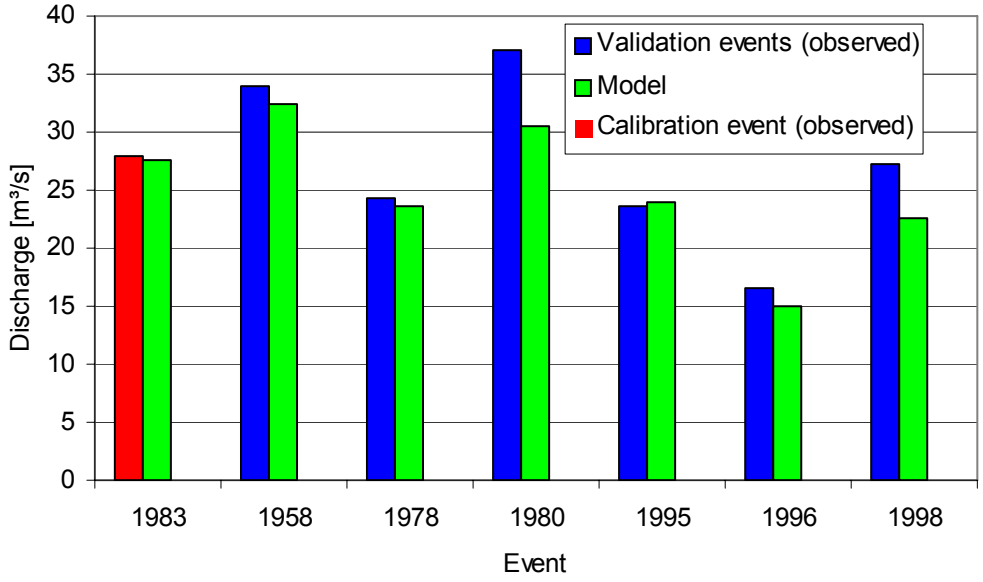


figure 16: Singular parameter set calibration and validation performance of WaSiM for various peak flow values

2.5.2 The improved strategy: Dynamic parameterisation.

Two of the three events with the worst validation performances for the singular parameter set approach are characterized by one common property. These events

are the largest in terms of the peak discharge values. Most likely these two events are governed by processes that are not of predominating importance for the 1983 event which has been used for calibration. This explains the poor validation performance. Therefore these two events (1958 and 1980) are separated from the rest of the events which are remaining untouched in the following process of setting up the dynamic parameterisation. The 1958 event is recalibrated and the model performance is validated with the 1980 event. The model performance for these two events (table 11) shows that the error is reduced for both events in the recalibrated version. For 1958 this is a truism because the event is calibrated. More important is that the relative peak error for 1980 significantly decreases from 6.64 % to 1.98 %. This is - besides their higher peaks - another strong indicator for the fact that these two events differ from the rest in terms of the dominating processes. For the 1998 event no such relation was found. It is most likely that the poor performance for this event is the consequence of input data quality. This result recommends taking into consideration the specific characteristics of flood events, originating from basically different patterns of flood formation.

table 11: Relative peak errors for 1958 and 1980 events for the singular parameter set and "dynamic" approach respectively

Relative Peak Error [%]			
Event	Validation "singular"	Recalibration	Validation "dynamic"
1958	1.59	0.2	-
1980	**6.64**	–»	**1.98**

Considering various flood types for model calibration yields a principal disadvantage: The model has to be run simultaneously for all parameter sets. The drawbacks are obvious: Longer CPU times and various model results which may lead to confusion and difficulties in the evaluation. In the exemplary case, two completely different time series result from the two considered classes: One for the 1983 parameterisation and the other for the 1958 event parameter set. This is inconvenient because the operator has to decide which model results are representative during the forecast. A new task arises: the results of the model runs have to be appropriately merged to characterise an unequivocal flood

hydrograph (Q) for all considered events. Fusing the output of the two model parameterisations with a sigmoid-type function (figure 17) solves the task. The exact formula used in this study is defined as:

$$Q = Wn * Qn + We * Qe \qquad \textbf{(equation 7)}$$

with Wn = the weight for the smaller event class (1983) model output

Qn = discharge of the model calibrated for the smaller event class

We = the weight for the extreme event class (1958) model output

Qe = discharge of the model calibrated for the extreme event class

The weights for the merging process are defined:

$$We = \frac{1}{1 + e^{-\frac{Qe-\mathrm{P}}{\Lambda}}} \qquad \textbf{(equation 8)}$$

$$Wn = 1 - We \qquad \textbf{(equation 9)}$$

with P = the threshold for the parameterisation. This parameter defines the intersection of the merging sigmoid. It can be interpreted as the threshold between the extensions of validity of the two model parameterizations. For the application with the test data P = 30 m³/s marks the transition point between the two parameterisations. Λ is the parameter which defines the steepness of the sigmoid. With increasing Λ the sigmoid becomes shallower arched. For our application we use a steep sigmoid (Λ = 0.4). This allows for sharply distinguishing between the two event classes.

The model performance for calibration and validation of the univocal model is shown in figure 18. The comparison with figure 16 reveals significant model improvement in portraying the peak flow. The relative RMSE for the peak flow is 2.6 % for the singular parameter set, while the dynamic model parameterisation approach yields 1.8 % RMSE for the validation.

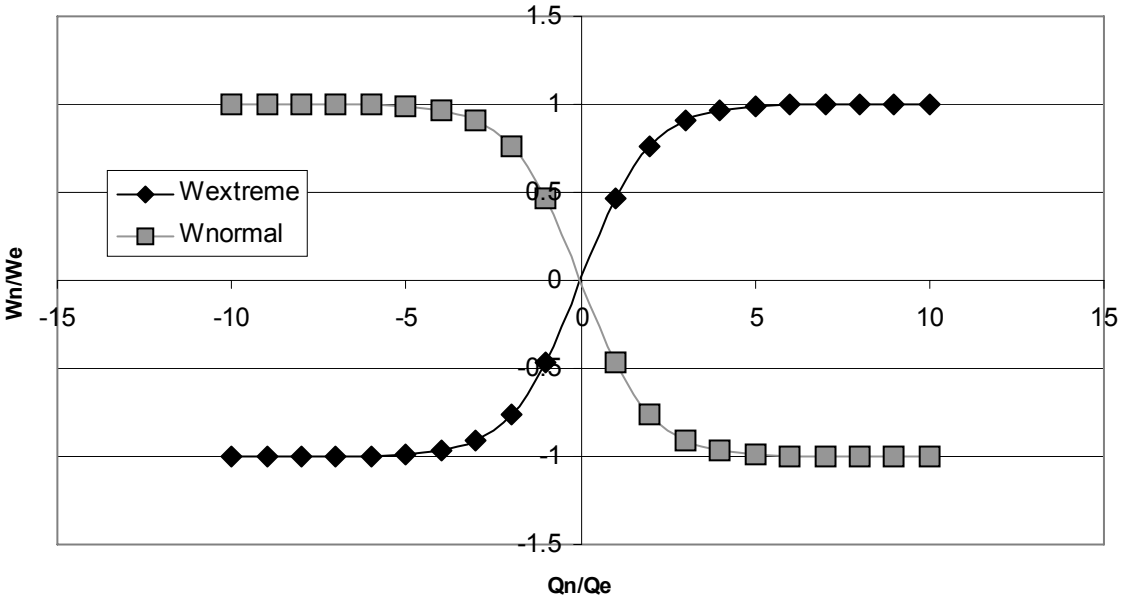


figure 17: Exemplary sigmoid function for the merging of different parameterisations

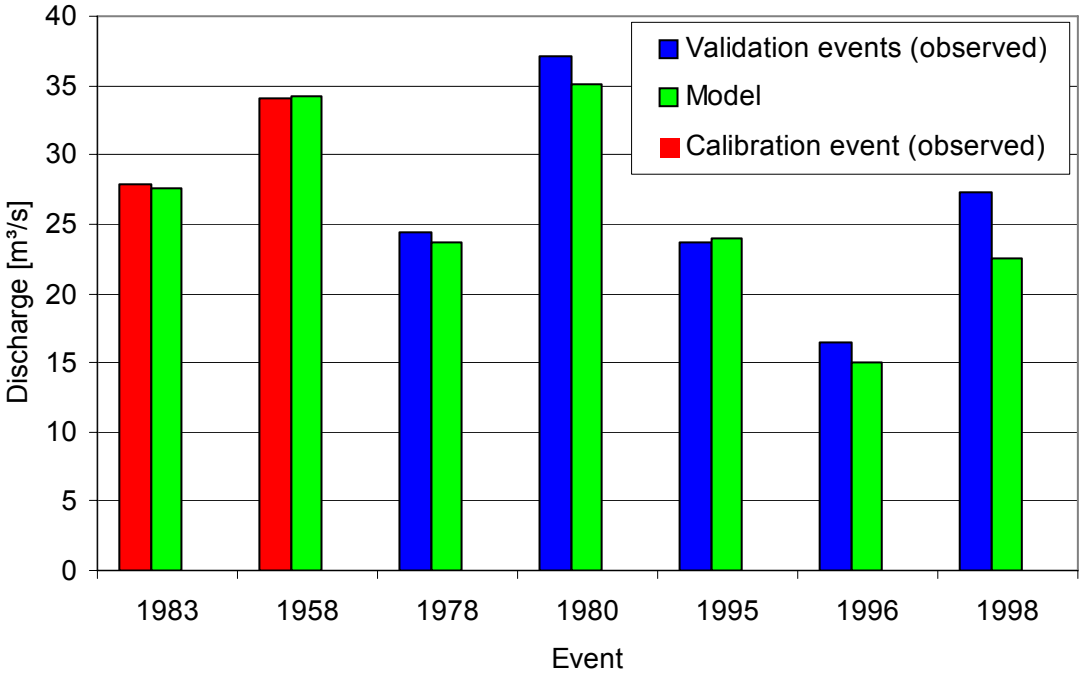


figure 18: Validation and calibration performance of merged model

This result, confirmed by the successful application of the dynamic parameterisation in section 5.2, recommends the strategy as a general means for improving model set-up in the context of flash flood forecasting. With the help of this new parameterisation strategy it is possible to describe the flood formation process with multiple parameter sets. If more than the two sets shown here are used, a multiple sigmoid function or a fuzzy weighting system can be used for the merging process.

3 Portraying rainfall-runoff processes by artificial neural networks

Various types of artificial neural networks are currently being applied for rainfall-runoff modelling purposes (section 1.2). After first introducing the most widely used class of artificial neural netwoks - the multi layer feed forward net - this chapter explains polynomial neural networks. The focus is set to the structure of both net types as well as their specific training procedures.

3.1 Multi Layer Feed Forward Nets

Multi layer feed forward nets (MLFN) are widely applied for rainfall-runoff modelling. As outlined in section 1.2, they have been introduced into the flood forecasting context in the 1990`s. Lately, Schmitz *et al* (2005) and Cullmann *et al* (2006) promoted this net type in their flood forecasting approach. This study evaluates their performance for flash flood forecasting .

3.1.1 Principles of multi layer nets

The principal functioning (figure 19) of multi layer nets is based on the neurons in the hidden and output layers. They transform their respective inputs to outputs through two separate stages. First, for each neuron, each of its inputs is multiplied by the corresponding weight (w) and secondly, the total sum of these products plus a constant known as bias (b) yields the node output in the hidden and output layers.

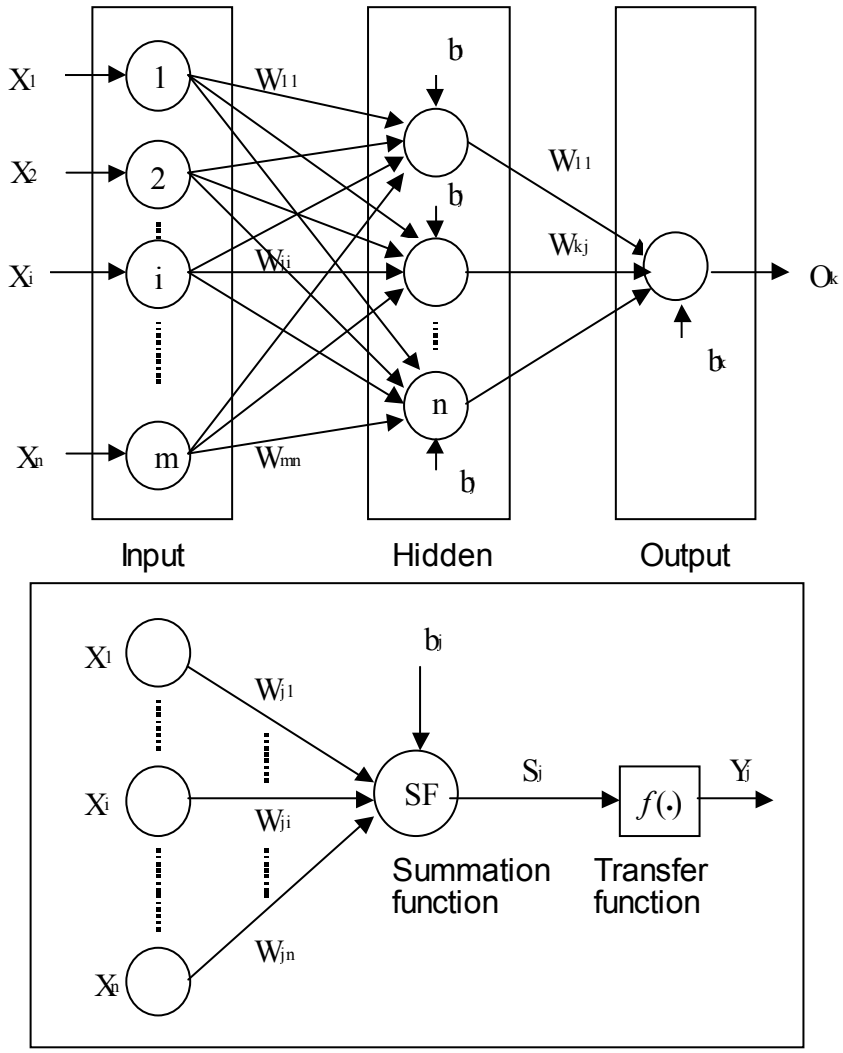


figure 19: Typical three-layer network structure (top) and operating scheme (bottom)

An elementary neuron with n inputs is shown in the lower part figure 19. Each input is weighted with an appropriate *w*. The sum of the weighted inputs and the bias forms the input to the transfer function *f (.)*. Any differentiable transfer function *f (.)* is suited to generate output. The output S from the j^{th} node in the hidden layer - after the summation operation - is defined as follows:

$$S_j = \sum_{j=1}^{n} W_{ji} X_i + b_j$$

(equation 10)

with: n = Number of elements in the vector

X_i = the input signal from i^{th} node

W_{ji} = MLFN weights

b_j = MLFN bias

Then the net output Y_j from the j^{th} output node is:

$$Y_j = f(S_j) \qquad \textbf{(equation 11)}$$

3.1.2 Structure of multi layer neural networks

Network geometry determines the number of connection weights and how these are arranged. Generally, a fix number of hidden layers is defined and the number of nodes in each layer is then chosen. It has been shown that three layer networks (1 input -1 hidden - 1 output) with sigmoid transfer functions in the hidden layer and linear transfer functions in the output layer can approximate virtually any function of interest to any degree of accuracy, provided that a sufficient number of neurons are available in the hidden layer (Hornik *et al*, 1989).

The number of nodes in the input layer is determined by the number of input vectors, whereas the number of nodes in the output layer equals the number of model outputs. The critical aspect is the choice of the number of nodes in the hidden layers and hence the number of connection weights. The importance of finding a balance between having sufficient free parameters (weights) to enable representation of the function to be approximated and having too many free parameters, which can result in over-fitting, is well known and has been discussed widely in the literature (e.g. Maren *et al*, 1990; Rojas, 1996). The number of hidden layer nodes significantly influences the performance of a network: with too few nodes the network will approximate poorly, while with too many nodes it will over-fit the training data. Consequently, an optimally designed hidden layer geometry

- ➔ Reduces the computational effort necessary for training the net;
- ➔ Ensures the best possible generalization performance;
- ➔ Avoids the problem of over fitting;

No unified theory exists for determining optimal network geometry. In the present study three layer feed forward networks are used. Following Hornik *et al*, (1989), this ensures the required net performance in the flood forecasting context. The number of input neurons equals the number of input vectors while one fixed node represents the model output.

The transfer function converts the effective incoming signal of node j, S_j (equation 11) into the output signal (Y_j). Multi layer nets typically use sigmoid transfer functions in the hidden layers. These functions are often called "squashing" functions, since they compress an infinite input range into a finite [-1; 1] output range. In this study, non-linear bipolar tan-sigmoid transfer functions are used for hidden layer nodes (equation 12), whereas linear transfer functions are characterizing the output layer node.

$$f(S_j) = \frac{2}{1+\exp^{-S_j}} - 1 \qquad \textbf{(equation 12)}$$

with: , S_j = input

$f(S_j)$ = output of the transfer function.

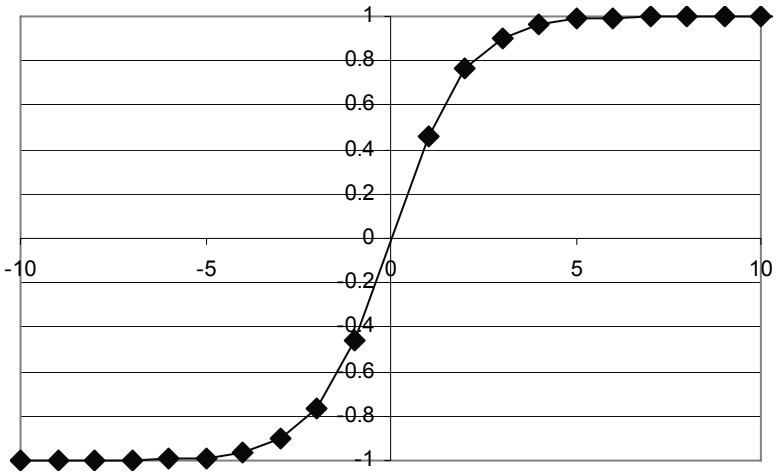


figure 20: tan-sigmoid function (equation 12) used in the study

The characteristics of the tan-sigmoid function are:

- ➔ Upper and lower bounds exist;
- ➔ It is monotonically increasing;
- ➔ It is continuous and differentiable everywhere.

The simple linear transfer function used for the output layer of the MLFN is expressed as:

$$f\left(S_j\right) = S_j \qquad \textbf{(equation 13)}$$

with: S_j = input

3.1.3 Training of multi layer nets

The optimisation of the multi layer network weights is known as 'training' or 'learning'. This process can be compared to the parameter estimation phase of conventional models. The aim is to find a global solution to what is typically a highly non-linear optimisation problem (White, 1989). Therefore, the general theory of non-linear optimisation just seems to be the method of choice

(Battiti 1992). The suitability of a particular method is generally a compromise between computational cost and training performance (Parisi *et al*, 1996).

There are numerous approaches for network training. The most popular algorithms are based on supervised training, which has also been used in the present study. Paradigms of supervised learning include error-correction learning, reinforcement learning and stochastic learning. An important issue concerning supervised learning is the problem of error convergence, i.e., the minimization of the objective function. The aim is to determine a set of weights which minimizes the mean square error. One well-known method, which is common to many learning paradigms, is the least mean square convergence, which was adopted in the present study.

After a careful and thorough investigation of different training algorithms, the Levenberg-Marquardt back-propagation algorithm (Hagan *et al*, 1994), was chosen for optimising the weight and bias parameters during the training process. It is a quick and stable second order non-linear least square technique (Toth *et al,* 2000).

The Levenberg–Marquardt algorithm is a modification of the classic Newton algorithm for finding an optimal solution to a minimization problem. It is designed to approach second-order training speed and accuracy without having to compute the Hessian matrix. The Hessian matrix contains second derivatives of the network errors with respect to the network weights, while the Jacobian matrix contains first derivatives of the network error matrix with respect to weights. If the performance function has the form of a sum of squares (as is typical in training feed-forward networks), then the Hessian matrix can be approximated as:

$$H = J^T J \qquad \textbf{(equation 14)}$$

with: J = the Jacobian matrix and H = the Hessian matrix

and the gradient can be computed as:

$$g = J^T e \qquad \textbf{(equation 15)}$$

with: e = a matrix of network errors.

3.2 Polynomial Neural Nets

Shin and Gosh (1992), Foka (1999) and Ma and Khorasani (2005) employ PoNN (Polynomial Neural Networks) for time series prediction and achieve promising results. Their work points towards the possibility of using polynomial nets as a Taylor approximation of the rainfall-runoff function. Consequently, this work tests polynomial nets for their potential applicability in the flood forecasting context. In the following, "characteristic feature" describes an input vector to a net, containing important information about the rainfall-runoff process. A more detailed description of characteristic features is given in section 4. A general description of polynomial neural networks follows:

3.2.1 Fundamentals of polynomial neural networks

The overall polynomial approximation of the predicted discharge Q at the considered catchment outlet is described in this section. The polynomial network is a feed-forward network with a single hidden layer. The output of the "hidden" layer is the product of the input terms while the output of the network is the weighted sum of these products (figure 21). Polynomial nets have only one layer of adaptive weights, this results in a very effective training with short CPU times compared to the classical multi layer nets.

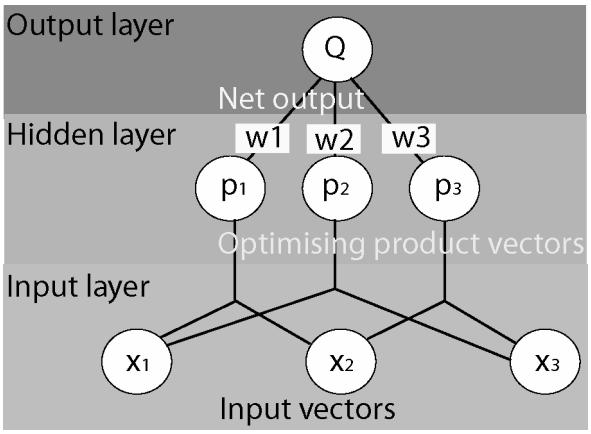


figure 21: Scheme of a polynomial net with [x1…x3] = input vectors, [p1….p3] = hidden layer, [w1….w3] = hidden layer weights, Q = output

The output of the network is given by:

$$Q = \sum_{i=1}^{N} w_i p_i$$

(equation 16)

with: Q = PoNN output (discharge)

w_i = linear weights of the PoNN (determined in the training)

p_i = product vectors

N = number of product vectors (3 in figure 21)

The PoNN product vectors result from a permutation of the inputs x (figure 21). They are exemplarily derived for a third degree PoNN according to:

$$p = x_k^a x_l^b x_m^c$$

(equation 17)

with: x = input for k = [1…N], l = [1…N], m = [1…N]. The exponents *a, b, c,* are all elements of [0; 3].

The power *a, b, c* of equation 17 satisfies the criterion :

$$a + b + c \leq g$$

(equation 18)

with:g = degree of the polynomial applied.

Usually the use of second or third degree polynomials is sufficient to guarantee a satisfactory forecast ability of the PoNN. In this case $a+b+c \leq 3$ sets the frame for the definition of product vectors, i.e. three features are simply multiplied or one feature² is multiplied with another feature, or just one feature³ is used. The PoNN transforms the input vectors directly into the flood forecast for a pre-defined river gauge at the catchment outlet. However, the total number of product vectors rises rapidly when high degree polynomials or a large number of input vectors are used. This is easily understood from the relation expressed in equation 19. Here N rises rapidly with increasing g or n.

$$N = \binom{n+g}{g}$$

(equation 19)

3.2.2 Training of polynomial nets

Stepwise serial regression (SSR) is a new algorithm which has been especially developed for the training of polynomial nets in the rainfall runoff context (Görner *et al* 2006). It consists of a combination of regression methods using both Efroymson`s algorithm as described in Miller (1999) and stepwise regression presented in Meyer-Brötz and Schürmann (1970). In the training process of the PoNN, first all possible product vectors *p* are calculated by means of permutation of the all characteristic features (equation 17). These vectors are then stored in the vector matrix *P*. The training process of the PoNN essentially consists of optimising the net structure and weights. This is achieved by the transformation of *P*, into a new - much smaller - optimised matrix *O*, i.e. *O* represents the structure of the trained net. The corresponding linear weights w_i of the optimised vectors p of *O* are determined by SSR. Herein, the general estimator for the target value $\Delta(x)$ is defined as follows:

$$\Delta(x) = w_0 + \sum_{i=1}^{n} w_i x_i \qquad \textbf{(equation 20)}$$

with: w = weight and x = regressor

Minimizing the objective function in the selection process (the deviation of target value and the dependant variable (Δx)) can be expressed as follows:

$$E\{[z - \Delta(x)]^2\} \Rightarrow Min \qquad \textbf{(equation 21)}$$

The training process itself starts with the selection of the first *n* of *N* vectors of the vector matrix *P*. These vectors represent the initial matrix O, which is yet to be optimised. Now the iterative optimisation (training) is ready for start: The stepwise regression arranges the n vectors in O with respect to reducing of the total error (MSE of observed and modelled flow). I.e. the first vector is the most important for the description of Q, the second vector is the second most important and so on and so forth. This process is also known as maximum scatter minimisation. The criterion for the scatter minimisation (ΔR) is calculated from the n^{th} Element in O according to:

$$\Delta R = \frac{E(x_i z_i)}{E(diagonal \in O)} \qquad \textbf{(equation 22)}$$

with: $E(x_i z_i)$ = element xz

$E(diagonal \in O)$ = diagonal element in O

Based on this ranking, the worst 30 % of the vectors are rejected and not considered further. Now vectors from *P* are drawn to fill the places of the rejected vectors in *O*. Again, the n vectors are ranked and reduced by rejection of the least 30%. This procedure is repeated until all *p* in *P* are pruned to the *n* optimal vectors of *O*. Thus, the training consecutively runs through the total amount of all the product vectors *p*. The key advantage of this training method is the information content of the features, which, contrary to the classical MLFN, can be interpreted in a physically meaningful way. The optimal size of *O* is a function of the catchment characteristics and described in section 3.4.

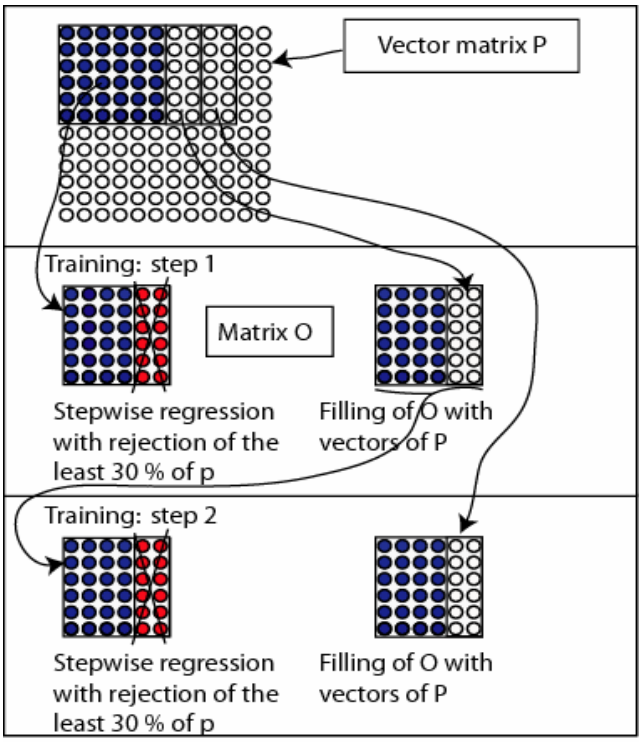


figure 22: The first two training steps of serial regression polynomial net training

3.3 Problem specific comparative analysis of multi layer net and polynomial network structures

To determine the optimal net structure in the context of flood forecasting, preliminary testing involves the comparison the two methods considered.

- Multi layer feed forward nets are a widely applied type. Their adoption to flood forecasting has been promoted by Schmitz *et al* (2005) and Cullmann *et al* (2006)

- Polynomial neural nets have been proposed for river discharge modelling by Foka (1999)

The two different net types are compared at Kriebstein gauging station. This gauge is the outlet of a 1757 km² flash flood prone catchment, situated in the East German ore mountain range, a more detailed description of the area is presented in chapter 5.1. The tested three layer net is generated according to the principles laid out in Schmitz et al (2005). It is trained with the strategy presented in section 3.1.3. The three layer net is fed with 30 input vectors containing information about rainfall, temperature and pre-event state of the watershed. The polynomial network is used with the same input vectors, it is generated and trained according to the standard procedures described in section 3.2. The testing is carried out with a lead-time of 24 hours for both net types. The input vectors are derived from data consisting of 1 km² grids of precipitation, air humidity, wind speed, global radiation and temperature. 47 years of data are available for hourly time steps. These data are transformed into runoff by a calibrated WaSiM model. The ability to portray this runoff is the evaluation criterion used in this study. All the 47 years of input data have been used to discriminate between the methods concerning training speed, operational speed, stability, forecast reliability, and easiness of model set up. The approaches are trained on 80 % of the available data, whilst 20% of the database is reserved for validation purposes.

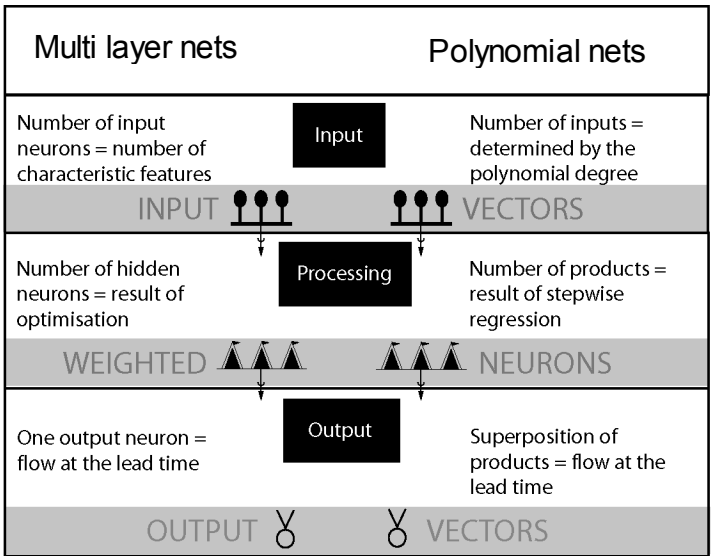


figure 23: Comparison of architectures multi layer net – polynomial net

A scheme shows the principles of the two approaches (figure 23). The architecture is generally different, but some parallels exist and ease direct comparison in the tests. The number of input neurons of the multi layer net equals the number of characteristic features used to portray the rainfall runoff process for the considered watershed. For the polynomial net, the input information consists of the same characteristic features, permutated according to the polynomial degree. The processing of the input information is solved in the hidden layer of the three layer net. This layer consists of a number of hidden neurons. Extensive testing is necessary to determine the optimal number of the hidden layers. The number of hidden neurons of an multi layer net can be compared to the number of products used for the forecast in the polynomial net. This value is found by optimisation procedures. The output of both polynomial and multi layer nets is one-dimensional; it consists of the flow at the considered lead-time.

As a first result of the tests, figure 24 shows that the Nash-Sutcliffe efficiency of the validation period varies significantly depending on the number of hidden layer nodes in the multi layer net. The result indicates that the predictive power of the MLFN is unstable for varying net architectures. This is not only the case for the Nash-Sutcliffe efficiency, but also proves for the mean square error in [%] of the total flow during the validation period, shown in figure 25. The behaviour of the forecast performance criterion is unforeseeable for increasing

numbers of hidden layer nodes. As the performance depends strongly on the input vectors and weights, this behaviour is characteristic for each watershed. This attribute poses a serious hurdle in the way of easy general application of a MLFN based flood forecasting system because for each watershed this preliminary and tedious testing would have to be repeated.

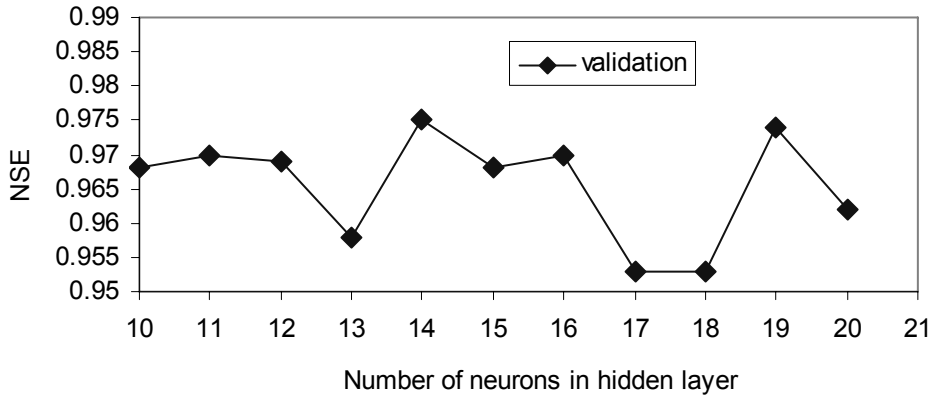


figure 24: Nash-Sutcliffe efficiency for the test data set versus number of neurons in hidden layer for a 24 h lead-time prediction with the three layer net

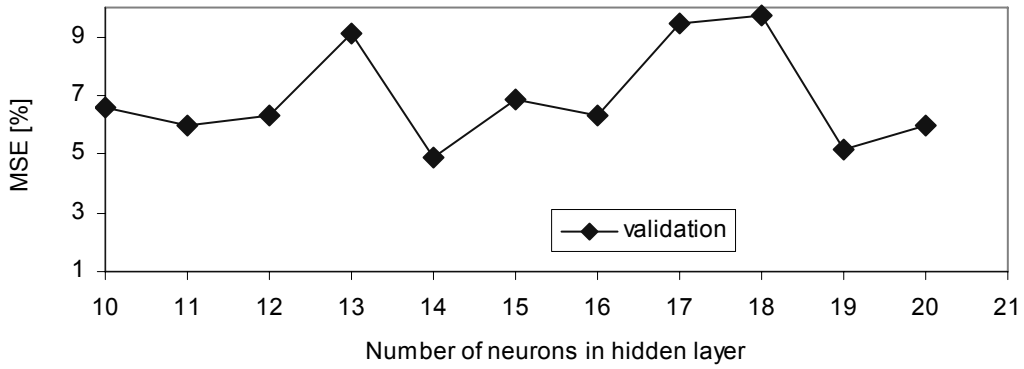


figure 25: Mean square error for the test data set versus number of neurons in hidden layer for a 24 h lead-time prediction with the three layer net

As explained above (figure 23), the number of product vectors used in the polynomial net can be compared to the number of hidden layer neurons of a multi layer network. This number impacts on the predictive power of the polynomial net. However, the generalisation performance of a polynomial net is

a clearly defined function of the number of product vectors employed. This allows for easily defining the net structure in the general set up of a polynomial net for different catchments. For the polynomial net, the validation graph in figure 26 illustrates how the predictive performance rises with increasing number of product vectors. It reaches an optimum, which is a plateau in the test case (vector numbers ranging from 120 to about 450). Generally the optimal configuration is the one which fully describes the dynamics with the least number of product vectors. The optimum is a function of effort (increasing numbers of vectors lead to more computational effort) and predictive reliability. The optimal set up may be readapted for each new forecasting task according to the users requirements and resources. Generally, increasing the number of deployed product vectors improves the nets training results. Obviously, the ability to generalize decreases if over-fitting characterises the system. This is visualized in figure 26, where product vector numbers > 550 show significant loss in the Nash-Sutcliffe efficiency (NSE) for the validation period (predictive ability). Again, this is confirmed by a similar behaviour of the mean square error MSE (figure 27).

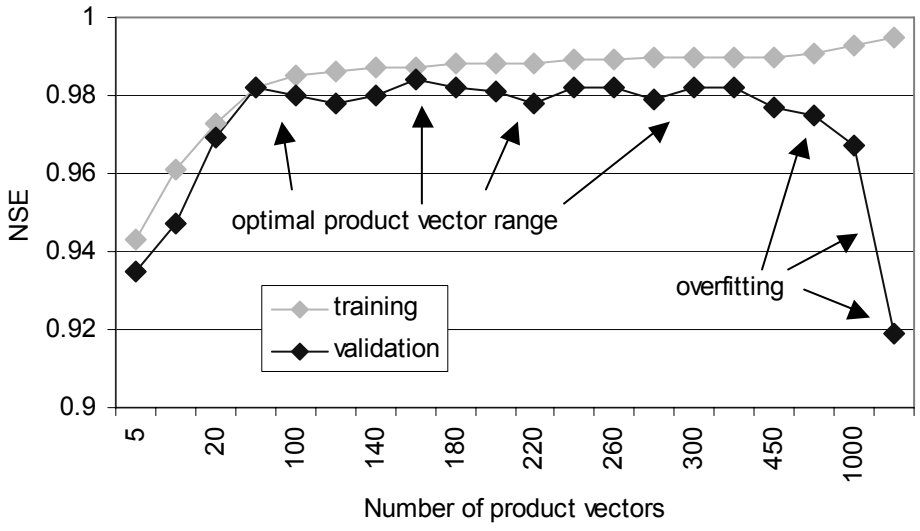


figure 26: Nash-Sutcliffe efficiency for training and test data versus number of product vectors for a 24 h lead-time prediction with the polynomial net

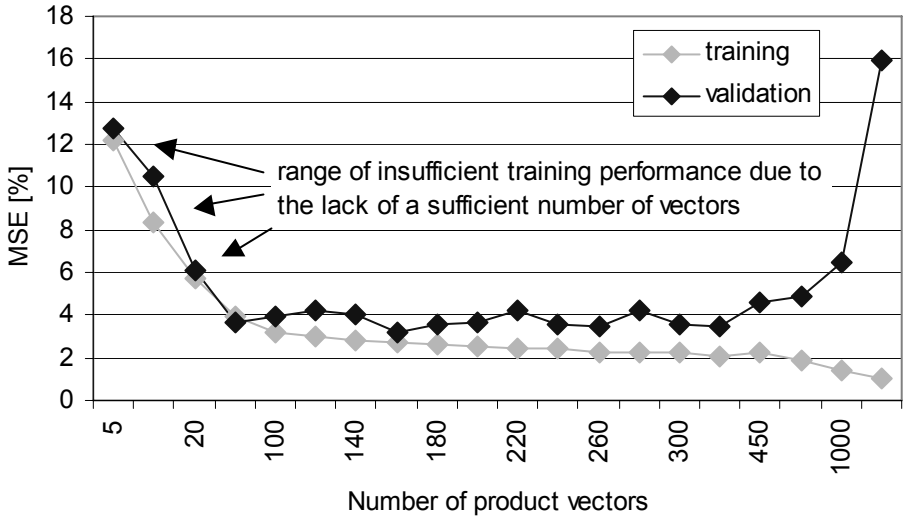


figure 27: Mean square error for training and test data versus number of product vectors for a 24 h lead-time prediction with the polynomial net.

In a second test, optimal 3 layer net architectures are compared to a 3rd grade PoNN encompassing 220 product vectors. The number of product vectors guarantees both good training results and a stable predictive performance (figure 27). The architecture of the multi layer nets is separately optimised for each lead-time. This results in the varying number of hidden layer neurons listed in column 2 of table 12. The optimal number of hidden neurons of the multi layer net varies between 14 and 16 (figure 24). For the optimal multi layer architectures, the Nash-Sutcliffe efficiency decreases with increasing forecast lead-time, i.e. the predictive power of the multi layer net decreases when longer lead times are considered (table 12). Contrary to this finding, polynomial nets are characterized by a stable forecast performance (Nash-Sutcliffe efficiency) with increasing lead-times. Obviously, this is a vitally important aspect and suggests preferring polynomial nets to multi layer nets in the flood forecasting context. This second test also reveals that the training time is constant for polynomial nets. In contrast, multi layer nets exhibit very inhomogeneous training times (table 12). The mean multi layer training time is double the time needed to train polynomial nets. The forecast CPU requirements, which is most important for the operational applicability is the same for the two methods.

table 12: Training performance of multi layer and polynomial nets

Lead-time	Multi layer net			Polynomial net		
	Nr. hidden layer nodes	Training [min]	NSE	Grade/ Nr. of features	Training [min]	NSE
6	14	111	0,97	3/220	45	0,97
12	14	26	0,93	3/220	45	0,98
18	16	111	0,86	3/220	45	0,98
24	14	237	0,85	3/220	45	0,97
36	15	28	0,86	3/220	45	0,97
48	16	34	0,84	3/220	45	0,97
Mean	15	90	0,88	3/220	45	0,97

The general results of the testing are: The predictive power of polynomial nets is superior to the one of classical multi layer feed forward neural networks. Polynomial nets are easy to set up because of the clear relationship between structure and performance (figure 27). Their fast and stable training, together with moderate operational CPU requirements (table 12) allow for a general and easy application of polynomial net based forecast tools. The comparative testing identifies polynomial nets as the method of choice for the set up of online flood forecasting applications. The convincing ground for this statement is the stable performance criterion of the polynomial net based system for increasing lead-times. This characteristic is of paramount importance for a reliable forecast and will become more and more important in the future as rainfall forecasts will become more and more accurate.

3.4 Optimal polynomial network forecast strategy for flash floods

In this section three tests serve as a means for defining the optimal net structure in terms of the power of the polynomial, the best forecast strategy and the training data requirements. All the tests are based on the 585 km² catchment of the Freiberger Mulde at Nossen gauging station. In the catchment, the agriculturally used area dominates, up to 30 % of the catchment are covered with forests. The travel time within the catchment is approximately 18 hours. The tests have been conducted at Nossen gauge to ensure a good predictive performance even for smaller catchments and show the general nature of the PAI-OFF approach proposed in the next chapter. The meteorological and runoff database is analogue to the one used in the preceding chapter.

The first test evaluates the dimensionality of the polynomial approach. The net is used in a second as well as a third degree approach, i.e. it is tested for $a+b+c = 2$ and $a+b+c = 3$ for equation 18. The result of this test is depicted in figure 28. The blue dots mark the performance of the second degree net. The green third degree approach (figure 28) clearly outperforms the second degree approach for all lead times. Based on this finding, polynomial nets are generally used in the third degree in the further testing and application.

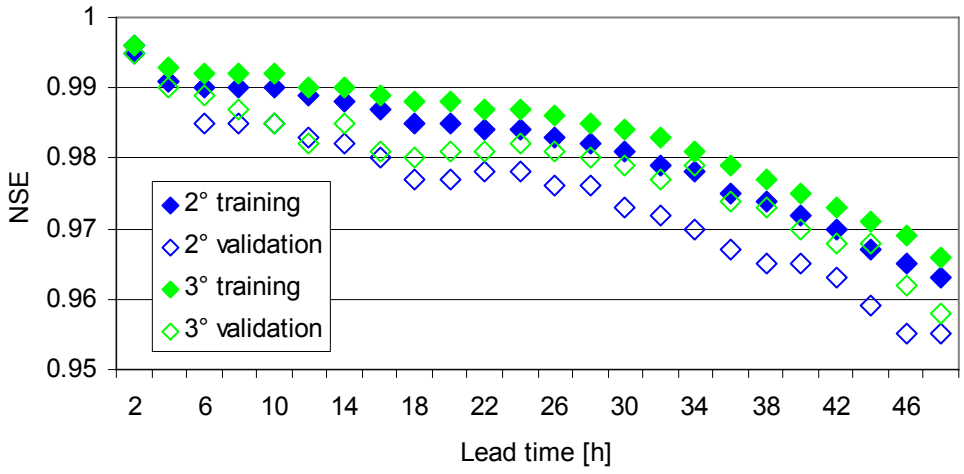


figure 28: Testing for the optimal polynomial degree (Nash-Sutcliffe efficiency)

It is principally possible to forecast either the absolute flow at the desired lead time or the flow increment that separates the actual discharge from the flow at the lead time (figure 29). The forecasting approach developed in this section adheres to the strategy of incremental predictions for lead-times less than 12 hours. The calculation of the forecast at time t for time $t + \Delta t$ (lead-time increment) considers the flow increments ΔQ with respect to the actual discharge at time t. For lead-times of more than 12 hours, the prediction does not principally depend on differences to the actual flow any more. Therefore it is favourable to directly predicts the flow Q for lead-times > 12h. This principle is shown in figure 29. In this figure, the green flow refers to the delta prognosis for up to 12 hours while the blue flow is the absolute value prediction used for longer lead-times.

The strategy described above is derived from a sensitivity analysis of the Nash-Sutcliffe criteria of both Q and ΔQ prediction for lead-times of up to 48 hours. The result of this analysis is shown in figure 30, where the green dots represent the results for ΔQ, while blue denotes results for Q.

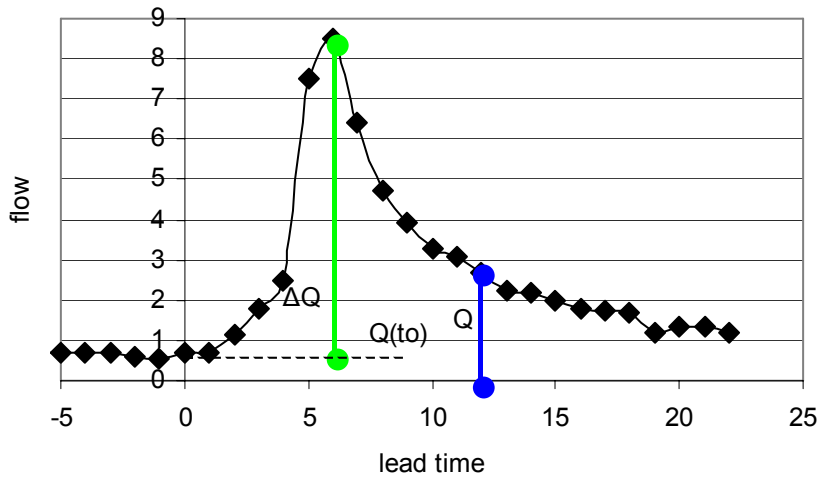


figure 29: Prediction strategy for Q and ΔQ respectively.

It can easily be noticed that for longer lead-times the blue squares, denoting the validation performance for the direct flow prediction, stay in the range of 0.98 while the green squares of the ΔQ validation performance criterion steadily decrease – down to values below 0.97. This difference is small, but nevertheless

significant as the Nash-Sutcliffe efficiency is calculated over a period of 47 years. For smaller lead times, the green validation squares of the delta method yield better results than the direct prediction of Q.

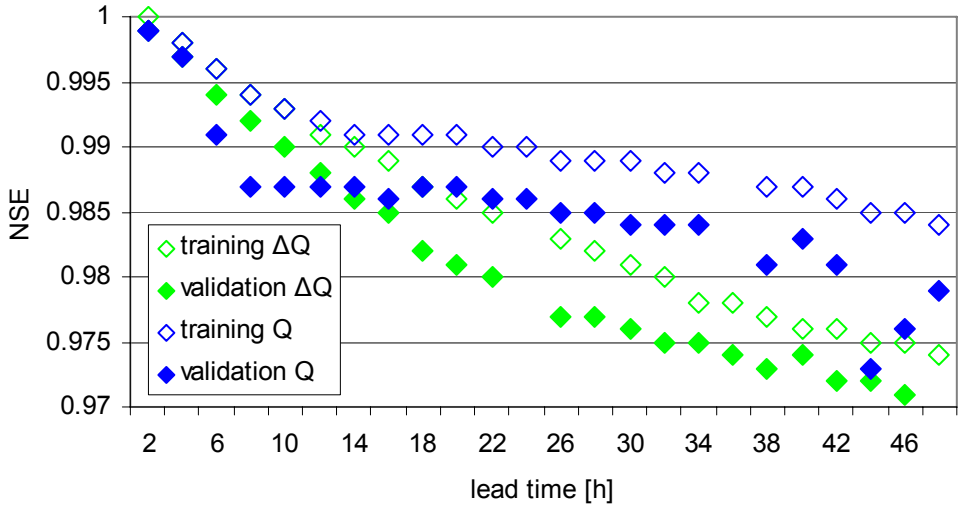


figure 30: Prediction strategy: Q versus ΔQ (Nash-Stutcliffe-Efficiency)

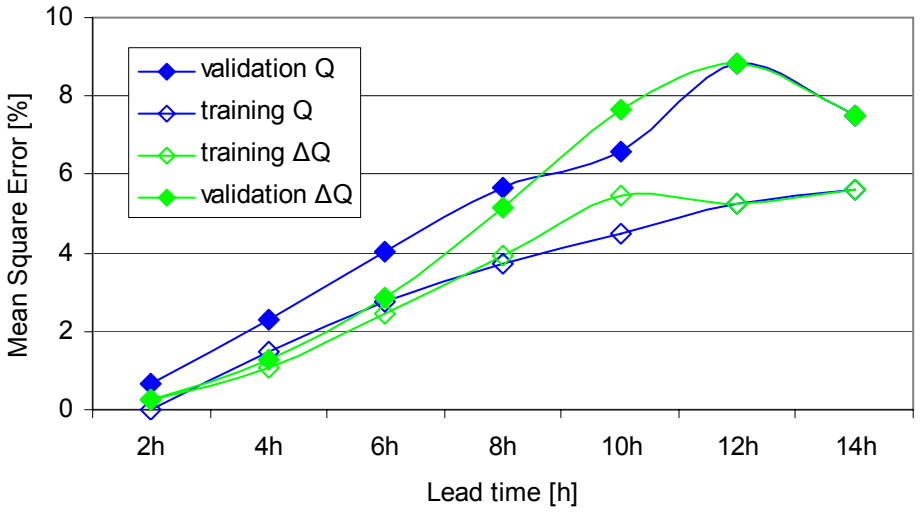


figure 31: Prediction strategy: Q versus ΔQ (Mean square error)

In figure 31, a zoom on the first 7 lead time increments of the test confirms the statements made earlier. For short lead times < 12 h the mean square error of the ΔQ prediction strategy clearly outperforms the direct prediction of Q.

A further important feature in the set-up of the polynomial net is the size of the training database. In a third test, this size has been varied from 10 to 47 years. 20 % of the training database is always reserved for testing the predictive power. One year of test data comprises a rough average of 20 flood events. The Nash-Sutcliffe efficiency yields excellent values for training databases with 20 years of input data. But if the mean square error is additionally considered, it becomes clearly visible, that the predictive power best if more than 40 years of data (600 events) are used in the training of the polynomial network.

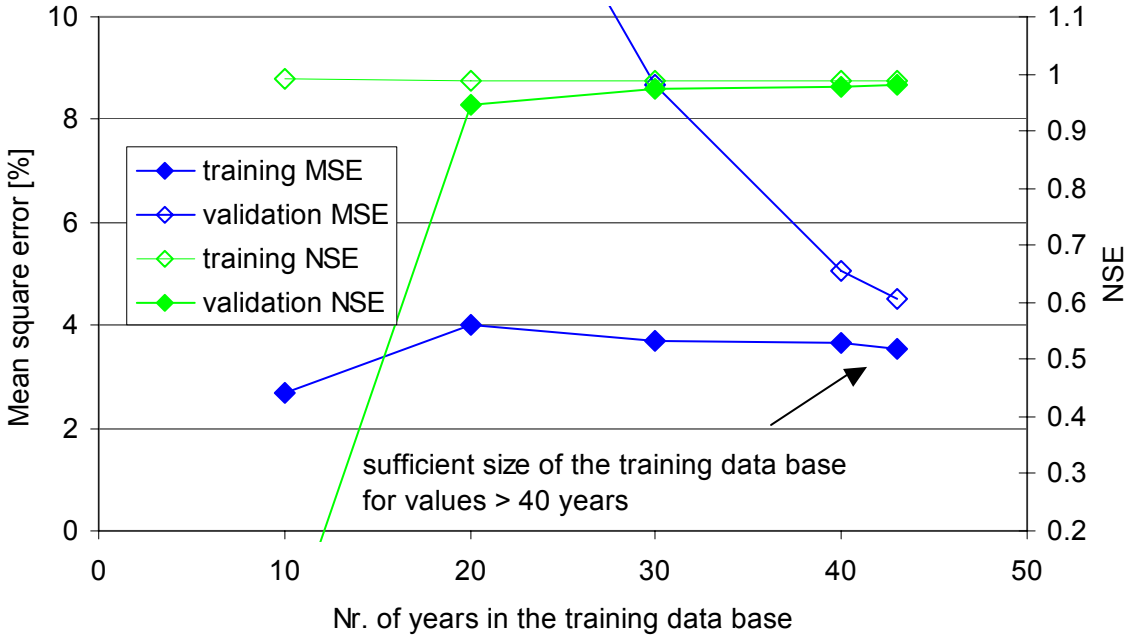


figure 32: Effect of the size of the training database on training and validation criteria (Nash-Sutcliffe efficiency and mean square error)

4 The PAI-OFF flash flood forecasting methodology

PAI-OFF (Process modelling and artificial intelligence for online flash flood forecasting) is a new methodology for improved online flash flood forecasting. The approach is founded on the adequate incorporation of the catchment specific flood relevant processes into a neural network. It covers both hydrological history - describing the catchment pre-event conditions - and specific rainstorm characteristics. These are the crucial foundations of any successful online flood forecasting. Various parameter sets have to be taken into account for accurately forecasting all the naturally occurring patterns of flood formation (section 2.5). From the many flood forecasting models analysed in section 1.2 we have learnt that highly accurate distributed process models are - amongst other drawbacks - too slow for improving the forecast in terms of lead time and/or a closer examination of the forecast uncertainty. Artificial neural network based approaches are fast enough, but if they are exclusively trained with historical data, they lack predictive power and fail to accurately forecast unseen extreme events.

Offering a way out of the dilemma, PAI-OFF combines the two approaches, transfusing the predictive power of physically based models into a fast and robust polynomial network based flash flood forecasting tool. The polynomial approach is chosen on the basis of the findings of sections 3.4 and 3.3. The combination of physically based process modelling and the polynomial neural network is achieved by training the net on the basis of a comprehensive database, containing all possible patterns of natural flood formation and all important pre-event conditions. This database is built by means of the physically based model. Therefore, all flood prone meteorological constellations are collected. They form the input side of the database. These data are then transformed into the corresponding runoff by means of a dynamically parameterised, physically based catchment model. This runoff data represents the output side of the polynomial networks training database (consider section 4.5 for further details about the database). Consequently, the trained polynomial net comprises the physically based models fidelity to the process for all patterns of flood formation and all the possible pre-event conditions. In the operational

phase, the trained PoNN substitutes the catchment model with implicit adequate parameter sets for all possible constellations of flood formation. PAI-OFF is easy to use and extremely fast. It therefore allows for analysing the forecast uncertainty by online Monte-Carlo simulations. This becomes possible because PAI-OFF is about 500 times faster than a classical hydrological model.

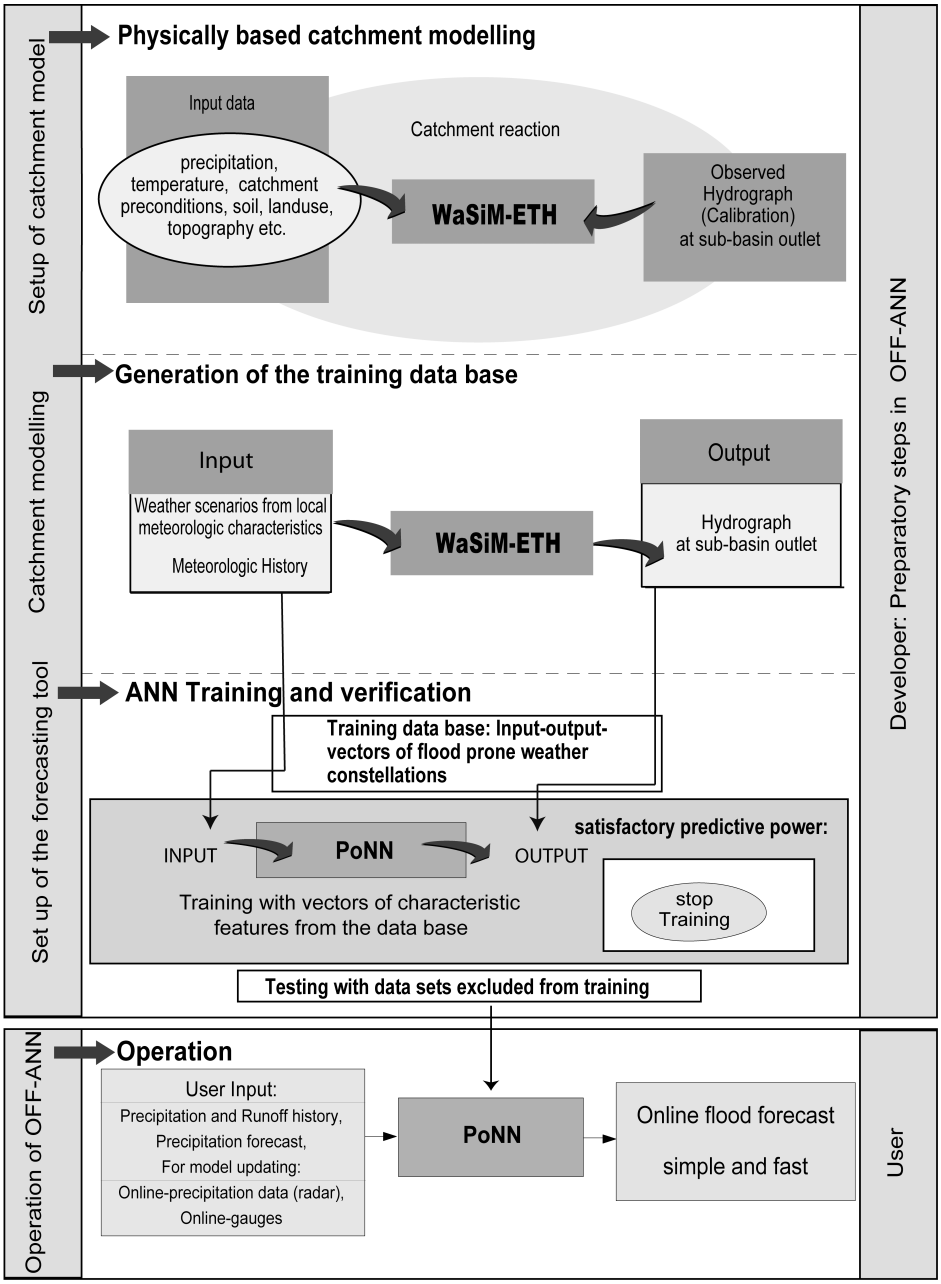


figure 33: Flow chart of the PAI-OFF methodology

Summarizing the three step preparatory phase:

1. The first step of the PAI-OFF methodology consists in setting up a physically based catchment model of the considered watershed. For this purpose, any sound and reliable hydrological model may principally be used;

2. After successful validation for a specific catchment, the physically based model is used for simulating the whole range of physically possible, flood relevant meteorological/hydrological situations which are meaningful for flood formation and propagation in the area (figure 33). The input (rainfall) and obtained output (flood discharge) values – together with the corresponding pre-event (initial) hydrologic catchment conditions – reflect the hydrologic behaviour of the catchment;

3. In the third step the polynomial neural network is trained to portray the complete hydrological process information considering the rainfall-runoff transformation. Here the incorporation of watershed pre-event conditions plays a key role for correctly portraying all the patterns of flood formation contained in the database.

The PAI-OFF flood forecast is a single step process once the polynomial net is trained (figure 33, operation). Generally, as all other neural network based solutions, the polynomial net used in PAI-OFF requires a number of input vectors (characteristic features). These selected features provide the net with the necessary information for the flood forecast (figure 33, user input). The purpose of feature selection is choosing information that fully portrays the dynamics of rainfall-runoff processes in the watershed. Generally, for any kind of hydrological model, it is desirable to use as much information as possible. Polynomial neural nets are also obeying this rule. For operational online flood forecasting though, input data has to be integrated and transformed into characteristic features for several reasons: One motivation is the sheer dimensionality of distributed meteorological input data. Handling high-resolution (e.g. 1 km²) grid based meteorological data would result in an enormous effort organizing the database and feeding the input vectors to the net This would lead to computing times foiling the general advantage of PAI-OFF: the fast forecast. Instead, relevant features are carefully selected, integrating the

data whilst keeping the information content. With this strategy, the number of net inputs can be kept at a minimum. This directly leads to the second reason for integrating input data: The number of input vectors is inversely proportional to the predictive power of any neural network based solution, i.e. more input = worse validation performance of the net. The strategy of integrating data in meaningful characteristic features helps improving the operational performance of the system. A third reason for this strategy is reducing redundant data in the training process. The training process is significantly slowed if redundant data has to be trained. The characteristic features cover the two most important aspects considering rainfall runoff dynamics: pre-event state of the catchment as well as precipitation as the dominant driving force. This information is usually derived from observations made before the oncoming event. The most important features contain information about predicted precipitation in terms of intensity, duration and location. In PAI-OFF, the characteristic features are divided in two classes (figure 34): The state features are used for describing the pre-event state of the catchment and the hydrologic response features characterise the runoff formation process as a result of precipitation input.

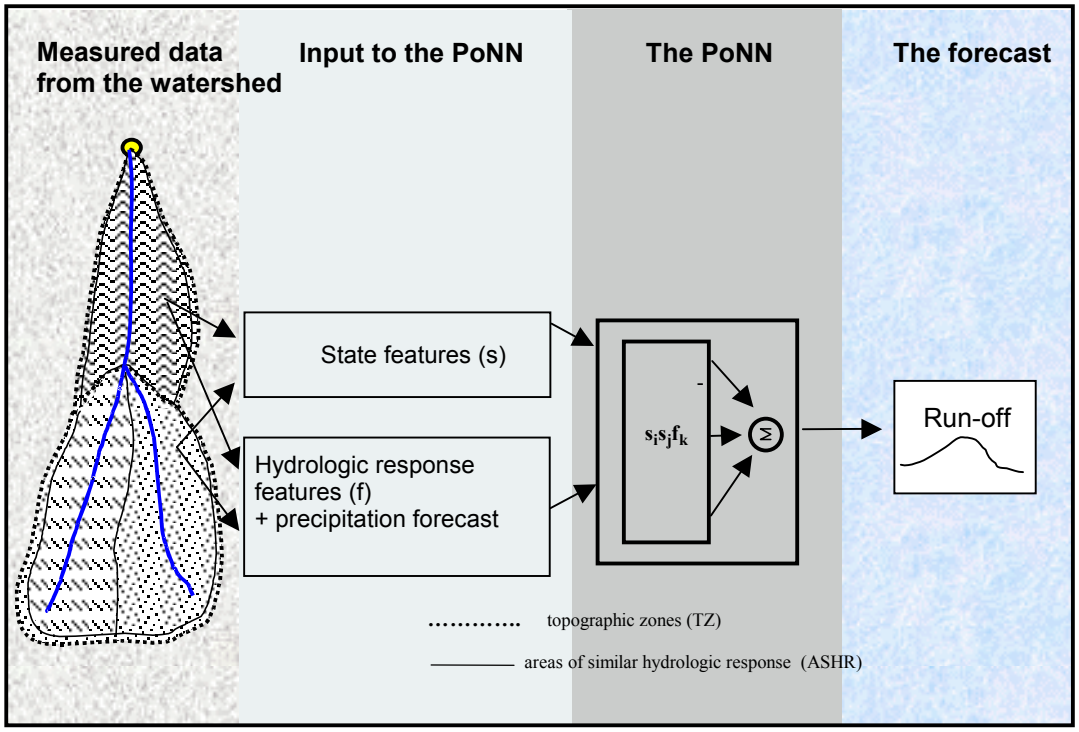


figure 34: Basic working principles of the polynomial net rainfall-runoff module

These two types of characteristic features are jointly processed in the course of the forecast by permutation (equation 17).

It is important that these vectors must exclusively consist of data which is commonly available (e.g. at water management agencies or flood alarm centers) in order to guarantee the overall applicability of the PAI-OFF methodology. In the following sections the hydrologic response features as well as state features are described in detail.

4.1 Hydrologic Response Features (HRF)

Hydrologic response features (figure 34) allow for incorporating hydrologic process knowledge into the net by explicitly taking into account the mechanisms of the catchment specific potential flow paths. So far, neural network based approaches have always been used as lumped models in rainfall runoff modelling. By means of the hydrologic response features, it is now possible to train the polynomial nets not only on the basis of observed/simulated input/output relationships (lumped model) but also with some pre-processed hydrologic evidence. Operating the nets with this hydrologic "expert knowledge" –which is included for the specific catchment during the set up of PAI-OFF - allows for the portrayal of high resolution input data and results in a substantial overall improvement of the predictive performance. The hydrologic response features lead to a better portrayal of rainfall patterns and the predicted movement of rainstorm fields over the catchment. They are offered with a given positive or negative time lag and to the power 2 and 3 in the training process. This is relevant, because the hydrologic response features can thus account for non-linear response characteristics of the area.

Determining hydrologic response features requires subdividing the catchment into areas of similar response time with respect to the basin outlet – the so-called area of similar hydrologic responses (AHR). In the operational PAI-OFF system each area of similar hydrologic response bears one hydrologic response feature. For each of the zones, a typical convolution kernel is evaluated on the basis of simulations of characteristic storm scenarios with the catchment model. Thus, the potential runoff contribution of these zones to the total flow at the watershed

outlet is already pre-described by the convolution kernel assigned to each area of similar hydrologic response.

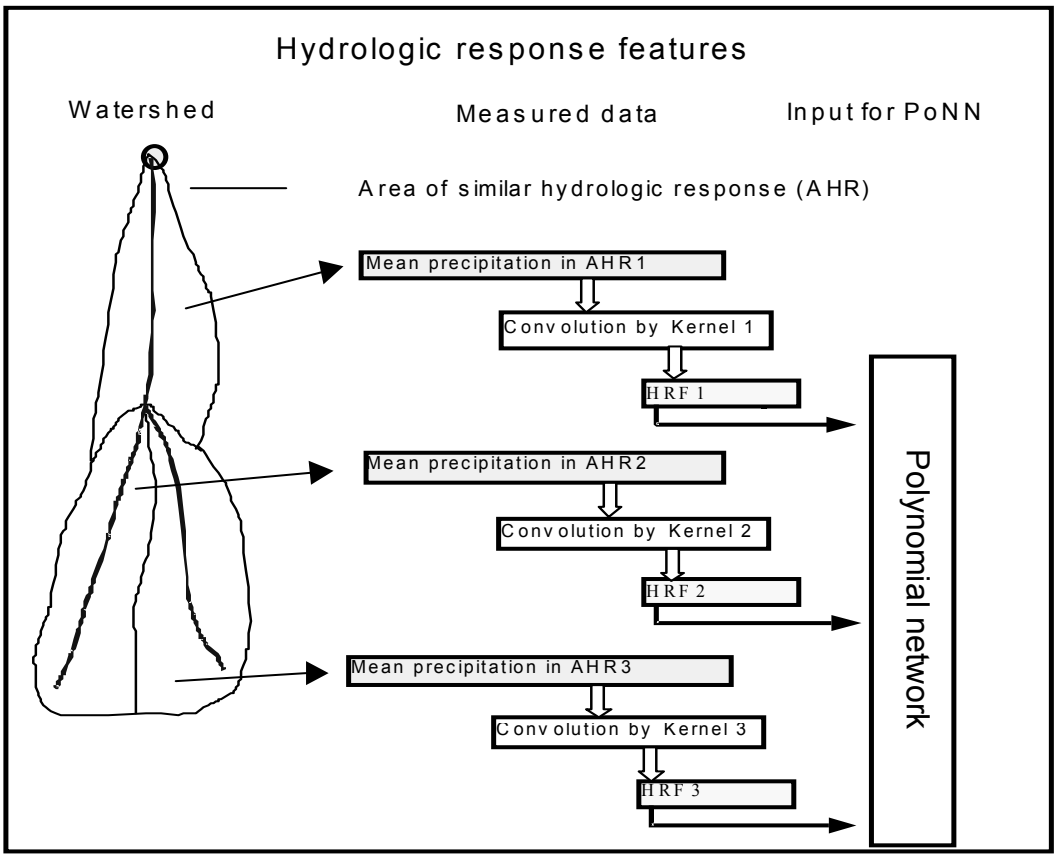


figure 35: Incorporation of hydrologic response features into the PAI-OFF net

The determination of the appropriate HRF for a given catchment is realized during the catchment modelling preparatory phase (figure 33):

The delineation of the area of similar hydrologic responses starts with the time-area (isochrone principle) method. It is generally used to derive a discharge hydrograph from a given excess rainfall hyetograph.

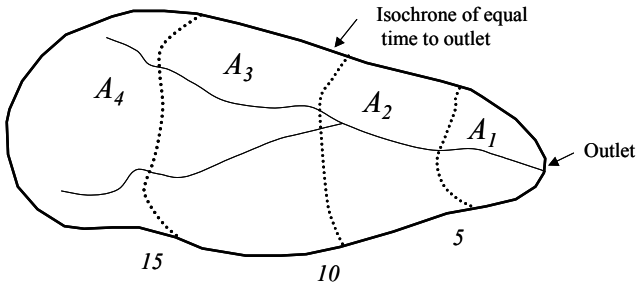


figure 36: Schematic subdivision of a watershed by means of the isochrone method

In figure 36, a plan view sketches an exemplary watershed with 4 travel time zones of 5 hours each. The procedure of travel time determination is well documented in Ponce (1989) and McCuen (1998). Further details are explained in Maidment (1993). He gives an example of the time-area model using grid based digital elevation models (DEM).

So far, the catchment is divided in zones of equal response times. Zones A3 and A4 in figure 36 represent the catchment areas of both rivers. This is contradictory to the principles of similar hydrologic response, because the reaction of areas attributed to different rivers might be totally different for a natural catchment. The zones of similar travel times are therefore further subdivided according to the river network.

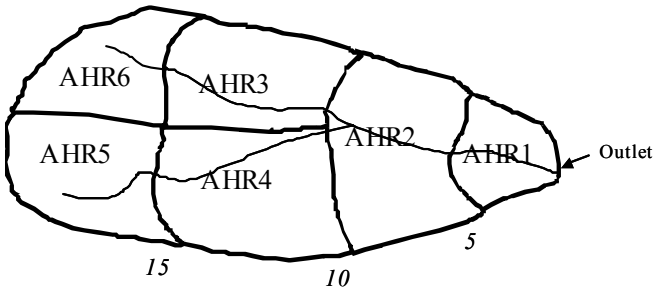


figure 37: Exemplary areas of similar hydrologic response

This leads directly to the areas of similar hydrologic response shown in figure 37. Once the areas of similar hydrologic response are defined, 300 rainstorm events from the training database are selected according to the following criterion: The events must have a minimum discharge with a return period of 20 to 25 years. These events form the kernel database, which is used to stimulate all hydrologic response areas separately. The calibrated catchment model is used to

transform the rainfall stimuli into runoff. A standard Duhamel-integration followed by minimum square deviation regression (Dyck and Peschke, 1995) is performed on this data and leads directly to the deterministic response kernel (figure 38). This resulting kernel is likely to be too rough for direct use as input for the PoNN training. Therefore a Gaussian filter is used to transform this deterministic kernel into a smooth analytical function according to equation 20:

$$S_t = \sum_{i=-n}^{n} W_i X_{t+i} \qquad \textbf{(equation 23)}$$

withS_t = smoothed function

X_t = original time series

W_i, i= weights

The Gaussian filter characteristics are that the weights equal the ordinates of an appropriate Gaussian probability density function (Mitchell *et al* 1966). The weights for a Gaussian window are calculated with:

$$W\ (k\ +\ 1)\ =\ e^{-\frac{1}{2}\left(\alpha\left(\frac{K\ -\ \frac{N}{2}}{\frac{N}{2}}\right)\right)^2} \qquad \textbf{(equation 24)}$$

with: $0 \le k \le N$ and $\alpha \ge 2$, where α is the reciprocal of the standard deviation and N is the filter length.

Generally, the length of the filter is inversely related to the value of α, a larger value of α produces a narrow filter length. For the purpose of filtering the deterministic response kernels, the following characteristics are used: filter length (N) = 5 and standard deviation = 0.65. The figure 38, depicts a hydrologic response kernel typically used in PAI-OFF . The green function is the exemplary result of a regression with more than 150 events. The kernel is rough, especially for flow times > 30h. The analytical kernel (red function in figure 38) allows for a smooth training of the PoNN.

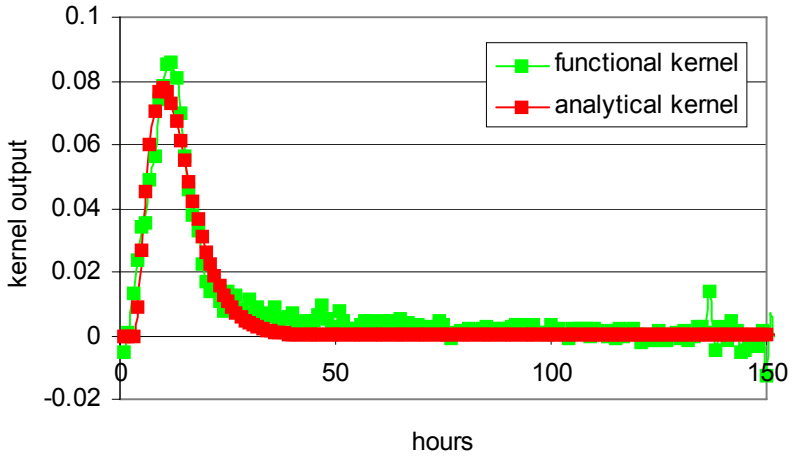


figure 38: hydrologic response feature kernel (functional-analytical)

For all areas of similar hydrologic responses, one unit hydrograph type kernel is used to determine the hydrologic response feature (figure 35), characterizing the catchments response to precipitation. Each hydrologic response feature is calculated according to:

$$F(t_0) = \sum_{\ell=0}^{L} P(t_0 - \ell)\, K(\ell)$$

(equation 25)

with: P = Precipitation at time t and ℓ = actual lag (time step)

L = maximum lag defining the temporal extension of the considered feature

The parameter ℓ (actual time lag) equals the number of considered time steps within the total length L of the feature, it is 2 for a time 2 hours before the start time of the forecast. L is a parameter that controls how much of the preceding history of an event is taken into account. It is defined in the set-up phase of PAI-OFF and then remains constant for any operational purpose. The minimal length L is defined as the maximal process memory. If this is unknown, it can be found during the training process by means of the stepwise serial regression.

$K(\ell)$ yields:

$$K(\ell) = c\left(\frac{\ell - L_0}{K_2}\right)^{(N-1)} e^{-\frac{(\ell - L_0)}{K_1}}$$

(equation 26)

with the parameters *c, K_1, K_2, N* derived by regression from the kernel database, L_0 represents the travel time from the flow time zone outlet to the reference gauge. All of the features (hydrological response features and state features) are denoted by $F(t_0)$ in the following. This reflects their equal treatment in the training of the net when they are permutated according to equation 4.

4.2 State Features (SF)

The formation of flood events is influenced by the hydrological and meteorological history of the catchment prior to a flood event, this is also known as pre-event state of the watershed. In PAI-OFF, the state features account for these catchment pre-conditions. They are selected according to basic hydrological principles. One of the most fundamental hydrological rules says: the discharge is the most reliable indicator for the basin internal state. This rule is integrated in PAI-OFF by means of various state features derived from the measured discharge.

Features derived from the observed discharge:

The discrete actual discharge is used as a feature:

$$F(t_0) = Q(t_0)$$

(equation 27)

with $Q(t_0)$ = actual flow at time t_0

Furthermore, the mean runoff in a defined time period is related to the basin internal state. The corresponding feature is the mean runoff in the interval $\{t_0, t_0 - L\}$, which writes as:

$$F(t_0) = \frac{1}{L}\sum_{\ell=0}^{L} Q(t_0 - \ell)$$

(equation 28)

with: ℓ = actual lag and L =.maximum lag, spanning $\{t_0, t_0 - L\}$

Here L can be used for introducing various spans for the features. It can be advantageous to employ various mean runoff features for the PAI-OFF net, e.g. one for a mean flow over the last three days, one for the mean flow over the last week and a third one for the mean monthly flow. This is allows for taking into account the specific internal hydrologic pathways of the considered catchment during the setting up of PAI-OFF.

As the observed discharge is most influenced by the conditions that prevailed just before the event, a weighted mean is introduced in PAI-OFF. It gives most importance to the discharges just before t_0:

$$F(t_0) = \frac{\sum_{\ell=0}^{L} Q(t_0 - \ell)\, e^{-\frac{\ell}{\tau}}}{\sum_{\ell=0}^{L} e^{-\frac{\ell}{\tau}}}$$

(**equation 29)**

with: τ = constant denoting the initial time step for the weighting in hours. This constant reflects the length of hydrologic memory of the catchment. It becomes thus possible to distinguish catchments with highly dynamic runoff processes from catchments with inert hydrological processes. If the user does not have access to such kind of information, the parameter can be optimized by means of stepwise serial regression during the set up of PAI-OFF.

Equation 29 results in a weighted mean. But, for flood forecasting it is essential to evaluate information about the gradient of the flow too. To this end, a special feature based on the Lagrange interpolation polynomial allows considering the gradient over the 5 preceding discrete values, starting from the actual flow. It is calculated as:

$$F(t_0) = \frac{1}{60}(-137\, Q(t_0) + 300\, Q(t_0 - 1) - 300\, Q(t_0 - 2) + 200\, Q(t_0 - 3) - 75\, Q(t_0 - 4) + 12\, Q(t_0 - 5))$$

(equation 30)

Extreme flows are an additional valuable indicator for evaluating the catchment internal state history within the interval $\{t_0, t_0 - L\}$. PAI-OFF therefore allows for incorporating this information by the state features for low and high flow:

$$F(t_0) = \min\left[Q(t_0) + [\overline{Q}(t_0 - \ell) - Q(t_0)]\, e^{-\frac{\ell}{\tau}} \right]_{\ell=0}^{L} \qquad \textbf{(equation 31)}$$

$$F(t_0) = \max\left[Q(t_0) + [\overline{Q}(t_0 - \ell) - Q(t_0)]\, e^{-\frac{\ell}{\tau}} \right]_{\ell=0}^{L} \qquad \textbf{(equation 32)}$$

Another important source of information describing the basin internal state is the precipitation history. Consequently, a significant number of state features are based on observed precipitation:

Features derived from observed precipitation:

The mean precipitation in the interval $\{t_0, t_0 - L\}$ yields information about the state of wetness of the considered catchment. It is calculated with:

$$F(t_0) = \frac{1}{L} \sum_{\ell=0}^{L} P(t_0 - \ell) \qquad \textbf{(equation 33)}$$

with: P = actual precipitation at time t_0

In analogy to the weighted discharge, a feature allows for describing precipitation as a function of time to underline the importance of the last time steps before the onset of the event:

$$F(t_0) = \frac{\sum_{\ell=0}^{L} P(t_0 - \ell)\, e^{-\frac{\ell}{\tau}}}{\sum_{\ell=0}^{L} e^{-\frac{\ell}{\tau}}} \qquad \textbf{(equation 34)}$$

The reaction of a watershed to precipitation input depends strongly on the temporal distribution of the precipitation intensity. I.e. the discharge from a watershed that receives 30 mm in 3 equal shares of 10 mm differs a lot from that of a watershed which is stimulated by a 5 mm, 20 mm, 5 mm rainfall distribution. This simple rule also applies for the basin state description. Consequently, the relation of peak and mean values in the interval $\{t_0, t_0 - L\}$ is evaluated in PAI-OFF:

$$F(t_0) = \frac{max\left[P(t_0 - \ell)\right]_{\ell=0}^{L}}{\frac{1}{L}\sum_{\ell=0}^{L} P(t_0 - \ell)} \qquad \textbf{(equation 35)}$$

So far, the precipitation based features describe the wetness state of the catchment by mean values and the relation of peak to mean values. The temporal distribution of the rainfall volume within the considered pre-event period yields additional information. It makes a great difference if the previous history of the considered event is characterized by more or less equally distributed precipitation or by periods of intense rainfall or dryness. The number of time steps with precipitation ≥ threshold Θ in $\{t_0, t_0 - L\}$ reveals information about the density of the preceding rainfall:

$$F(t_0) = \sum_{\ell=0}^{L} if\left(P(t_0 - \ell) > \Theta, 1, 0\right) \qquad \textbf{(equation 36)}$$

Long dry periods influence the basin internal state. Consequently the basin response to a rainstorm after a long dry period differs greatly from the reaction caused by an event hitting the catchment under normal conditions. The time steps without precipitation are weighted for the interval $\{t_0, t_0 - L\}$.

$$F(t_0) = \frac{1}{L}\, max\left[d_k \; e^{-\frac{\ell_k}{\tau}} \right]_k \qquad \textbf{(equation 37)}$$

with: d_k = length of period *k* without precipitation

Last but not least, the precipitation peak value of the preceding history might have a significant influence on the formation of a flood event. Thus a feature allows for taking into account the filtered and weighted peak rainfall in the interval $\{t_0, t_0 - L\}$.

$$F(t_0) = max\left[\overline{P}(t_0 - \ell) \; e^{-\frac{\ell}{\tau}} \right]_{\ell=0}^{L} \qquad \textbf{(equation 38)}$$

where: $\overline{P}(t) = \frac{1}{5}\sum_{\ell=-2}^{2} P(t + \ell)$ **(equation 39)**

Apart from discharge and precipitation, other factors influence the hydrological state of a catchment. Evapotranspiration drains the soils of catchments and might therefore be an important factor controlling the onset of a flood event. PAI-OFF consequently offers the possibility to directly include information on global radiation and air humidity for portraying this process.

Features based on global radiation:

Global radiation is the driving force of evapotranspiration. To this end is considered in two features. The mean feature and the weighted mean feature according to equations 40 and 41.

$$F(t_0) = \frac{1}{L} \sum_{\ell=0}^{L} G(t_0 - \ell) \qquad \textbf{(equation 40)}$$

$$F(t_0) = \frac{\sum_{\ell=0}^{L} G(t_0 - \ell)\, e^{-\frac{\ell}{\tau}}}{\sum_{\ell=0}^{L} e^{-\frac{\ell}{\tau}}} \qquad \textbf{(equation 41)}$$

where *G(t)* denotes the global radiation at time *t*.

Features based on air humidity:

In PAI-OFF air humidity is represented by means of two features in analogy to the global radiation:

$$F(t_0) = \frac{1}{L} \sum_{\ell=0}^{L} M(t_0 - \ell) \qquad \textbf{(equation 42)}$$

$$F(t_0) = \frac{\sum_{\ell=0}^{L} M(t_0 - \ell)\, e^{-\frac{\ell}{\tau}}}{\sum_{\ell=0}^{L} e^{-\frac{\ell}{\tau}}} \qquad \textbf{(equation 43)}$$

M(t) denotes the relative air humidity. This parameter may be replaced by the vapour pressure in case this data is more easily accessible.

Features based on the vegetation period:

The vegetation activity represents another important aspect for correctly portraying catchment pre-event conditions is. It influences the flood formation in two ways: It affects evapotranspiration and it changes the basin properties with respect to interception and overland flow resistance. Vegetation activity is defined as a function where 0 is no activity and 1 represents full vegetation activity. This function reflects specific characteristics of the considered catchment, not only with respect to the climatic conditions, but also regarding human impact on the catchment state, such as deforestation, mining or agricultural practices. An exemplary function is depicted in figure 39.

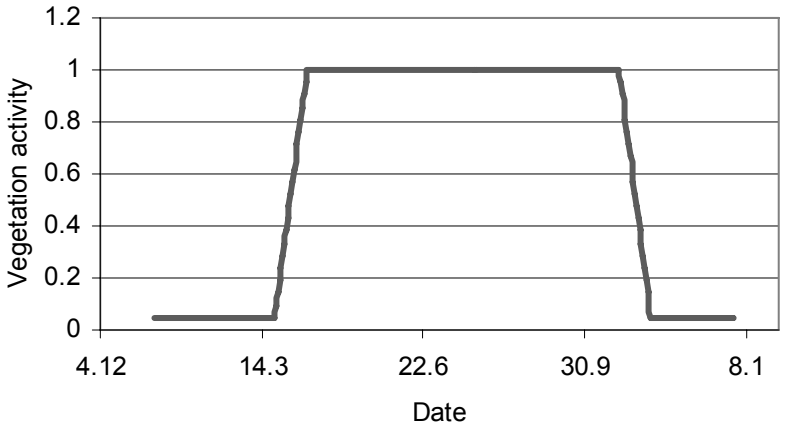


figure 39: Exemplary vegetation activity function for a typical East German mountainous watershed

PAI-OFF offers three features for characterizing the vegetation activity :

The discrete actual value, where *V(t)* = the vegetation index [0;1] at time *t*:

$$F(t_0) = V(t_0)$$ **(equation 44)**

The mean activity in the interval $\{t_0, t_0 - L\}$:

$$F(t_0) = \frac{1}{L}\sum_{\ell=0}^{L} V(t_0 - \ell)$$ **(equation 45)**

The weighted activity:

$$F(t_0) = \frac{\sum_{\ell=0}^{L} V(t_0 - \ell)\, e^{-\frac{\ell}{\tau}}}{\sum_{\ell=0}^{L} e^{-\frac{\ell}{\tau}}} \qquad \textbf{(equation 46)}$$

Features derived from measured temperature:

Without accounting for temperature, PAI-OFF would not be able to model the snowfall, snow storage and snowmelt process. The possibility to include this information is offered by means of two relevant features:

Mean temperature in the interval $\{t_0, t_0 - L\}$:

$$F(t_0) = \frac{1}{L} \sum_{\ell=0}^{L} T(t_0 - \ell) \qquad \textbf{(equation 47)}$$

Weighted mean with exponential kernel:

$$F(t_0) = \frac{\sum_{\ell=0}^{L} T(t_0 - \ell)\, e^{-\frac{\ell}{\tau}}}{\sum_{\ell=0}^{L} e^{-\frac{\ell}{\tau}}} \qquad \textbf{(equation 48)}$$

T(t) is the temperature at time *t.*

The listed features encompass all the relevant processes in a typical Central European medium range mountainous watershed. Thus, it is possible to transfer the PAI-OFF methodology to any catchment with the same characteristics. If PAI-OFF is used in catchments with different basic rainfall-runoff characteristics, features can be adapted in the course of the set-up of the system (e.g. it might be favorable to increase the total feature length for flat catchments with deep soils. PAI-OFF thus reflects the longer hydrological memory of such catchments). If this is not leading to a satisfactory performance, new features can be used for exploiting available flood relevant data routinely recorded in order to include the important specific information of the catchment. This applies for example for plain catchments in Northern Germany. Here, the runoff

process is influenced by water exfiltrating from the shallow groundwater storage. Additional features like e.g. groundwater depth can provide the necessary additional information. For the application in different climatic zones it is important to adapt the vegetation activity. The scheme shown in figure 40 demonstrates the incorporation of state features for the whole catchment and additional features from two orographic zones (OroZone 1 and OroZone 2) into the PAI-OFF system. (Orographic zones are explained in the following section)

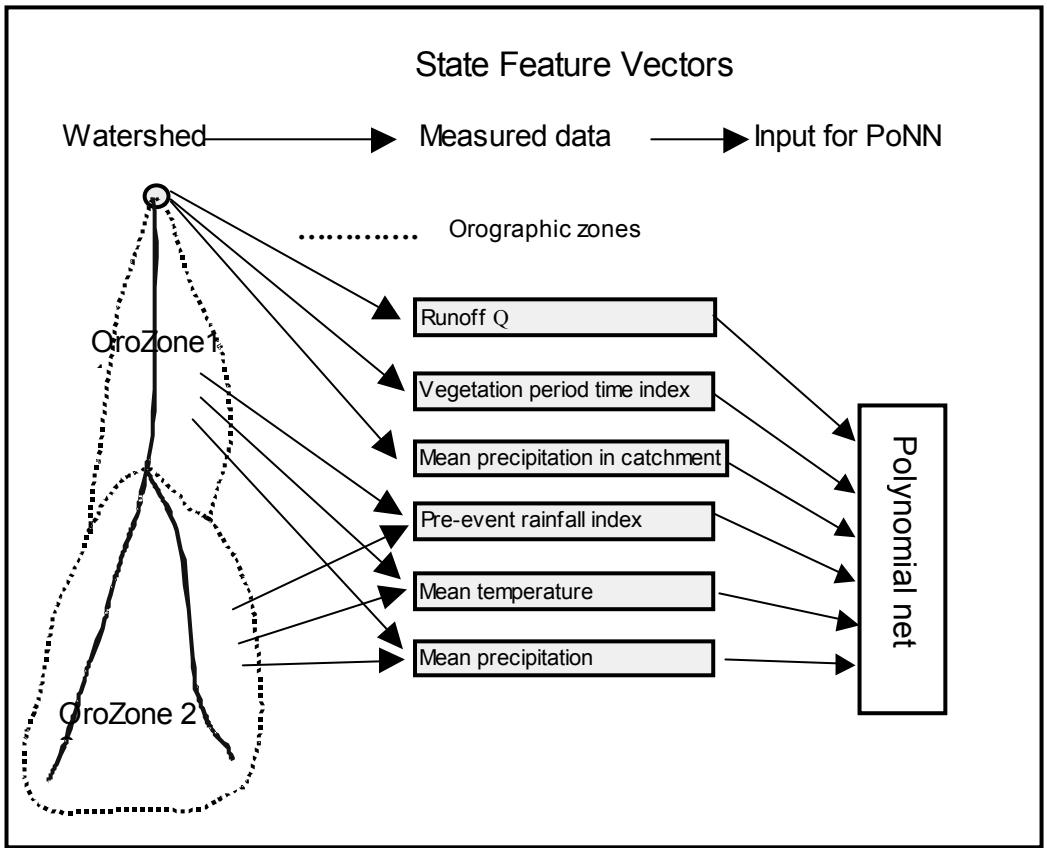


figure 40: Incorporation of state features into the PAI-OFF net

4.3 General feature selection strategy

Input vector selection is one of the key points of the PAI-OFF methodology. The PoNN includes the recorded information which is relevant for the flood formation in the considered catchment via the characteristic features. Generally, the considered watershed characterizes the input vector selection process. i.e. the

selection has to match the specifics of each modelled catchment. A comprehensive proceeding encompasses the principles of selecting representative features for any type of watershed, guaranteeing the overall applicability of the PAI-OFF methodology:

The selection process is a three-step procedure: The first step allows for considering important basic response characteristics at the entire basin scale, i.e. it covers the entire range of possible flood formation patterns in the whole catchment. Here precipitation and temperature as well as the observed discharge time series play the key role. These data yield integral information about the current state of the basin and its impact on the potential flood formation in reaction to a rainstorm. Consequently, a wide frame of features considering mean flows, low flow indices etc. is best suited for this first step.

In a second step, the study area is divided to consider sub basin scale hydrological particularities, again with the focus on characterizing the pre-event conditions. Therefore the orographic zones are delineated. An orographic zone is a part of the considered watershed which is defined by similar characteristic attributes. These attributes can vary according to the location of the catchment and the catchment characteristics, i.e. in a mountainous catchment the orographic zones are most likely to be defined according to elevation bands whereas in more uniform terrain the orientation might play the key role. The zones allow for taking into consideration special conditions of the watershed– i.e.orographic rainfall due to a mountain ridge within the basin etc.. The zones are particularly important if the rainfall-runoff process is distinctly influenced by the temperature regime, i.e., the processes of snow storage and snow melt affect the flood formation. Within the orographic zones, characteristic state features are derived by the formulae listed in the preceding section.

In a last step the watershed is divided into areas of similar hydrologic response. Within these zones precipitation is converted to hydrologic response features according to the procedure described in section 4.1. This opens the possibility to incorporate catchment specific flow paths into the neural network.

Generally it is not always indicated to use all the potential features in the net training process. On the contrary, if some features contain redundant information, it is best to eliminate these before the training process starts. This

makes the training faster and enhances the predictive power. Such a selection is always made in the set-up process. It is achieved by a preliminary application of the stepwise serial regression presented in chapter 3.2. Therefore, all features are offered to the regression. Then the ones which are contributing most to minimizing the objective function (equation 22) are selected for further use. If ΔR (equation 21) does not clearly reveal a threshold value for discriminating between important and redundant features, a semi-logarithmical graphical evaluation can help detecting the important features.

4.4 Pre-processing of relevant input information

Like all other neural networks, polynomial nets are sensitive to the data input characteristics (Kaastra and Boyd, 1995). Data pre-processing can have a significant effect on the model performance as different vectors span different ranges. Ensuring that all features receive equal attention during the training process (Maier and Dandy 2000) can be achieved by standardisation.

For all datasets considered in the PAI-OFF methodology, all characteristic features are normalized linearly in the range of [0,1] before the training. The normalization is based on the following standard equation:

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \qquad \textbf{(equation 49)}$$

with: X_{norm} = the normalized variable

X_i = the variable

X_{min} = the minimum value of the related variable

X_{max} = the maximum value of the related variable

4.5 Operational principles of PAI-OFF

The operation of PAI-OFF is easy, fast and reliable. A key point to ensure this quality is a comprehensive training database covering all possible types of flood formation patterns and all flood relevant catchment pre-event conditions. The

training database consists of two parts: the meteorological scenarios (input) as well as the runoff (output) section. The most important input data are precipitation and temperature, but generally, taking into account additional data (e.g. global radiation, wind speed, air humidity) allows for a more detailed portrayal of the catchment internal state. The PAI-OFF database consists of continuous hourly data for all the meteorological parameters. This data must be provided preliminary to the set up phase. The available meteorological data is typically a collection of observed events or periods. Usually, this does - by far - not satisfy the requirements of a comprehensive database. Therefore, a combination of historical data and synthetic rainfall events can be used to enlarge the database and fill "gaps" in the range of unobserved extreme precipitation events. To this end, PAI-OFF offers a characteristic rainfall generator (see appendix 2). With this tool, synthetic rainfall events can be inserted into existing time series of precipitation data. A prerequisite for the generation of precipitation events is a comprehensive meteorological characterisation of the considered watershed. This is necessary to provide the rainfall generator with the catchment specific information that is required to produce the precipitation events for the database. In detail the information provided to the generator are: maximum possible hourly rainfall rate, volume of the rainfall event, duration of the rainfall event, distribution and distribution characteristics of the rainfall (uniform, normal, skewness, noise), extension of advective events, path and speed of convective events (figure 63, appendix 2). Once the observed plus the generated events cover the whole range of possible flood patterns, the validated catchment model is transforming the meteorological section of the database to the output section. For each rainfall scenario an output vector characterizes the catchment response in the form of a discharge hydrograph at the basin outlet. According to flash flood characteristics, it is important that the time step of the data base be in hourly - or a finer resolution. If the available resolution is coarser, downscaling has to be employed to transform the data. A detailed description of data disaggregation goes beyond the scope of this work. For rainfall Lanza *et al,* 2001; Gyasi-Agyei, 2005; Sivakumar *et al,* 2001; Pegram and Clothier, 2001; Mackay *et al,* 2001; Koutsoyiannis and Onof, 2001 present a vast range of applicable approaches.

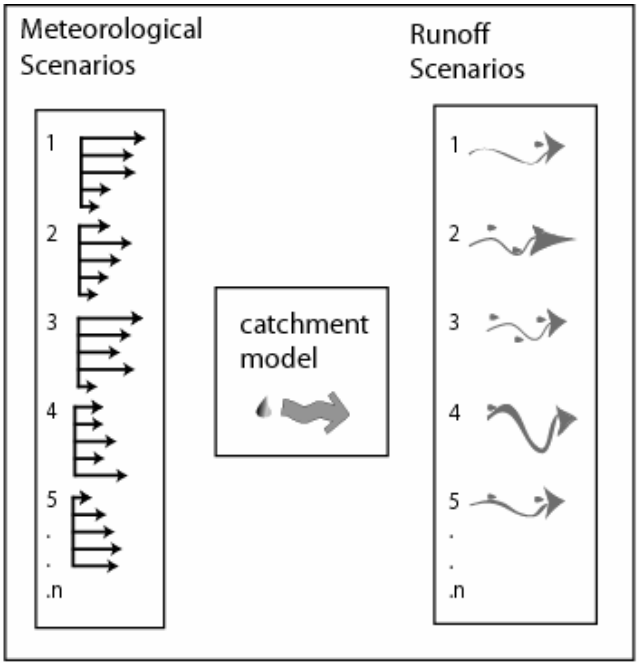


figure 41: Principle structure of the training database.

In figure 41 the structure of the database is exemplarily shown for 5 scenarios. The database consists of *n* pairs of meteorological scenarios and the corresponding runoff scenarios. Here, the number n depends on the result of the meteorological characterisation; it is the number of all possible, meaningful scenarios of rainfall and other meteorological variables influencing the runoff formation. As a result of the complexity of the processes involved, *n* results in values >1000 scenarios for a typical Middle European mountainous watershed. The vectors in the meteorological scenarios of figure 41 can be interpreted as different input parameters (i.e. precipitation, temperature, air humidity, etc.). The right hand side of the figure represents the catchment reaction, which is different for each input scenario. The established input and output sections of the training database subsequently serve for training the Polynomial Neural Network (PoNN) according to section 3.2.

5 Test application: Online flood forecasting at Kriebstein gauging station

The system performance is tested in this section. The target is predicting the 2002 flood with a peak flow error of less than 10 %. This flood is the largest event ever recorded at the gauging station. In figure 42 Kriebstein dam is shown under normal flow conditions. In the right fore of the picture, the spillways are active, inundating the power station during the 2002 flood. The dam is located about three km upstream of Kriebstein gauging station, the reference gauge of this application (figure 45).

figure 42: Kriebstein dam during the 2002 flood event (www.wikimedia.de/ www.doebelnerleben.de)

5.1 Study area

Kriebstein catchment area is situated in the East German Ore Mountains (coordinates: 51°04´- 50°23´ north; 12°47´ - 13°41´ east), it covers 1757 km².

The ridge of the mountain range - running from southwest to northeast - represents the upper boundary to the south of the watershed (figure 43). The catchment and consequently the flow direction are predominantly aligned from south to north. Two major rivers characterize the catchment. The eastern parts of the catchment are drained by the Flöha river, the Zschopaus main tributary. It merges with the Zschopau river after 75 km river length. The Zschopau river has a total length of 115 km to the gauging station in Kriebstein, the outlet of the considered area.

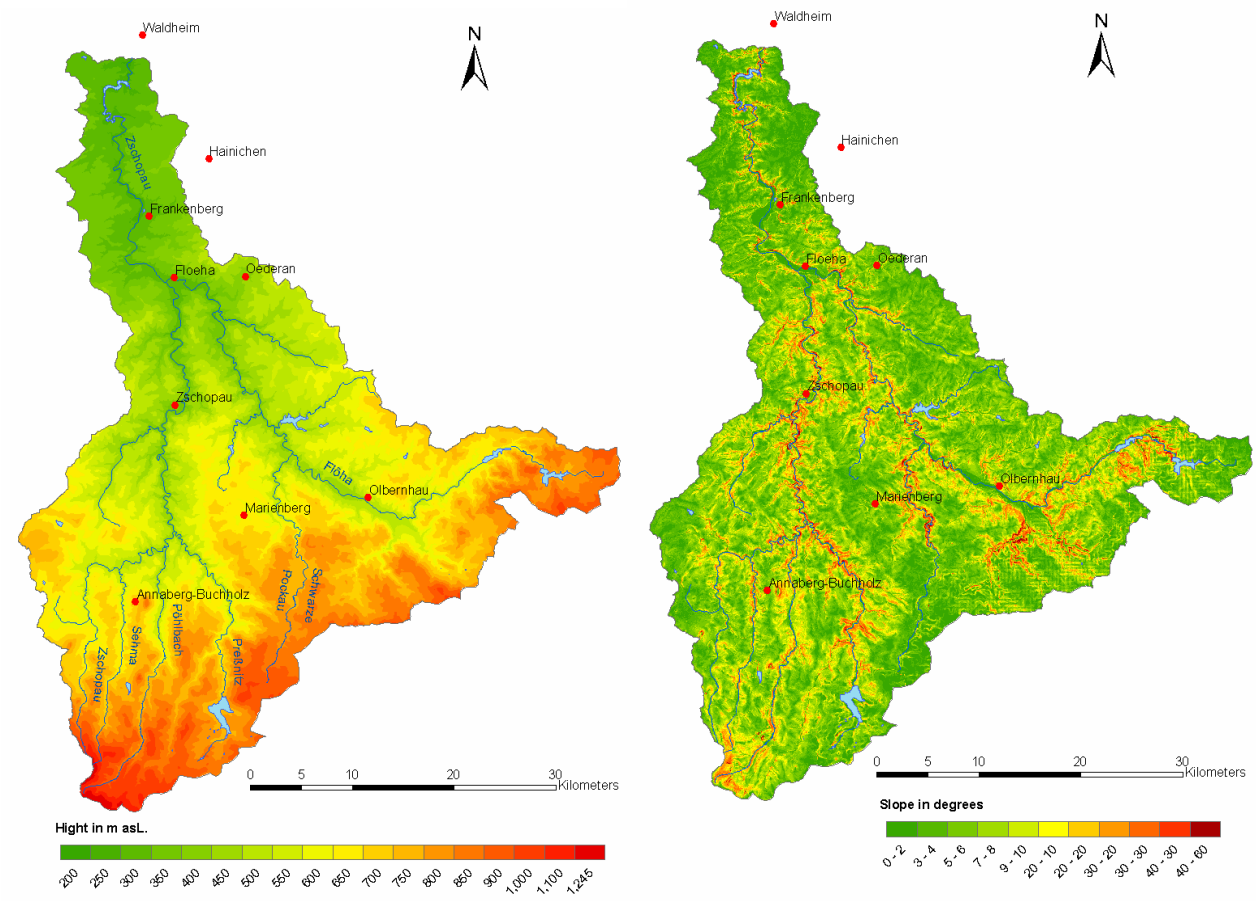


figure 43: The digital elevation model (left) and the slope distribution (right) of the test catchment

Characteristic flow figures are given in table 13 for two headwater gauging stations of the main stem and the most important tributary. These two headwater catchments account for almost 60 % of the total catchment size and largely dominate the runoff formation process of the catchment. In the table, the characteristic flows are also given as standardized values for a 600 km²

catchment size for better comparison of the two sub-catchments. The two catchments exhibit a very similar characteristic flow pattern. Only the absolute flows vary slightly due to the more pronounced conditioning of this figure by local rainfall distributions for the considered event.

table 13: Flow characteristics for the headwaters of Flöha (Borstendorf gauge, 644 km") and Zschopau (Hopfgarten gauge, 529 km²)as well as the standardized 600 km² catchment.

	Zschopau [m³/s]	Zschopau 600 km² [m³/s]	Flöha [m³/s]	Flöha 600 km² [m³/s]
Absolute low flow	0.1	0.1	0.2	0.2
Mean low flow	1.4	1.6	1.67	1.6
Mean annual flow	7.9	9.0	9.2	8.5
Mean winter flow	9.7	11.0	11.2	10.4
Mean summer flow	6.2	7.0	7.1	6.6
Mean high flow	81.2	92.1	91.3	85.1
Absolute high flow	420	476.4	540	503.1

Meteorology

Mean annual precipitation in the catchment ranges from 600 to 1200 mm. The higher rates are caused by orographic precipitation near the Ore Mountain ridge. This is due to the strong tailback effects observed when the predominant westerly winds hit the mountains. The lower rates are observed in the northern part of the watershed, which is quite flat and does not benefit from orographic precipitation. The lowest monthly precipitation is recorded in February and November. Highest rainfall rates are measured in July, when most solar energy is available for intensive convective precipitation. From April throughout September, nine heavy precipitation events occur on a monthly average. Depending on altitude and exposition, the snow coverage ranges from 25 to 200 cm in average for the winter months. The snow cover might last from November throughout April on higher and northerly exposed surface areas. Mean temperature is around 10°C in the lower areas of the catchment. This figure falls to about 6°C in areas near the ridge. For the vegetation period the temperatures

average between 14 and 17°C. table14 exhibits the means monthly temperatures of for Fichtelberg a station representative for the high catchment areas. The corresponding evaporation rates range from 300 mm/a to 600mm/a, this is roughly half of the precipitation quote.

table 14: Mean monthly temperatures of Fichtelberg station (1992-1998).

°C	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
Fichtelberg	-3.55	-3.13	3.78	8.43	11.68	16.50	15.23	14.78	11.50	3.50	-0.13	-0.95

Geology and morphology

The Ore Mountains emerged as a result of the alpidic formation, when high pressure was exerted by the North American continental drift. As a result of the alpidic block massive, constantly eroded by the rivers, the Ore Mountains today present themselves as a furrowed plateau gently inclined from north to south. The southern part of the watershed is steeper (15%), while the northern part of the catchment only exhibits slopes of around 3 %. The bedrock of the Ore Mountains consists of by gneiss, mica slates and granite. Near the ridge, at altitudes around 900 MSL, the slope abates; elevated plains favour the development of fens. Valleys with steep slopes and broad bottoms characterize typical landscapes. In the lower reaches, the valleys open up, the floodplains broaden and the slopes are less steep.

Soils

Characteristic soils of the Zschopau catchment are dominated by silt, loam and clay. The soils are largely a product of gneiss weathering. In the upper regions soil depth varies from a few cm to about 1m. The thin soil layers are mainly attributed to the steep slopes of the upper valleys. Thick soil layers can be found in valley bottoms and on elevated plains. Soil depths of more than 3m can be found in the lower parts of the catchment. The main soil types are: silt (42%), clay sand (37%), 21% sandy loam and -marginal- organic soils in fens. In the northern parts of the catchment the fraction of clay and silt increases. For a

better understanding of the soil morphology, the lithofazies are illustrated in figure 44.

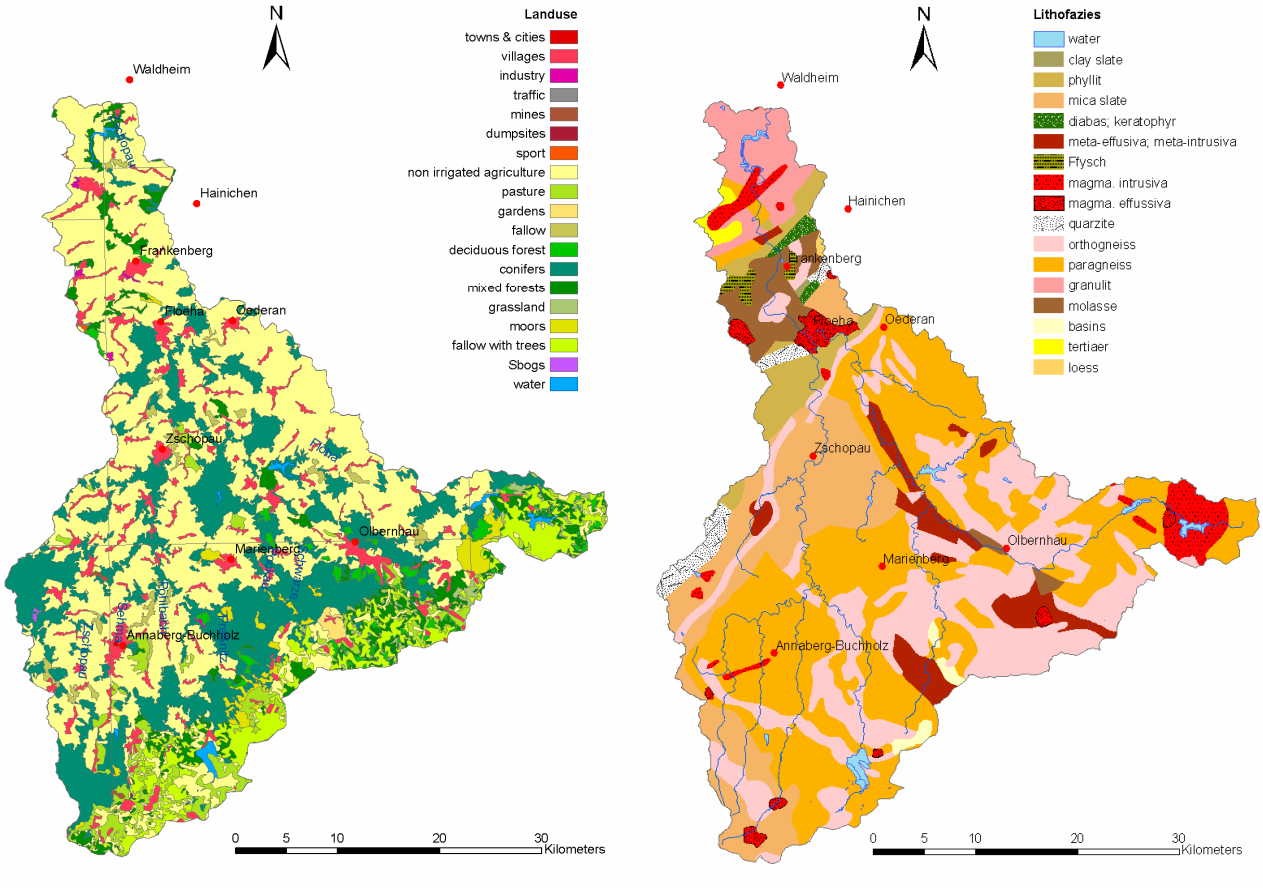


figure 44: Landuse (left) and Lithofazies (right) of the test catchment

Landuse

Agriculture is the dominant landuse in the catchment, it accounts for 64 % of the area (figure 44). On high or steep areas stock farming prevails. In the lower, warmer and gently shaped parts crop farming is the predominant agricultural characteristic. The later accounts for about 80 % of the agricultural activities within the watershed. Roughly 5 % of the catchment is covered by human settlements. 29 % is attributed to forest. Here again a pronounced dependency on the altitude divides lower deciduous and mixed forests from the purely coniferous stands on the high catchment areas. The effective root zone in the

forests reaches down to 90 cm. This values is somewhat higher for crop stands with up to 120 cm.

figure 45: Zschopau (mean flow) at Kriebstein gauge cross section (in flow direction).

5.2 Hydrological catchment modelling

The WaSiM model is used to describe the basin response characteristics for the test application. The catchment model is set up according to the dynamic parameterisation strategy (described in section 2.5.2).

5.2.1 Data and model structure

Consistent meteorological data is available for the years 1953 throughout 1999 for the whole catchment on a 1 km² grid. This data includes precipitation, temperature, wind speed, global radiation and relative air humidity. WaSiM is driven by this data. Wind speed, global radiation and air humidity are used

exclusively in the Penman evaporation module. The temperature additionally serves to run the snow module. Furthermore soil and landuse grids are required to run WaSiM. This data is derived from the Bük 200 and Corine (2000) respectively. Runoff data is available for the events shown in table 15 at Kriebstein gauge

table 15: Available runoff events for model testing (events with hourly measured data are italic).

02.07 – 06.08.1954	01.01 - 03.02.1968	*02.05 - 23.05.1978*	*14.03 - 25.04.1988*
01.01 - 20.05.1955	*28.03 - 19.05.1969*	*18.07 - 08.08.1980*	01.01 - 06.02.1989
01.07 - 02.09.1955	*06.04 - 01.06.1970*	17.03 - 27.03.1981	*29.06 - 20.07.1992*
19.07 - 09.08.1957	*06.03 - 21.05.1973*	01.01 - 26.02.1982	*08.03 - 29.03.1993*
20.06 - 01.08.1958	10.01 - 07.02.1974	*29.07 - 26.08.1983*	*27.05 - 01.07.1995*
14.10 - 25.11.1960	01.12 - 31.12.1974	*23.05 - 13.06.1986*	*26.08 - 16.09.1995*
05.03 - 06.06.1965	04.01 - 29.01.1975	01.01 - 16.01.1987	*29.06 - 27.07.1996*
16.01 - 24.02.1967	*25.07 - 22.08.1977*	*30.01 - 24.04.1987*	*07.03 - 04.04.1998*
10.03 - 20.04.1968			

For thorough modelling, events with hourly raw data quality are considered exclusively. This fact significantly reduces the number of available events. The suitable events are shaded in table 15. One exception is made here. The late 1974 event is included in the parameterisation study because of its exceptional tempestuousness.

WaSiM is set up following the natural structure of 17 subcatchments (figure 46). Seven of these subcatchments are equipped with gauges (table 16). This structure also allows for a distributed “internal” validation of the catchment model whenever data is available. Where possible, all available runoff data of the subcatchments are used in the calibration process of the model.

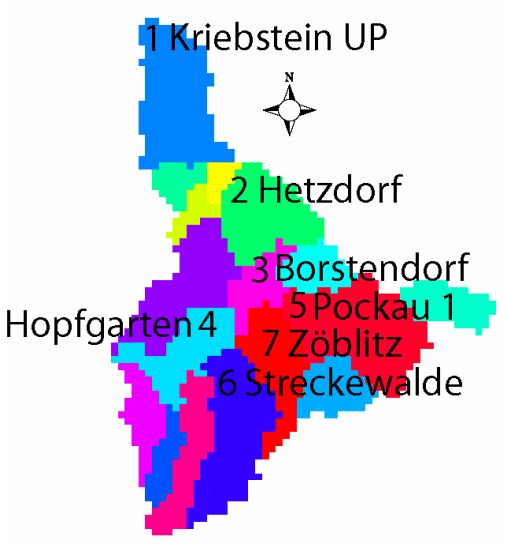


figure 46: Schematic illustration of the 17 subcatchments and 7 gauging stations used in the study

table16: Gauged subcatchments of the study area

Nr.	Gauging station	River	Area [km²]
1	Kriebstein UP	Zschopau	1757
2	Hetzdorf	Flöha	760
3	Borstendorf	Flöha	644
4	Hopfgarten	Zschopau	529
5	Pockau1	Flöha	385
6	Streckewalde	Peßnitz	206
7	Zöblitz	Schwarze Pockau	129

5.2.2 Model performance for a singular parameter set in the test catchment

The event data shown in table 15 has been checked for a consistent model input data quality. After this check only 13 events remained for the testing. The events are shown in figure 48. The 1996 event is automatically calibrated minimizing the rooted mean square error of observed and modeled peak flows with the aid of the SCEUA-method. A number of 40 different start values was used to avoid

"getting stuck" in a local minimum during the optimization process. The 1996 event has been selected because it is a typical medium-sized event. It was caused by advective rainfall and possesses no singular characteristics. It is therefore well suited for calibration purposes. The performance of the calibrated model is shown for all the available internal gauging stations in figure 47.

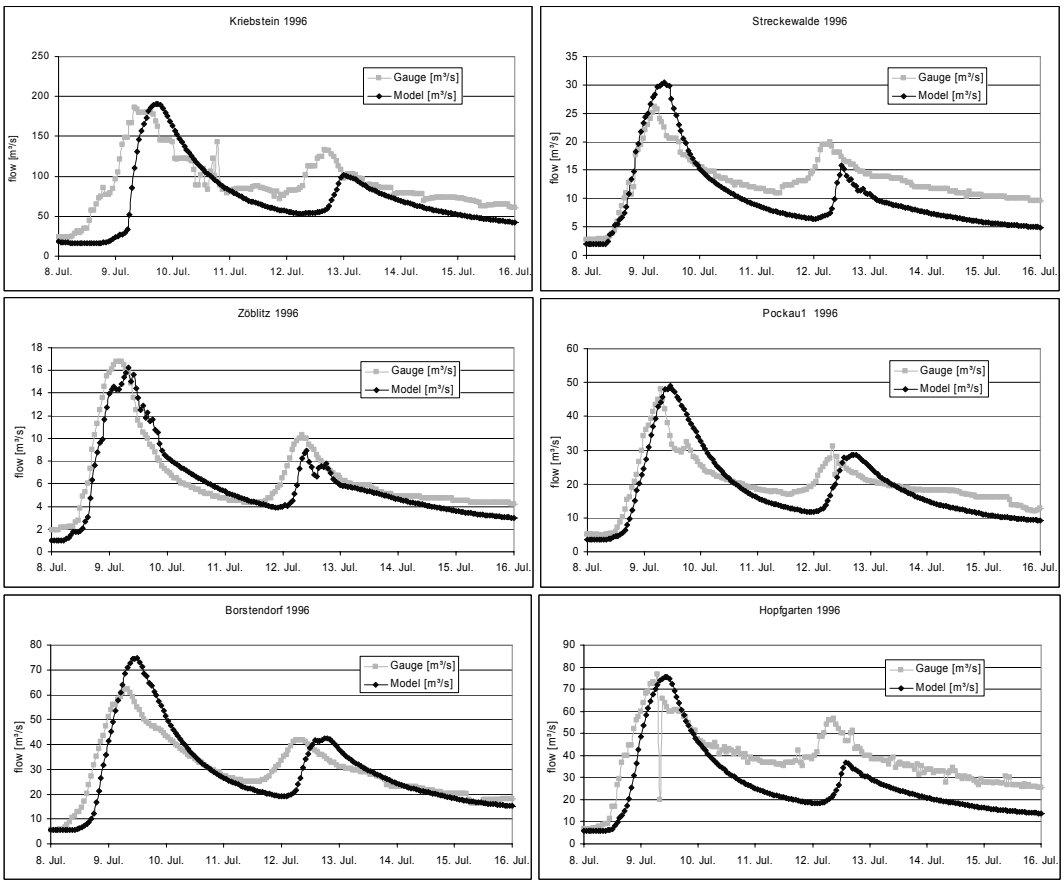


figure 47: Calibrated WaSiM performance for 1996

The time lag at Kriebstein gauge catches the eye at first sight. It is most probably due to the erroneous interpolation of rainfall data for the lower part of the catchment. The headwaters are less exposed to this source of error as shown in figure 47. The validation of the calibrated model is shown in figure 48. Validation of the smaller and medium sized events yields a convincing agreement between observed and modelled flow peaks at Kriebstein gauging station as long as the model is validated for similar maximum peak discharges. For the four highest events of figure 48 the model calibration leads to unsatisfactory validation results.

table 17: Mean relative peak error of the 12 validated events

	All events	Large events	Small events
Mean relative peak error [%]	36	78	15

In table 17 the visual interpretation of figure 48 is confirmed. The RSME for the four largest events is more than 5 times higher than the validation error criterion of the eight smaller flood events. This outcome is not surprising if we recall that singular model parameter sets are not able to account for all possible types of event characteristics. The processes dominating extreme floods are different than the governing processes of common floods. The model is calibrated for the 1996 event, the optimal model parameter set reflects the specifics of this rather common flood event with a return period of ~ 15 years. Therefore the model completely fails to represent the extreme floods with return periods > 30 years.

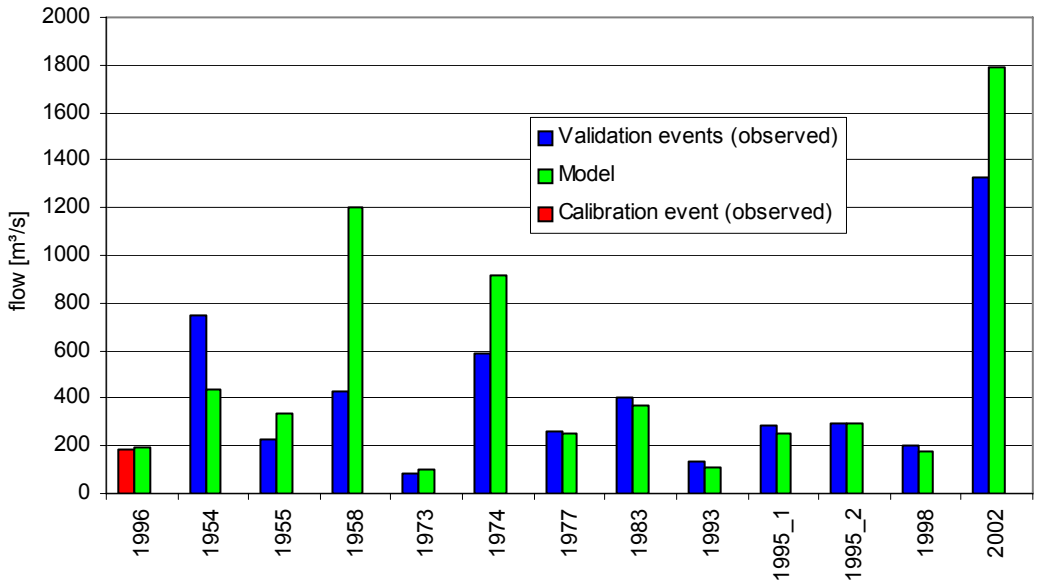


figure 48: Calibration and validation performance of the singular parameter set for small events.

WaSiM As a consequence of the awareness that an approach based on a singular parameter set is unattractive for flood forecasting the dynamic calibration strategy (section 2.5.2) is introduced for the test application:

5.2.3 Dynamic parameterisation for the test application

As described above, it is extremely difficult, if not impossible to describe the whole spectrum of flood events with one single model parameter set. For modelling the discharge at Kriebstein, the considered events have therefore been divided into two classes. Separation criteria are the peak discharge rates. The first class comprises common flood events up to a return period of about 20 to 25 years. The second class of events represents the “extreme” floods, it comprises the four highest events ever recorded (the flood events which could not be validated successfully with a single parameter set type of model in 5.2.2). For the two classes, two WaSiM models are calibrated and validated separately. The parameter set for the common event class was calibrated for the flood event in the summer of 1996, according to the procedure described in 5.2.2. The extreme flood parameter set was calibrated for the 2002 flood event and validated for the three other data sets (1954, 1958 and 1974). The calibration criterion is the weighted sum of the timing error (measured-modelled) of the peak discharge and the RMSE of the peak discharge value. This criterion reflects the basic purpose of flood forecasting: Human life and material values are to be protected from loss or damage due to flood events by the earliest possible warning in terms of the timing and magnitude of a dangerous event.

The calibration performance of the model describing the extreme flood events is shown in figure 49. The slight deviations in the rising limbs between observed and modelled hydrographs are the result of inadequate input data interpolation, not covering the first little “pre-peak”. The systematic underestimation in the falling limb of the hydrograph is not covered by the calibration criterion and can therefore not be addressed with this parameter set. However, this does not interfere with the aim of predicting the rise and the peak of the considered flood events. Here, the calibration is excellent and robust in the context of the convincing peak fit for all the sub basins. The performance of the parameterisation for the common events is not shown, it is equal to the singular parameter set performance and shown in figure 48.

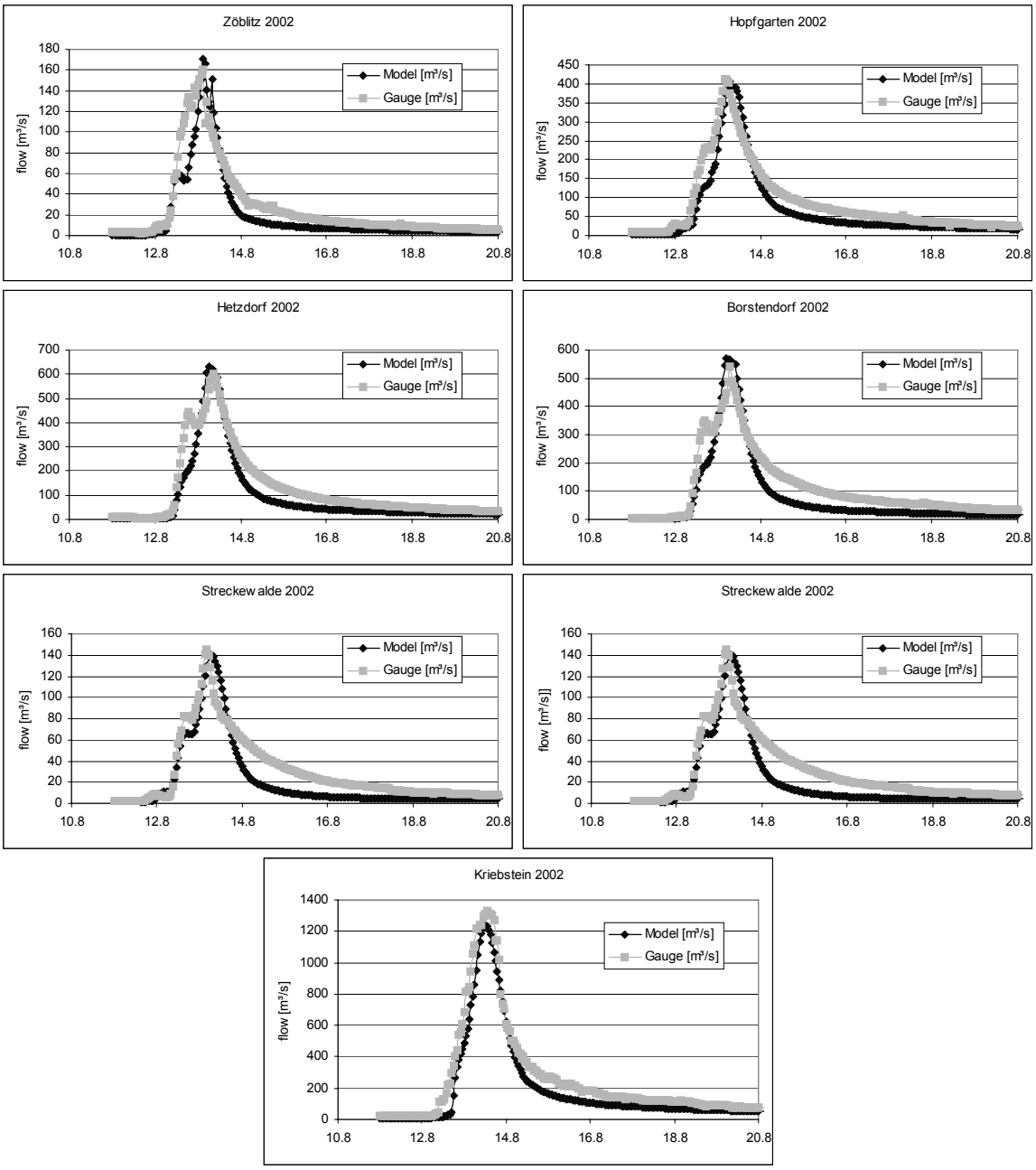


figure 49 WaSiM performance for the 2002 extreme event at the 7 gauging stations (calibration)

Validation the extreme class parametrisation leads to the results shown in figure 50. The events 1954 and 1974 are well portrayed with a mean relative peak error of about 10%. The validation of the 1958 event is of poorer quality. Taking into consideration the poor data situation and the difficulty to re-check the input data

this result can not be explained satisfactorily. Nevertheless, the theory about the bad data quality of the 1958 event is supported by the validation performance of the singular parameter set WaSiM. Here also, the 1958 event is heavily over-predicted.

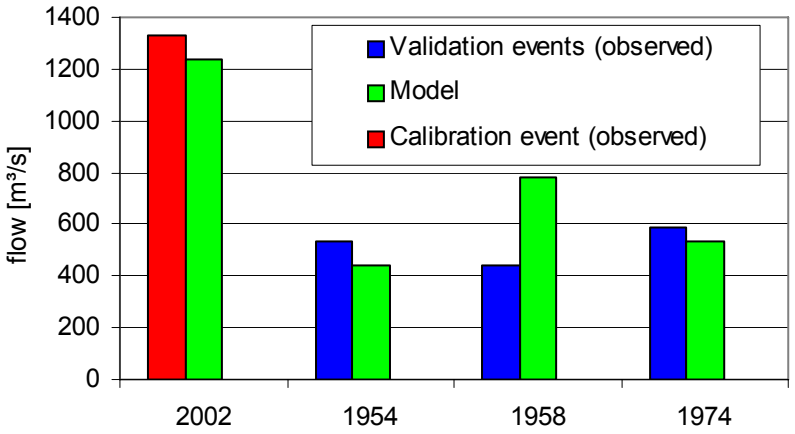


figure 50: Calibration (red) and validation (green) for extreme events for the dynamic method.

At this stage the principle of sigmoidal weighting (described in section 2.5.2) is applied to merge the results of the different parameter sets. An exemplary event from the extreme class is shown in figure 51. It can be seen that the merged model for this event (the dark blue line) well follows the parameterisation for the extreme event class for the peak flow. In the recession phase, when the observed flow falls back into the range of the small event class calibration this becomes more dominant in the merged model.

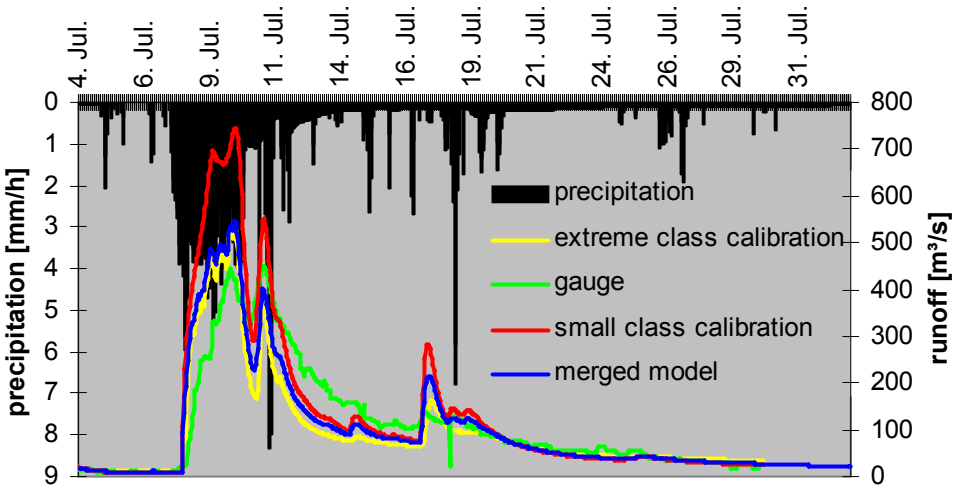


figure 51: Sigmoid-merged hydrograph of the 1954 event

The excellent validation results for all events are shown in table 18 and in figure 52. The overall relative peak error is decreasing from 36 % to 16 % when applying the strategy of dynamic parameters. This is a direct consequence of the additional process information introduced into the calibration of the extreme class of events. For this class the peak error decreases from 78 % to 15 %. For completing this statement, figure 66 in appendix 4 shows the calibration and validation performance of the extreme event parameter set for all 13 events used in this study. The overall relative peak error of this approach is 31 %. This is somewhat smaller than the value given for the small class, but it has to be seen in the light of the very large errors for the few extreme events occurring for the small class parameter validation. Nevertheless, this result confirms the statement that the dynamic parameter set is superior to singular set with respect to the ability to correctly portray different types of flood peaks.

table 18: statistical errors of the 11 validated events.

	All events	Large events	Small events
Mean relative peak error [%]	16	25	15

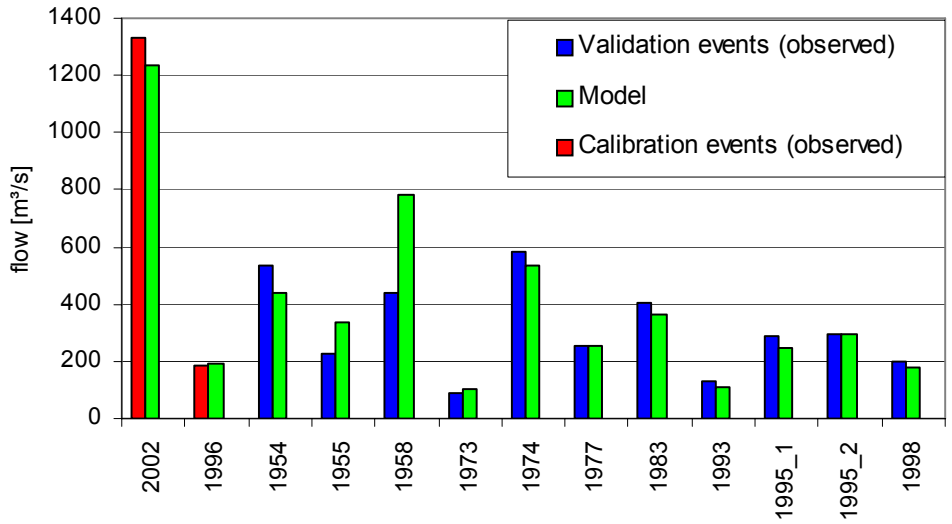


figure 52: Dynamic calibration (red and validation (blue) Kuerlb

This application clearly shows the potential of the dynamic parameterisation and recommends this strategy for the generation of the PAI-OFF net training database described in the following section.

5.3 Training database

The training data base is built according to the procedure described in section 4.5. First, the meteorological data is prepared. Here, 47 years of measured wind speed, relative air humidity, temperature, sunshine duration and precipitation data are kriged to a 1km² grid that covers the whole catchment. This data serves as a framework for the generation of the rainstorms, which are inserted into the natural precipitation time series. Because the test application adheres to the dynamic calibration approach described in section 2.5.2, this process is executed twice for all of the 1757 cells of the investigated area. The data set for modelling the common class parameter set contains events of up to 25 years return period. The extreme class parameter data set consists of rainstorm data which lead to flood event with a return period > 25 years. The database now consists of two sets of 47 years of meteorological data; this leads to a total of 94 years (823440 hourly time steps) on the input side of the training database. Advective as well as convective storms are taken into account in the database. These events are generated for given rainfall intensities and storm durations on the basis of the

probable maximum precipitation of the area (DWD 1997). The rainstorm scenarios mirror the catchment typical meteorological behaviour, i.e. the parameters of the generated rainfall hydrograph which portray the local situation with respect to:

- ➔ The event shape and skewness;
- ➔ The drift direction and velocity of advective storm events;
- ➔ The location of the rainstorm centre and the radius of convective events.

Basic details of the precipitation generator are described in appendix 2. The most important features of the generated precipitation series are shown in table 19. Here the main difference between the normal and extreme set is the generated number of events per year. This arises from the simple fact that the extreme events are characterized by much longer recession phases. No further events are inserted into these periods because this would lead to an unrealistic soil moisture state of the catchment for the training data. The total training data comprises less than the 823440 potential time steps resulting from the two model parameterisation variants for 47 years of hourly data. This is a result of concentrating this exemplary study to summer flood events. This practical approach allows for a faster training, whilst the forecast performance for the test flood does not suffer from the restriction to the snow-free periods, because it occurred in August.

table 19: Characteristics of the two generated precipitation series.

Series	Years	Nr. of years	Events per year	Total nr. events	Total training time steps
Normal	1953-99	47	28-30	1351	240264
Extreme	1953-99	47	13-18	345	236656
∑	-	94	-	1696	476920

In the second step of setting up the database WaSiM is used to transform the meteorological input data to the corresponding hydrograph according to figure 41. Because of the rapid catchment response to rain storm events (especially in the steep upper parts of the catchment) we use a time step of 1 hour for all processes involved. This refers especially to rainfall series, wind velocity,

radiation, temperature and air humidity. The transient character of seasonal variation with respect to plant growth and its impact on soil water dynamics and evapotranspiration is taken into account by dividing the year into four distinct vegetation periods. The soil hydraulic parameters originate from the official soil map of the area (Bük 200) and the land use data was derived from Corine (2000). The Zschopau river down to Kriebstein is characterized by a steep bottom slope without backwater effects, which allows employing the translation – diffusion approach for the flood routing. As it has been discussed in section 2.5.2 it is not feasible to operate a model with a single set of parameters in the context of reliable flood forecasting. Therefore, this test application takes full advantage of using the strategy of dynamic calibration. The result of the two WaSiM model runs (one for the extreme meteorological data set with the extreme class parameter set and one for the common data and parameters set) is then merged according to equation 7. Now the training database is fully established and the characteristic input vectors have to be defined in order to train the polynomial neural network.

5.4 Characteristic features for the Kriebstein polynomial net

This section describes the selection of input for the net training. All the possible features described in section 4 are tested by means of the stepwise serial regression method used also for training the net. With this method it is easy to discriminate the important features from redundant information. As illustrated in section 4.3, the feature selection process starts with characteristic features representing the entire test watershed. For Kriebstein the result of stepwise serial regression yields the features listed in table 20. In the second step, the catchment it is divided into orographic zones. This is necessary because the temperature and precipitation regimes vary with the height above mean seal level. For the test watershed three orographic zones (OroZone) are used. Their delineation is based on a well-balanced representation of different characteristic temperature zones in the model by means of elevation band thresholds. This approach results in the Orozone 10000, which covers the area below 400 MSL OroZone 10400 reaches from 400 to 800 MSL, and the uppermost OroZone 10800 represents the parts of the watershed with heights > 800 MSL.

table 20: State features covering the whole catchment in the test application

Number	Description
1	Mean actual precipitation for the whole catchment
2	Mean monthly precipitation for the whole catchment
3	Vegetation activity index
4	Absolute monthly low flow
5	Absolute monthly peak flow
6	Mean weekly flow for the whole catchment
7	Monthly mean flow
8	Gradient of flow over the last 6 hours
9	Actual flow at Kriebstein gauge

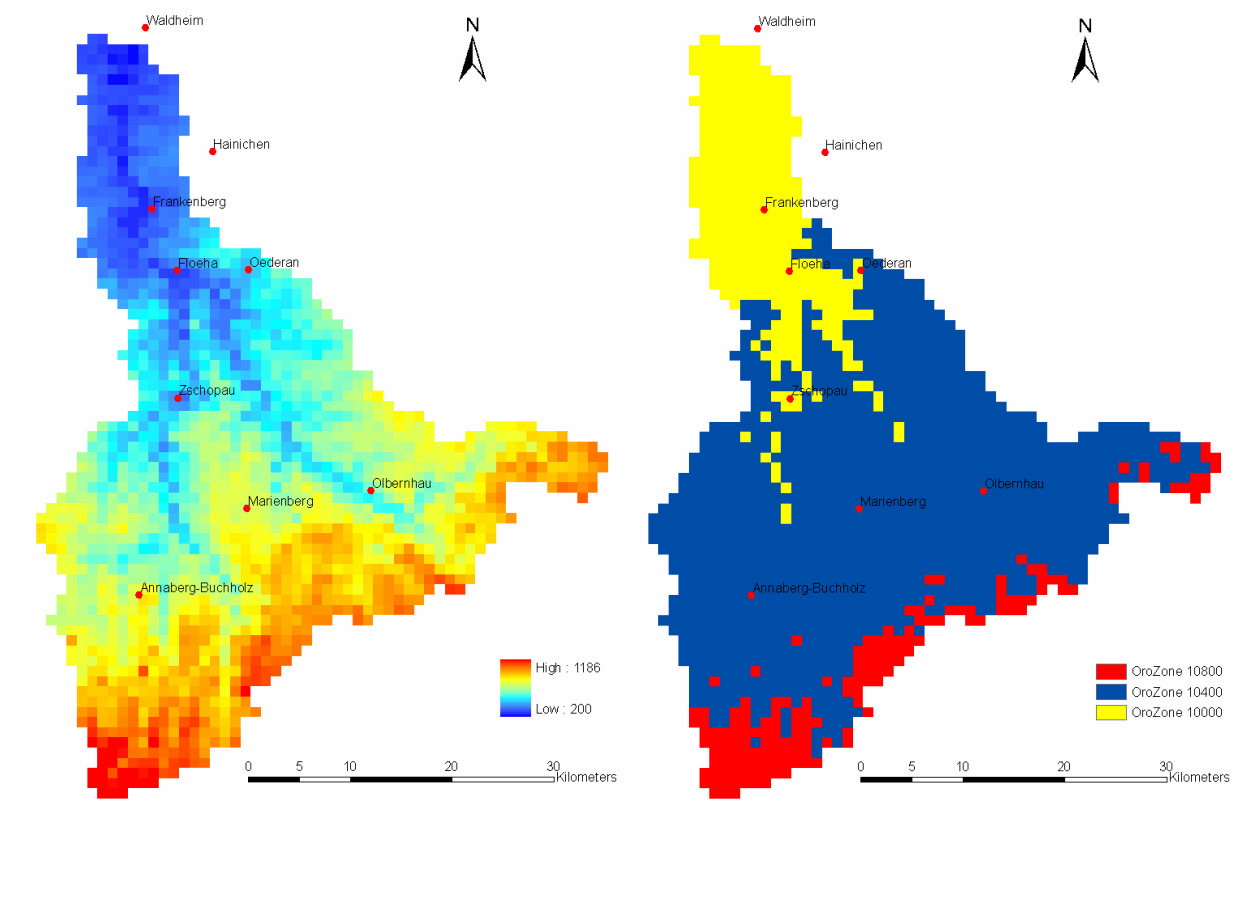


figure 53: Kriebstein: digital elevation model (left) and orographic zones (right)

The features for the zones are again evaluated by means of stepwise serial regression. In table 21, the significant features are listed for all zones.

table 21: Orographic zones based state features used in the test application of PAI-OFF for Kriebstein gauging station.

Number	PAI-OFF internal code	Description
10	WMN10000	Mean weekly precipitation in OroZone 10000
11	WMN10400	Mean weekly precipitation in OroZone 10400
12	WMN10800	Mean weekly precipitation in OroZone 10800
13	WMT10000	Mean weekly temperature in OroZone 10000
14	WMT10400	Mean weekly temperature in OroZone 10400
15	WMT10800	Mean weekly temperature in OroZone 10800
16	TMT10000	Mean daily temperature in OroZone 10000
17	TMT10400	Mean daily temperature in OroZone 10400
18	TMT10800	Mean daily temperature in OroZone 10800

It is eye-catching, that the stepwise serial selection of relevant state features (tables 20 and 21) only identifies features based on precipitation and temperature. These two processes are of overarching importance for the flood formation. One exception from this rule is the vegetation activity index, which is a significant source of information for the polynomial net in the test catchment. It is unnecessary to incorporate humidity and global radiation as input features into the forecasting system for the test site. The vegetation activity alone is suitable to describe the effects of transient climatic characteristics for the catchment, thus portraying the annual changes in the catchment. This is a typical phenomenon for the latitude of the test catchment. The vegetation activity is a mirror of global radiation. It also correlates with the dynamics of the daily humidity dynamics.

The third step divides the watershed into area of similar hydrologic responses (AHR). This is done on the basis of overland flow times (figure 54) calculated according to Schulla (1997). The travel times are then combined in groups starting from points along the channels in the catchment (table 23). Kriebstein

gauge is the first outlet of a area of similar hydrologic response. The other zones are then defined one by one moving up the channel network. The criterion herein is the formation of groups with not more than 4 hours internal flow time difference. This holds for all area of similar hydrologic responses except the one, which is situated just above the catchment outlet. This area of similar hydrologic response comprises eight hours instead of four. This is due to the relatively smooth topology, where overland flow is inert.

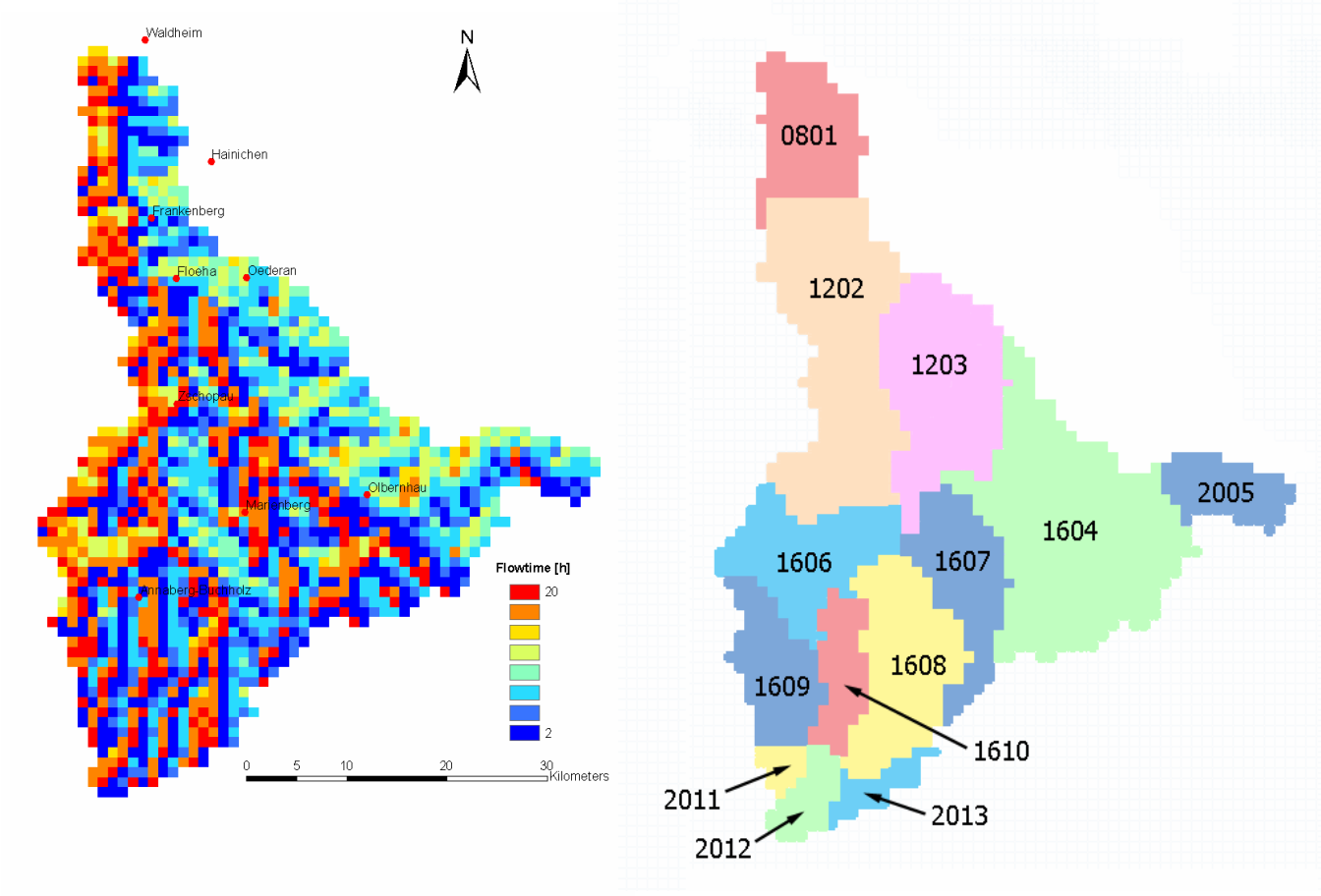


figure 54: Travel time (left) and areas of similar hydrologic response (right) for the test watershed

For each area of similar hydrologic response, one hydrologic response feature is determined according to the standard procedure described in section 4.1. The relevant features are listed in table 22.

table 22: Hydrologic response features used in the test application of PAI-OFF for Kriebstein gauging station.

Number	Description
1	Hydrologic response feature for area 801
2	Hydrologic response feature for area 1202
3	Hydrologic response feature for area 1203
4	Hydrologic response feature for area 1604
5	Hydrologic response feature for area 1606
6	Hydrologic response feature for area 1607
7	Hydrologic response feature for area 1608
8	Hydrologic response feature for area 1609
9	Hydrologic response feature for area 1610
10	Hydrologic response feature for area 2005
11	Hydrologic response feature for area 2011
12	Hydrologic response feature for area 2012
13	Hydrologic response feature for area 2013

After all the characteristic features are available, the product vectors are derived by permuting the total of the state and hydrologic response features by means of equation 17. The total number of possible permutations is restricted in the operational PAI-OFF by a simple rule (trimming): The combination of state features and hydrologic response features of differing subcatchments is suppressed. This trimming is based on to the fact that combining the hydrologic characteristics of a certain sub basin with the mean temperature of a far away and disconnected area does not yield any meaningful process information. In spite of this technique, the total number of vectors employed is still considerable (table 24).

table 23: Mean travel time of similar hydrologic response areas of figure 54

Time Travel Zone	Travel Time (h)	Time Travel Zone	Travel Time (h)
Zone 0801	08	Zone 1608	16
Zone 1202	12	Zone 1609	16
Zone 1203	12	Zone 1610	16
Zone 1604	16	Zone 2011	20
Zone 2005	20	Zone 2012	20
Zone 1606	16	Zone 2013	20
Zone 1607	16		

In the net training the hydrologic response features are used in the "raw" form p as well as in the powered forms p^2 and p^3. The application of the powered hydrologic response features improves both training (43 % MSE) and validation (38% MSE) of the PoNN if compared to the lagged kernels described Schmitz *et al* (2006). The set of features in the training contains i = 18 state features and j = 13 „raw" hydrologic response features (39 hydrologic response features when the powered features are counted). Appendix 3 contains a tableau of the features used in the application. The possible combinations for three state features and three hydrologic response features are exemplarily visualized in figure 65 (appendix 3).

table 24: Number of product vectors used in the training

Product vectors	Trimmed product vectors
45599	14957

All of the PAI-OFF features are now available. Together with the database described earlier they are the basis of the PoNN training with stepwise serial regression.

5.5 PAI-OFF Training

The training is performed using data from 1953-1971 and 1982-1999. This equals a total of 1060 flood events, describing all the possible flood formation constellations in the test catchment. The years 1972-1981 are excluded from the training and used for testing (validation) the predictive performance only. The validation part of the database consists of 150 flood events. The training in this test application adheres to the principles described in section 3.2. A separate PoNN is trained for each pair lead-time step from 2 to 48 hours separately. This results in 24 training loops for 24 separate nets. Each loop employs stepwise serial regression to train the matrix O, which encompasses 180 product vectors in the test case. In this section the focus is on determining how well the PAI-OFF PoNN portrays the hydrological catchment model.

Evaluating the PoNN performance with the part of the database, which is reserved for validation, gives evidence about the predictive performance of the final operating system. To this end, three criteria are used. The first criterion is the dimensionless Nash-Sutcliffe-Efficiency. It is one of the most commonly used measures to assess the predictive power of hydrological models. A second criterion (figure 55) is the peak-to-peak error (PPE), calculated from the database output and the predictions of PAI-OFF. This relative error describes the mean difference between the peak flow values for all the events used in the validation process. And last but not least, the PPT describes the error in the peak timing in hours.

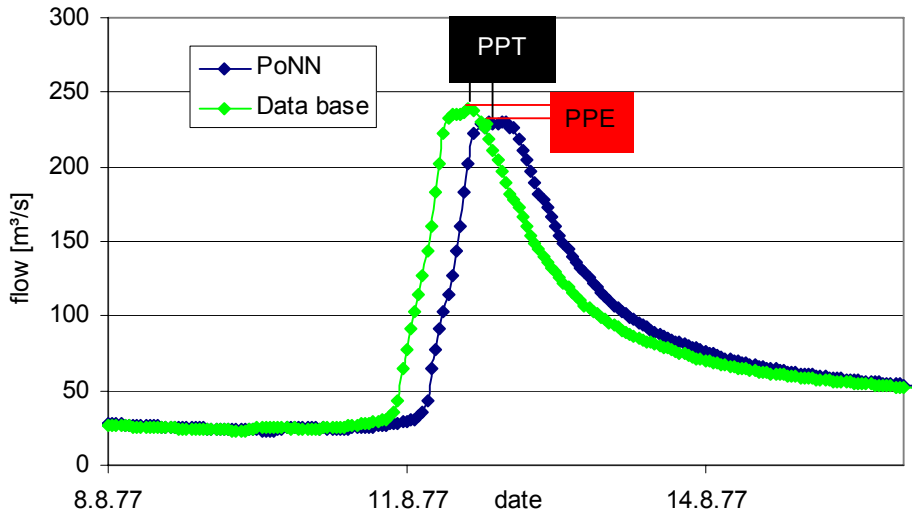


figure 55 Error criteria used to evaluate the predictive power of the PoNN

The training results shown in figure 57 are excellent, the Nash-Sutcliffe yields values > 0,97. This also holds for the validation period which encompass the events not used in the training process (figure 57). This result confirms the overall predictive power of the approach.

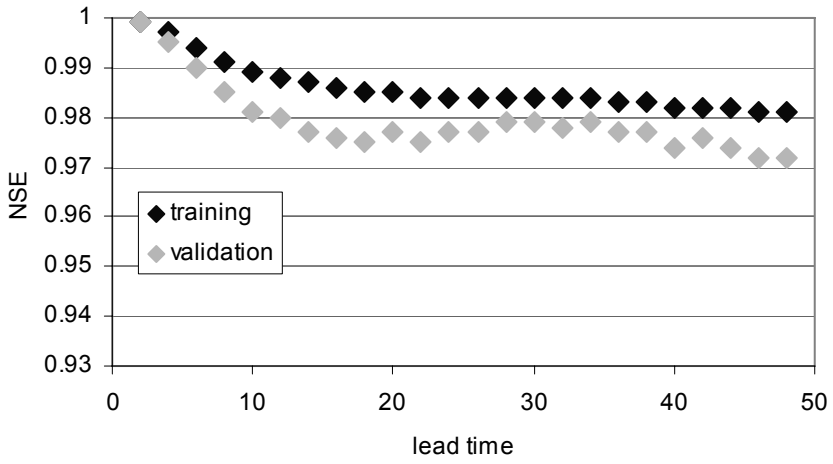


figure 56: Training and validation of the net (Nash-Sutcliffe-Efficiency).

The Peak-to-Peak Error, which is evaluated in figure 57, allows for a more detailed view the flood peaks. The picture reveals a slight underestimation of the peak flows for lead times > 10h. For shorter lead times the peaks are slightly overestimated. Nevertheless, as the mean relative error does not exceed 4%, the

PoNN fulfils the requirements of a reliable flood forecasting system in terms of the ability to correctly predict the peak flow.

In figure 58 the peak-to-peak times are evaluated. The overall behaviour of this error criterion well correlates with peak flow error (PPE). The figure reveals that the net predicts early peaks for lead times > 15, while for smaller lead times the net tends to predict late peaks. The mean error is less than 1h for all of the trained lead times. As the systems time step in the test application is one hour, the peak-to-peak time error of the system is negligible for the test events. All criteria considered, the system is able to predict the peaks at the right time with the desired accuracy.

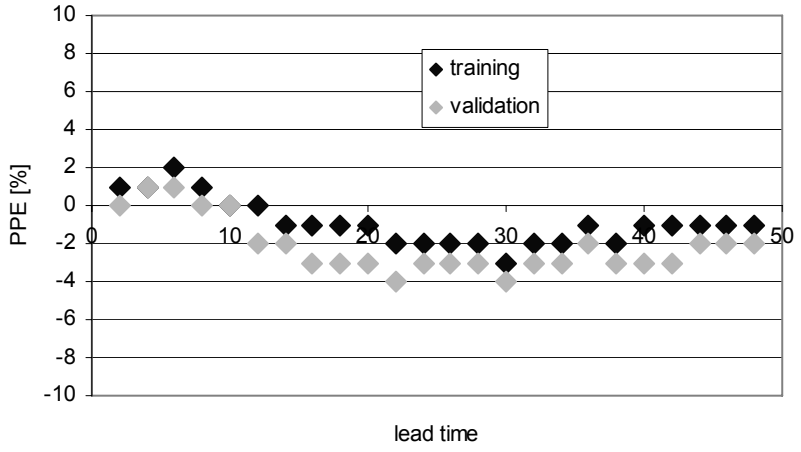


figure 57:Training and validation of the net (peak to peak error).

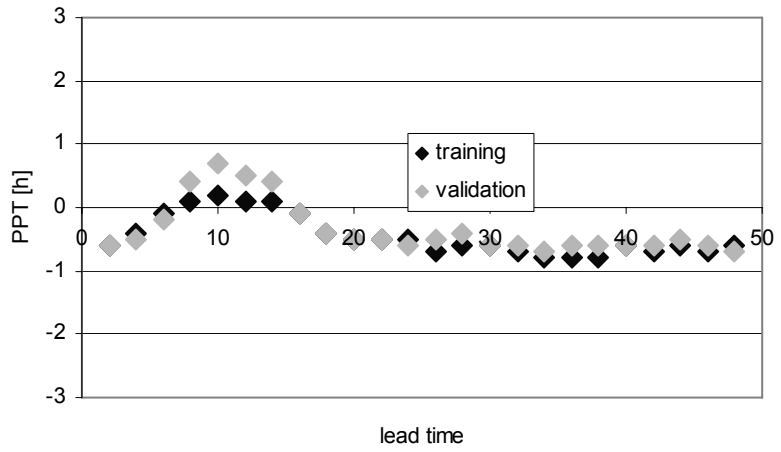


figure 58: Training and validation of the net (peak to peak time).

The résumé of this training and evaluation is: The system is well trained and able to generalise in the context of the database which consists of training and validation events generated with WaSiM. In the next step, an observed extreme flood event is predicted with PAI-OFF in order to put the system to the test in the context of applicability to real data.

5.6 PAI-OFF application: Predicting the 2002 flood event

The system is used to predict the 2002 flood event, which did not feature in the training process. The aim is to predict the peak flow rate with less than 10 % relative error. Focus is set to long lead times (24h, 36h and 48h), because these are the key to a timely warning. Furthermore, the catchment has a travel time of about one day. This makes it desirable to test the performance for lead times ≥ 24 hours. The forecast is separately given for the considered lead times. No updating or other corrective features are used, the forecast is a straight forward single step operation. This is the key to online Monte-Carlo evaluation of model uncertainty. A model which is constantly updated cannot be used for this purpose, because updating procedures usually interfere with internal state variables - such as water content of a storage – and therefore do not consistently predict over a whole forecast period.

In figure 59 the results are given for the 24 hour lead time forecast. The PAI-OFF prediction slightly under-estimates the peak flow, but stays within the 10 % confidence limits for the high discharge periods of the test flood.

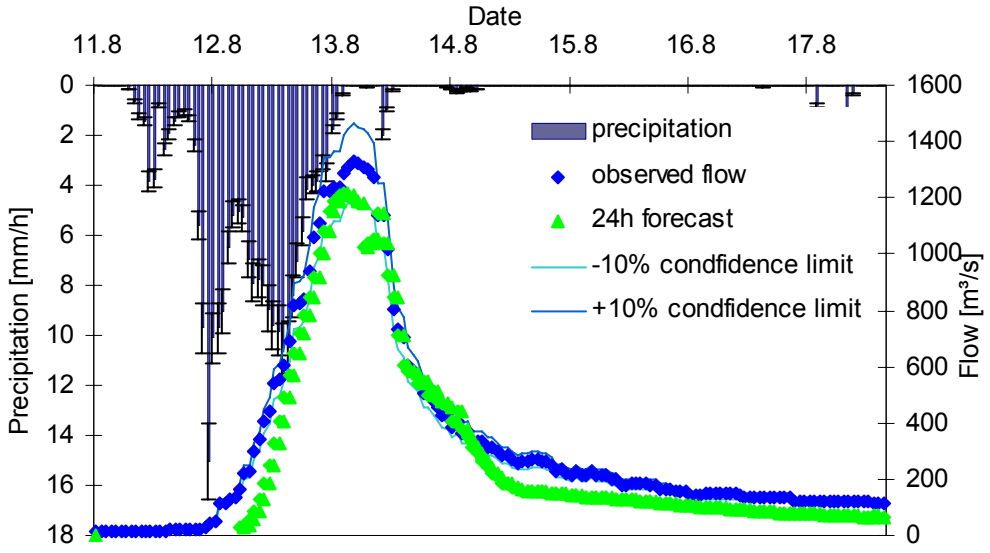


figure 59: Forecast performance of the 2002 flood event at Kriebstein gauging station for 24 hours lead time

This predictive performance holds also for the 36 hour forecast (figure 60). For this lead time, PAI-OFF slightly-over-predicts the peak flow, but again satisfies the 10 % error criterion. It is also notable how well the system is able to portray the recession phase of the event. This must be emphasized in the light of the fact that the test application abstains from updating procedures.

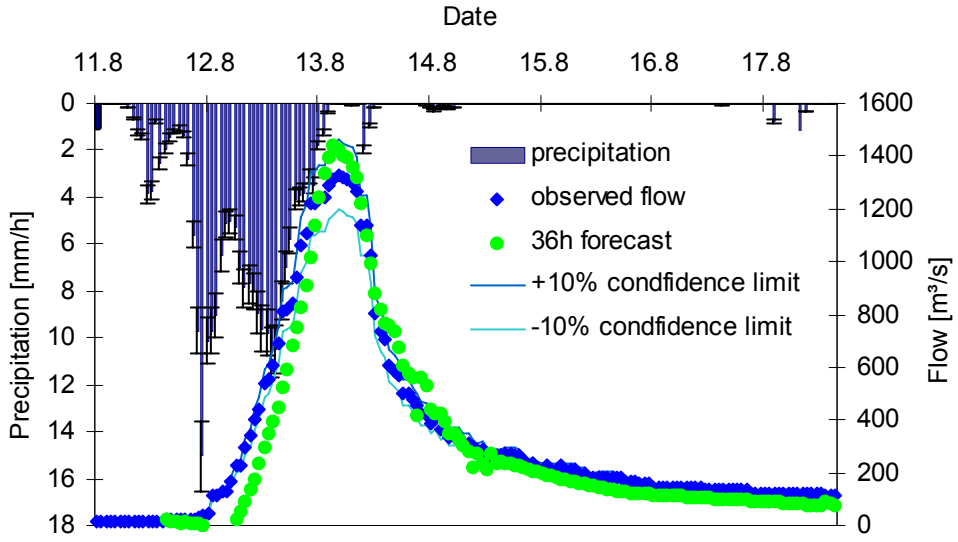


figure 60: Forecast performance of the 2002 flood event at Kriebstein gauging station for 36 hours lead time

Finally, figure 61 gives the picture of the 48 hour lead time prediction. The peak flow relative error for the forecast is only 4.3 %. At first sight it seems astonishingly that the 48 hour forecast outperforms the forecast for shorter lead times. The latter show peak errors close to the 10 % confidence limits. But really, this result confirms the stability and predictive power of the PAI-OFF. approach for increasing lead times.

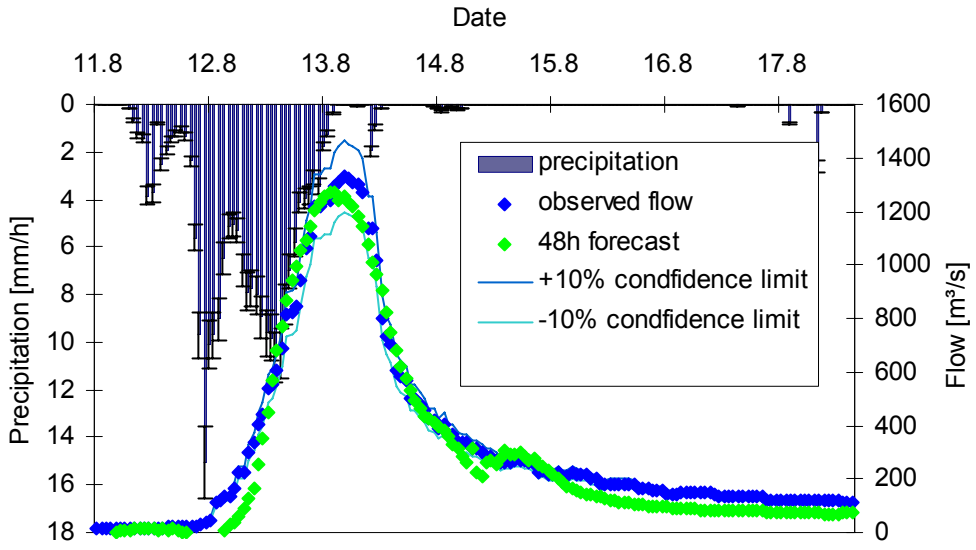


figure 61 Forecast performance of the 2002 flood event at Kriebstein gauging station for 48 hours lead time

Figure 63 allows for another view on the forecast performance for the 48 hour lead time. The four days with highest discharge of the 2002 flood event are separated and then sorted in descending order. It can be seen how well the forecast portrays the high discharges. Only when the discharge falls below 200 m³/s, PAI-OFF under-predicts the observed flow slightly.

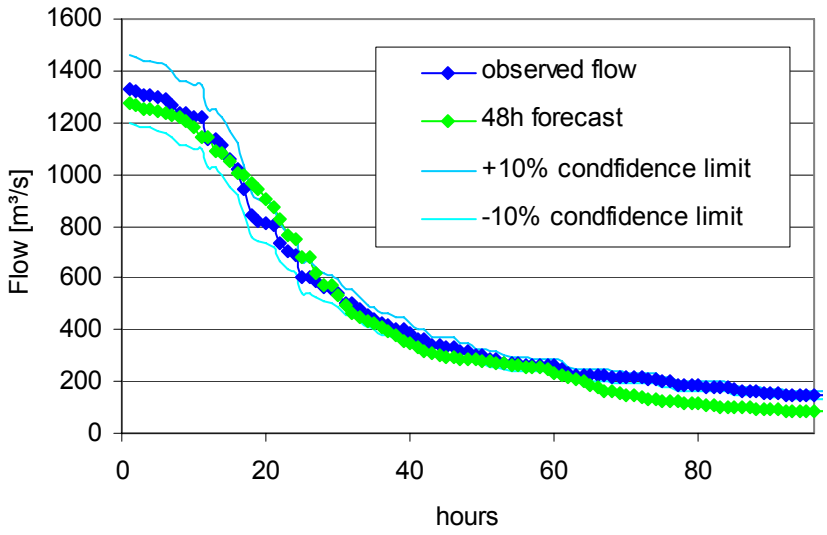


figure 62: 48h forecast sorted by discharge

The fact that the forecast with the shortest lead-time (24h) is not better than the forecast with 48 hours supports the features selected here. They fully describe the governing processes, even for a real, unseen event. In the case of an application with precipitation forecast from e.g. the MM5 climate-model, the longer lead times might show a higher error due to the larger uncertainty in the meteorological forecast. Here the uncertainty arising from precipitation data is the same for all lead times. The differences between observed and predicted flows in the rising and especially the falling limb of the hydrograph are a result of the training strategy, which concentrates on the best possible portrayal of the peak flows. The results have to be seen in the light that there is no updating of whatever kind involved.

Summarizing the convincing forecasts for the 2002 event yields:

- ➔ The 10 % error criterion is satisfied for all lead times;
- ➔ The flood dynamics is very well portrayed over the whole event without model updating;
- ➔ PAI-OFF is a robust tool with excellent predictive power.

6 Résumé

Flash floods are amongst the most dangerous natural hazards. Nevertheless, until now, current research effort did not lead to an approach mastering this phenomenon in the context of online forecasting (section 1.2). This is due to the contradiction that fast conceptual models are likely too simplistic to be used for extrapolation into the range of extreme flood events. Detailed, physically based models are maintaining more fidelity to the process, even in the context of predicting unseen extreme events. Unfortunately they are too slow for online forecasting, especially if the uncertainty of the forecast is to be taken into account by Monte Carlo analysis. For both conceptual and physically based models it is difficult, if not impossible, to find one singular set of parameters describing the full range of flood patterns occurring in a natural catchment.

To address this dilemma, dynamic parameterisation, an alternative way for model calibration is developed in the study. The considered flood events are therefore classified according to the dominating processes mirrored by the peak flow rates. WaSiM-ETH is calibrated for the different event classes separately, thus reflecting the specific process characteristics of the classes. Results of the various model runs are then merged with a sigmoidal weighting approach to form an unequivocal hydrograph. This significantly improves the overall model performance for the peak flow rate. For the exemplary application in the 129 km² catchment of the Schwarze Pockau River at Zöblitz gauge (section 2.5.2), this approach improves the overall performance from 2.6 % to 1.8 % for the peak error. Kriebstein gauge (the catchment is 10 times larger) is modeled in section 5.2.3. Here the same strategy allows for a reducing the mean peak RSME from 36% to 16%. The results for the smaller catchment are generally better than for the larger catchment. This is a consequence of the different data quality for the catchments. The convincing ability to reduce the peak error in both applications allows for a general use of this method in the context of flood modeling. Integrating various parameter sets for all important patterns of flood formation in a watershed becomes thus possible.

Artificial neural networks have been proposed as an alternative to the tedious application of detailed, physically based model, but until now, no efficient strategy has been proposed to overcome the fact that they are restricted to the range of their training data. This training data typically consists of observed time series which do not encompass all the possible patterns of flood formation. This prevents the nets from accurately predicting extreme floods.

PAI-OFF (Process Modelling and Artificial Intelligence for Online Flood Forecasting) is overcoming the above mentioned general restrictions and shortcomings of both hydrological models and neural networks. It is presented in chapter 2 of this work. This new methodology is based on a synthesis of physically based catchment modelling and artificial neural networks in the form of polynomial neural networks. Exploiting the advantages of the two modeling approaches allows for complying with the most important requirements of flash flood forecasting: low computation times, complete robustness and a straightforward operation alongside high predictive reliability. The PAI-OFF set-up is a three step process: First the considered catchment is described by means of a physically based, distributed process model. This model is thoroughly calibrated and validated according to the strategy described in chapter 2. The model is then used to transform all possible meteorological flood prone scenarios into the catchments integral response: the runoff. The meteorological input data, together with the corresponding runoff time series, form are used to build a data base. This data base characterizes the catchment in all flood relevant situations, taking into account the basin internal preconditions. It serves for training a polynomial neural network by means of the characteristic features described in section 4. Once the net is trained it can be easily employed for online flood forecasting. The outstanding computational efficiency of the PoNN (less than half a second per forecast), together with the new approach for parameterising the catchment model described in section 2.5.2 yield two advantages outperforming the currently available models:

- ➔ In PAI-OFF different model parameter sets can be considered in the set-up process. This capability of merging different process representations in one fast single operational system makes PAI-OFF unique, offering a possible way to the future generation of hydrological models. It allows

for specifically considering different types of flood formation patterns in one model.

➔ The immense speed of the forecast allows for online Monte Carlo scenarios. PAI-OFF is thus capable of evaluating the meteorological uncertainty online as the event approaches.

In the last chapter of this work, a test application confirms the predictive power of the PAI-OFF methodology in the 1700 km² Kriebstein test catchment. The system is used to forecast the 2002 flood at Kriebstein gauge for lead times of 24, 36 and 48 hours respectively. The target is to predict the observed flow with a relative peak error of less than 10 %. PAI-OFF impressively meets this requirement for all three lead times, proving its predictive power not only for synthetic data but also for a real extreme flood event which was not part of the training data.

References

Abbott M.B., Bathurst J.C., Cunge J.A., O',Connell P.E and Rasmussen J. (1986). An introduction to the European Hydrological system - Systeme Hydrologique Europeen. SHE 1: History and philosophy of a physically based distribute d modelling system. Journal of Hydrology 87: 45-59.

AG Boden. (2005). Pedological Mapping Instructions. Federal Institute of Geosciences and Resources. Schweizerbartsche Verlagsbuchhandlung.

AMS (American Meteorological Society) (2000). Glossary of Meteorology. Glickman T. S. (ed.), 2nd Edition. Boston: American Meteorological Society: 855 pp.

Anderson E. A. (1973). National Weather Service River Forecast System – Snow Accumulation and Ablation Model. NOAA Tech. Memo. NATIONAL WEATHER SERVICE. U.S. Dep. Of Commerce, Silver Spring.

Badoux A. (1999). Untersuchung zur flächendifferenzierten Modellierung von Abfluss und Schmelze in teilvergletscherten Einzugsgebieten. (Investigation on spatially distributed runoff and snowmelt modelling in partially glaciered catchments) Diploma thesis, Geographical Institute, ETH Zurich.

Battiti R. (1989). Accelerated back-propagation learning: Two optimization methods. Complex Systems (3): 331-342.

Benaman J. (2002). A systematic approach to uncertainty analysis for a distributed watershed model. Ph.D. Thesis, School of Civil and Environmental Engineering, Cornell University, Ithaca, N.Y.

Beven K.J. (1989). Changing ideas in Hydrology- the case of physically based models. Journal of Hydrology 105: 157-172.

Beven K. J.(2005). On the concept of model structural error. Water Sci. Technol. 52(6): 167-175.

Beven K. J. and Kirkby M. J. (1979). A physically based variable contributing area model of basin hydrology, Hydrol. Sci. Bull. 24(1): 43-69.

Beven K. and Binley A. (1992). The Future Of Distributed Models: Model Calibration and Uncertainty Prediction. Hydrol. Process. 6: 279-298.

Beven K. and Freer J. (2001). Equifinality, data assimilation and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249(1-4): 11-29.

Bonafe A., Galeati G. and Sforna M. (1994). Neural networks for daily mean flow forecasting. Hydrol. Engrg., Software V, Computational Mechanics publications, Southampton, U.K. 1: 131-138.

Braun L. (1985): Simulation of Snowmelt-Runoff in Lowland and Lower Alpine Regions of Switzerland. Zürcher Geographische Schriften, Heft 21, Geographisches Institut der ETH Zürich.

Bruen M. and Yang J. (2005). Functional networks in real-time flood forecasting-a novel application. Journal of Hydrology 28: 899-909.

Bük 200: Available at: http://www.bgr.bund.de/nn_524512/DE/Themen/Boden/ Produkte /Karten/BUEK__200.html.

Buslenko N. P. and Schreider J. A. (1964). Doe Monte Calo Methode und ihre Verwirklichung mit elektronischen Digitalrechnern. Teubner Verlag, Leipzig

Butts J. T., Kristensen M. and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. Journal of Hydrology 298: 242-266.

CAN (2005). Flash Flood Forecasting Over Complex Terrain: With an Assessment of the Sulphur Mountain NEXRAD in Southern California, NRC, USA.

Cappus P. (1960). Bassin experimental d'Alrance: Etude des lois de l'ecoulement - Application au calcul et a la prevision des debits. La Houille Blanche A: 493–520.

Castellano-Mendez M. (2004). Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods. Journal of Hydrology 296 38-58.

Christiaens K. (2002). Use of sensitivity and uncertainty measures in distributed hydrological modelling with an application to the MIKE SHE model. Water Resources Research 38(9): 8-1 to 8-15.

Ciarapica L. and Todini E. (1998). TOPKAPI - Un modello afflussi-deflussi applicabile dalla scala di versante alla scala di bacino. XXVI Convegno di Idraulica e Costruzioni Idrauliche.

Ciarapica L and Todini E. (2002). TOPKAPI: a model for the representation of the rainfall-runoff process at different scales. Hydrol. Process. 16(2): 207-229.

Cigizoglu H. K. (2003). Incorporation of ARMA models into flow forecasting by artificial neural networks. Environmetrics, 14(4): 417-427.

CORINE (2000). Available online: www.caf.dlr.de/caf/anwendungen/projekte/projektenutzung/corine/.

Crawford N.H. and Linsley (1966). Digital Simulation in Hydrology: Stanford Watershed Model IV. Stanford Univ., Dept. Civ. Eng. Tech. Rep.39.

Cullmann J, Schmitz G.H., Mishra V. (2006). Eine neue Strategie zur Hochwasservorhersage in schnellreagierenden Einzugsgebieten - Modellierungsaspekte. (A new strategy for flood forecasting in fast reacting watersheds- aspects of modelling) BFG-Veranstaltungen 2006/3, Koblenz.

Dawson C. W. and Wilby R. (1998). An artificial neural network approach to rainfall-runoff modelling. Hydrol. Sci. 43(1): 47-66.

Dawson C. W. and Wilby R. (2001). Hydrological modelling using artificial neural networks. Progress in Physical Geography 25(1): 80-108.

Doswell C. A., Brooks H. E. and Maddox R. A. (1996). Flash Flood Forecasting: An Ingredients-Based Methodology. Weather and Forecasting 11: 560-581.

Duan Q., Sorooshian S. and Gupta V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28(4): 1015-1031.

Dunne T. and Black R. D. (1970). Partial area contributions to storm runoff in a small New England watershed. Water Resources Research 6: 1296-1311.

DWD (1997). Starkniederschlagshöhen für die Bundesrepublik Deutschland (KOSTRA), Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main .

Dyck S. and Peschke G.(1995). Grundlagen der Hydrologie (Foundations of Hydrology) Verl. F. Bauw., Berlin., 387 ff.

Foka A. (1999). Time Series Prediction Using Evolving Polynomial Neural Networks. Dissertation, Uni Manchester.

Freer J.and Ambroise B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research 32(7): 2161-2173.

Georgakakos K. P. (1986). On the design of national, real-time warning systems with capability for site-specific, flash-flood forecasts. Bulletin of the American Meteorological Society, 67: 1233-1239.

Görner W., Cullmann J., Peters R. and Schmitz G. H. (2006). Nutzung künstlicher neuronaler Netze zur Bereitstellung von

Entscheidungsgrundlagen für operative und planerische wasserwirtschaftliche Maßnahmen. Projektbericht TU-Dresden.

Green W. A. and Ampt G. A. (1911). Studies on soil physics I. The flow of air and water through soils. J. of Agr. Sci. 4: 1-24.

Gupta V. K. , Mesa O. and Dawdy D. R. (1994). Multiscaling theory of .food peaks: Regional quantile analysis. Water Resources Research 30(12):3405-3421.

Gupta V. K. and Dawdy D. R. (1995). Physical interpretations of regional variations in the scaling exponents of flood quantiles. Hydrol Process. 9(3/4): 347–61.

Gurtz J., Badoux. A. and Lang H. (1999). Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. Hydrol. Process. 13: 2751-2768.

Gurtz J., Zappa M., Jasper K., Lang H., Verbunt M., Badoux A. and Vitvar T. (2003). A Comparative Study in Modelling Runoff and its Components in Two Mountainous Catchments. Hydrol. Process. 17: 297-311.

Gruntfest E. and Huber C. J (1991). Toward a comprehensive national assessment of flash flooding in the United States. Episodes 14: 26-35.

Gyasi-Agyei Y. (2005). Stochastic disaggregation of daily rainfall into one-hour time scale. Journal of Hydrology 309: 178-190.

Hagan M. T. and Menhaj M. (1994). Training feedforward networks with the Marquardt algorithm. IEEE Trans. on Neural Networks 5: 989-993.

Halff A. H., Halff H. M. and Azmoodeh M. (1993). Predicting runoff from rainfall using neural networks. Proc Engrg. Hydrol., ASCE, New York: 760–765.

Hall T., Brooks H. and Doswell C. D. III. (1999). Precipitation forecasting using a neural network. Weather and Forecasting, 14(3), 338-345.

Hettiarachchi P, Hall M. J. and Minns A. W. (2005). The extrapolation of artificial neural networks for the modelling of rainfall-runoff relationships. J. of Hydroinf. 07.4: 291-296.

Hirschboeck K. K. (1987). Catastrophic flooding and atmospheric circulation anomalies. In: Mayer L. and Nash D. (Eds) Catastrophic Flooding. Boston Allen and Unwin: 23-56.

Hjelmfelt A. T. and Wang M. (1993). Artificial neural networks as unit hydrograph applications. Proc. Engrg. Hydrol., ASCE, New York: 754–759.

Hock R. (1999). Distributed temperature-index ice- and snowmelt model including potential direct solar radiation. J. Glaciol. 45(149): 101-111.

Hornberger G. M. and Spear R. C. (1981). An approach to preliminary analysis of environmental systems. Journal of Environmental Management 12: 7-18.

Hornik K., Stinchcombe M. and White H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks 2(5): 359-366.

Hu T. S., Lam K. C. and Ng S. T. (2001). River flow time series prediction with a range-dependent neural network. Hydrol. Scienc. Journal 46 (5) : 729-745.

Imrie C. E., Durucan S. and Korre A. (2000). River flow prediction using artificial neural networks: generalizing beyond the calibration range. Journal of Hydrology 233: 138-153.

Iman R. L. and Conover W. J. (1982). A distribution free approach to inducing rank correlation among input variables. Communications in Statistics-Simulation and Computation, B 11:311-334.

Kaastra I., and Boyd M. S. (1995). Forecasting futures trading volume using artificial neural networks. The J. of Future Markets 15(8): 953-970.

Kang M. S. and Kang M.G. (2006). Application of Grey model and artificial neural networks to flood forecasting. Journal of the American Water Resources Association 42(2): 473-486

Kelsch M. (2001). Hydrometeorological characteristics of flash floods. In: Gruntfest E. and Handmer J. (Eds.) Coping with Flash Floods. NATO Science Series 2. Environmental Security, Boston Kluwer Academic Publishers Vol. 77: 181-193.

Kim G. and Barros A. P. (2001). Quantitative flood forecasting using multisensor data and neural networks. Journal of Hydrology 246: 45-62.

Kothyari K.C. and Garde R.J. (1991). Annual Runoff Estimation for Catchments in India. J. Water Res. Management and Planning 117(1): 1-10.

Koutsoyiannis D. and Onof C. (2001). Rainfall disaggregation using adjusting procedures on a Poisson cluster model. Journal of Hydrology 246: 109-122.

Kuligowski, R. J. and Barros A. P. (1998). Experiments in Short-term Precipitation Forecasting Using Artificial Neural Networks. Monthly Weather Review 126: 470-482.

Kull D. W. and Feldman A. D. (1998). Evolution of Clark's unit graph method to spatially distributed runoff. J. Hydrologic Engineering ASCE 3(1): 9-19.

Laio F., Porporato A., Revelli R. and Ridolfi L. (2003). A comparison of nonlinear flood forecasting methods. Water Resources Research 39(5): 1129, doi:10.1029/2002.

Lanza L.G.; Ramírez J.A. and Todini E.(2001). Stochastic rainfall interpolation and downscaling. Hydrology and Earth System Sciences 5(2): 139–143.

Liu Z. and. Todini E. (2001). Relationship between TOPMODEL and TOPKAPI and Physics. Mediterranean Storms. Proc. of the 3rd EGS Plinius Conference.

Liu Z. and Todini E. (2002). Towards a comprehensive physically based rainfall-runoff model. Hydrology and Earth System Sciences (6): 859-881.

Luce A, Haag I. and M. Bremicker (2006). Einsatz von Wasserhaushaltsmodellen zur kontinuierlichen Abflussvorhersage in Baden-Würrtemberg. (Daily discharge forecasting with operational water-balance models in Baden-Würrtemberg): Hydrologie und Wasserbewirtschaftung 50(2): 58-67.

Ludwig K, Bremicker M. and Haag I. (2006) Operationeller Vorhersagebetrieb mit dem Wasserhaushaltsmodell LARSIM in Baden-Würrtemberg. BFG Veranstaltungen 3/2006: 15-27.

Ma L. and Khorasani K. (2005). Constructive Feedforward Neural Networks Using Hermite Polynomial Activation Functions. IEEE Trans. on Neural Networks 16(4): 821-823.

Mackay N. G., Chandler R. E., Onof C. and Wheater H. S.(2001). Disaggregation of spatial rainfall fields for hydrological modelling. Hydrology and Earth System Sciences 5(2): 165–173.

Maidment D.R. (1993). Developing a spatially distributed unit hydrograph by using GIS. Proc. HydroHIS'93. IAHS Pub. No. 211.

Maier H. R. and Dandy G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 15: 101-124.

Maren A., Harston C. and Pap R. (1990). Handbook of Neural Computing Applications. Academic Press, San Diego, CA.

Martina.M., Todini E and Libralon A. (2006). A Bayesian decision approach to rainfall thresholds based flood Warning. Hydrol. Earth Syst. Sci. 10: 413–426.

McCuen R. H. (1998). Hydrologic Analysis and Design. 2nd. Ed. Prentice Hall New Jersey 814p.

Meetschen D. and Simmer C. (2006) Einsatz des Bonner Wetterradars als Hochwasserwarninstrument in den Einzugsgebieten von Agger und Erft. BFG Veranstaltungen 3/2006: 70-76.

Merz R. and Blöschl G. (2003). A process typology of regional floods. Water Resources. Research 39(12).

Meyer-Brötz G. and Schürmann J. (1970). Methoden der automatischen Zeichenerkennung (automatic character recognition). Akademie Verlag Berlin: 100-105.

Michaud J. and Sorooshian S. (1994). Comparison of Simple Versus Complex Distributed Runoff Models on a Midsized Semiarid Watershed. Water Resources Research 30(3): 593-605.

Mogil H. M., Monro J. C. and Groper H. S. (1978). The National Weather Services flash flood warning and disaster preparedness programs. Bulletin of the American Meteorological Society 59**:** 690-699.

Mohamed J. L. and Walsh J. (1986): Numerical Algorithms, Oxford Science Publications.

Monteith J. L. (1965). Evaporation and Environment. Symp. Soc. Exp. Biol. 19: 205-234.

National Weather Service. (2004a). Turn Around, Don't Drown. Available online at http://www.srh.noaa.gov/tadd/. Accessed June 19, 2004.

National Weather Service(2004b). National Weather Service Hydrologic Glossary. Available online at http://www.crh.noaa.gov/hsd/hydefa-c.html. Accessed May 14, 2004.

O'Connor J. E. and Costa J. E. (2004). Spatial distribution of the largest rainfall-runoff floods from basins between 2.6 and 26000 km2 in the United States and Puerto Rico. Water Resources Research, 40**:** W01107, doi:10.1029/2003WR002247.

Parisi R.,Claudio D., Orlandi G. and Rao B. D. (1996). A generalised learning paradigm exploiting the structure of neural networks. IEEE Transactions on Neural Networks 7 (6): 1451-1460.

Pegram G. G. S. and Clothier A. N. (2001). Downscaling rainfields in space and time, using the String of Beads model in time series mode. Hydrology and Earth System Sciences 5(2): 175–186.

Peschke G. (1987). Soil Moisture and Runoff Components from a Physically Founded Approach. Acta hydrophysica 31(3/4): 191-205.

Polger P. D., Goldsmith B. S., Przywarty R. C. and Bocchieri J. R. (1994). National Weather Service warning performance based on the WSR-88D. Bulletin of the American Meteorological Society 75: 203-214.

Ponce V.M. (1989). Engineering Hydrology. Prentice-Hall, New Jersey. 640p.

Post D.A. and Jakeman A.J. (1996). Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. Hydrol. Process. 10(6): 877-892.

Prudhomme C. J., D. and Svensson C. (2003). Uncertainty and climate change impact on the flood regime of small UK catchments. Journal of Hydrology 277: 1-23

Rabuffetti D. and Barbero S. (2005). Operational hydro-meteorological warning and real-time flood forecasting: the Piemonte Region case study. Hydrology and Earth System Sciences 9(4): 457-466

Rojas R. (1996). Neural Networks: A systematic Introduction. Springer-Verlag, Berlin.

Saghafian B. and Julien P.Y. (1995). Time to equilibrium for spatially variable watersheds. Journal of Hydrology 172: 231-245.

Sajikumar N. and Thandaveswara B. S. (1996). Artificial neural network for estimating annual runoff. In: Aladama A. A and Aparicio J. (Eds.) Computation in water resour., XI Vol. II, Computation Mechanics Publication, Southampton, England: 509-516.

Sajikumar N. and Thandaveswara B.S. (1999). A non linear rainfall-runoff model using artificial neural networks. Journal of Hydrology 216: 32-55.

Schmitz G. H., Cullmann J., Peters R., Görner W., Lennartz F (2006). PAI-OFF: A New Way to Online Flood Forecasting in Flash Flood Prone Catchments. (In review at Water Resources Research).

Schulla J.(1997). Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen, Diss., ETH Zürich, CH, 161p.

Schulla J. and Jasper K.(2001). Model Description WaSiM. Internal report, IAC, ETH Zürich, 166p.

Shamseldin A. Y. (1997). Application of neural network technique to rainfall-runoff modelling. Journal of Hydrology 199: 272-294.

Shin Y. and Gosh J. (1992). Approximation of Multivariate Functions Using Ridge Polynomial Networks. Proc. International Joint Conference on Neural Networks, Vol. 11: 380-385.

Sivakumar B., Sorooshian S., Gupta., Hoshin V.and Gao X. (2001). A chaotic approach to rainfall disaggregation. Water Resources Research 37(1): 61-72.

Skahill B. E. and Doherty J. (2006). Efficient accomodation of local minima in watershed model calibration. Journal of Hydrology (online).

Smith J., and Eli R. N. (1995). Neural network models of rainfall-runoff process. J. Water Res. Management and Planning, ASCE 121(6): 499-508.

Spear R.C., Hornberger G.M (1980). Eutrophication in Peel Inlet, II: Identification of critical uncertainties via generalised sensitivity analysis. Water Research 14: 43-49.

Tate E. and Cauwenberghs K. (2005). An innovative flood forecasting system for the Demer basin: A case study. Intl. J. River Basin Management 3(4): 1-5.

Todini E. (1996). The ARNO rainfall-runoff model. Journal of Hydrology 175: 339-382.

Tokar A. S. and Johnson P.A. (1999). Rainfall- runoff modelling using artificial neural networks. J. Hydrol. Engrg., ASCE 4(3): 232-239.

Tokar A. S. and Markus M. (2000). Precipitation-runoff modelling using artificial neural networks and conceptual models. J. Hydrol. Engrg., ASCE 5(2): 156-161.

Toth E., Brath A. and Montanari A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology 239: 132-147.

USGS (2005). http://ks.water.usgs.gov/Kansas/waterwatch/flood/definition.html

Wagener T., McIntyre N., Lees M.J., Wheater H.S., Gupta H.V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrol. Process. 17: 455-476.

White H. (1989). Learning in artificial neural networks: A statistical perpective. Neural Computation I: 425-464.

Wriedt G. and Rode M. (2006).Investigation of parameter uncertainty and identifiability of the hydrological model WaSiM-ETH. Adv. Geosci. 9: 145–150.

Zealand C. M., Burn D. H. and Simonovic S. P. (1999). Short term stream-flow forecasting using artificial neural networks. Journal of Hydrology 214: 32-48.

Zhu M., Fujita M., and Hashimoto N. (1994). Application of neural networks to runoff prediction. In: Hipel K. W. et al (Eds) Stochastic and statistical method in hydrology and environmental engineering, Vol. 3, Kluwer, Dordrecht, The Netherlands: 205–216.

Abbreviations:

AMS	American Meteorological Society
ANN	Artificial neural network
ANOVA	Analysis of variance
ASCE	American Society of Civil Engineers
CPU	Central processing unit
DEM	Digital elevation model
DYNIA	Dynamic identifiability analysis
HRF	Hydrologic response feature
GML	Gauß-Marquardt-Levenberg
IoA	Index of agreement
LM	Levenberg-Marquardt
LMS	Least mean square
MCS	Monte Carlo-Simulations
MLFN	Multi layer feed forward neural network
MSE	Mean square error
MSL	Meters above mean sea level.
NSE	Nash-Sutcliffe efficiency
PAI-OFF	Process modelling and artificial intelligence for online flood forecasting
PEST	Parameter Estimation
PoNN	Polynomial neural network
RMSE	Rooted mean square error
SAE	Sum of absolute errors
SCE	Shuffled complex evolution
SCEUA	Shuffled complex evolution
SF	State feature
AHR	Area of similar hydrologic response

Symbols:

a	*Exponent for permutation*
b	*Exponent for permutation*
B	*Bias*
c	*Exponent for permutation*
d_k	*Length of period k without precipitation [h]*
dr	*WaSiM parameter scaling the interflow*
Duration	*Criterion for describing rainfall events [h]*
e	*A matrix of network errors*
f (.)	*Transfer function*
$F(t_0)$	*Feature at reference time* t_0
Form	*Criterion for describing rainfall events [h]*
g	*Degree of the polynomial applied*
$G(x,y,t)$	*Global radiation as a function of space (x,y) and time (t)*
H	*The Hessian matrix*
J	*The Jacobian matrix*
K0	*Reference input parameter value of simple sensitivity analysis*
Kd	*Direct runoff parameter in WaSiM [h]*
Ki	*Interflow parameter in WaSiM [h]*
krec	*Parameter in WaSiM controlling the gradient of hydraulic conductivity*
l	*Actual Lag [h]*
L	*Maximum lag, defining the interval* $\{t_0, t_0 - L\}$ *[h]*
$M(x,y,t)$	*Relative air humidity as a function of space (x, y) and time (t)*
MPI	*Criterion for describing rainfall events [mm/h]*
Mv(x)	*Mean Sensitivity for period x*
n	*Number*

N	*Number of product vectors*
O	*Optimized vector matrix*
p	*Product vector*
Po	*Input parameter of simple sensitivity analysis*
P	*Vector matrix of net training*
$P(x,y,t)$	*Precipitation as a function of space (x, y) and time (t)*
$Peak$	*Criterion for describing rainfall events [mm/h]*
PF	*Criterion for describing rainfall events [mm]*
$P_G(t)$	*Mean precipitation for entire watershed as a function of time (t)*
$P_O(t)$	*Mean precipitation for orographic zone as a function of time (t)*
$P_L(t)$	*Mean precipitation for the area of similar hydrologic response as a function of time (t)*
Q	*Runoff [mm or m³/s]*
Qe	*Extreme model output [mm or m³/s]*
Qn	*Normal model output [mm or m³/s]*
$Q(t)$	*Flow at the considered gauge as a function of time (t)*
Si	*Sensitivity index* of simple sensitivity analysis
St	*Smoothed function*
$T(x,y,t)$	*Temperature as a function of space (x, y) and time (t)*
TI	*Criterion for describing rainfall events [mm]*
TL	*Criterion for describing catchment retention [h]*
$thick$	*Parameter in WaSiM*
$V(t)$	*Vegetation activity as a function of time (t)*
V	*Output variable of simple sensitivity analysis*
$Volume$	*Criterion for describing rainfall events [mm]*
W	*Weight*
We	*Weight for extreme parameters*

Wji	*MLFN weights*
Wn	*Weight for normal parameters*
Yj	*The output of jth node of a MLFN*
X	*Regressor*
Xi	*The input signal to ith node of a MLFN*
X norm	*The normalized variable*
X min	*The minimum value of the related variable*
X max	*The maximum value of the related variable*
ΔR	*Scatter minimisation*
Λ	*Sigmoid parameter (steepness)*
Θ	*Threshold value*
σ	*Confidence level for the Man-Withney-U test*

LIST OF FIGURES

LIST OF TABLES

APPENDIX

APPENDIX 1

table 25: Events used in the calibration study.

Event	Year	Duration [h]	Peak [mm/h]	Volume [mm]	Form [h]	TI [mm]	TL [h]	PF [mm]	MPI [mm]
1	1981	38	6.34	55.78	0.11	2.68	4.00	43.05	1.47
2	1982	9	10.70	16.02	0.67	0.47	1.00	69.25	1.78
3	1982	6	16.57	37.99	0.44	0.12	3.00	29.46	6.33
4	1984	14	6.08	46.39	0.13	0.66	10.00	22.38	3.31
5	1984	16	13.07	53.78	0.24	0.15	6.00	31.15	3.36
6	1985	14	3.46	25.83	0.13	1.45	1.00	82.41	1.84
7	1985	21	6.57	64.67	0.10	1.07	5.00	29.92	3.08
8	1986	10	7.62	36.55	0.21	0.92	3.00	76.98	3.65
9	1986	5	24.34	41.54	0.59	0.08	1.00	18.47	8.31
10	1987	11	6.40	27.59	0.23	0.63	2.00	59.64	2.51
11	1987	8	22.19	32.16	0.69	0.18	2.00	24.40	4.02
12	1989	6	11.33	12.70	0.89	0.44	1.00	27.36	2.12
13	1990	8	8.29	40.69	0.20	0.72	2.00	26.74	5.09
14	1991	18	6.95	56.91	0.12	1.44	3.00	41.00	3.16
15	1991	18	5.46	35.65	0.15	0.73	3.00	73.80	1.98
16	1993	2	15.60	16.83	0.93	0.13	2.00	40.51	8.42
17	1993	13	23.28	84.41	0.28	0.22	4.00	17.68	6.49
18	1993	24	6.68	62.29	0.11	1.65	7.00	41.01	2.60
19	1994	27	11.38	93.47	0.12	0.44	8.00	20.72	3.46
20	1994	8	2.25	32.96	0.07	2.22	3.00	68.30	4.12
21	1994	10	8.05	30.74	0.26	0.50	1.00	56.06	3.07
22	1994	9	32.75	73.56	0.45	0.09	4.00	15.02	8.17

23	1994	14	4.33	17.61	0.25	1.62	1.00	97.44	1.26
24	1994	5	14.66	28.26	0.52	0.07	1.00	47.21	5.65
25	1994	2	11.47	13.75	0.83	0.09	1.00	64.63	6.88
26	1994	10	12.79	34.69	0.37	0.31	1.00	12.75	3.47
27	1994	30	7.48	45.13	0.17	2.01	2.00	42.92	1.50
28	1994	12	7.31	31.80	0.23	0.82	3.00	41.28	2.65
29	1994	8	7.52	25.76	0.29	0.67	2.00	32.65	3.22
30	1994	10	7.04	33.55	0.21	0.57	4.00	45.95	3.36
31	1994	23	6.81	48.68	0.14	0.73	1.00	11.78	2.12
32	1994	28	3.96	43.87	0.09	6.07	2.00	39.57	1.57
33	1995	14	13.31	41.68	0.32	0.45	1.00	31.96	2.98
34	1995	26	6.94	38.74	0.18	1.87	2.00	85.70	1.49
35	1996	23	10.35	65.27	0.16	0.68	1.00	33.00	2.84
36	1999	11	13.24	55.21	0.24	0.45	1.00	47.68	5.02

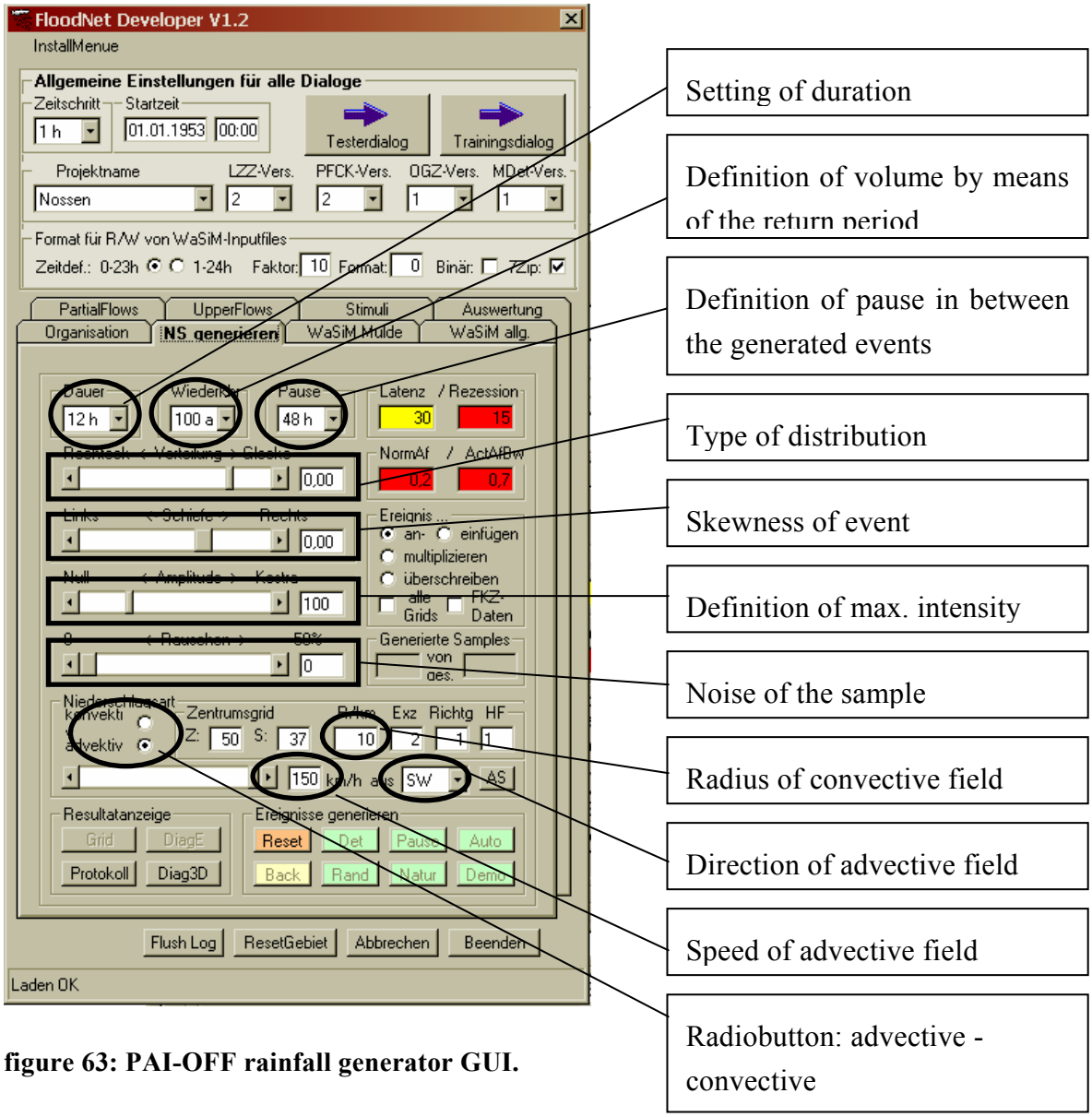
APPENDIX 2

The rainfall generator

The rainfall generator is a development of Görner *et al* (2006). It is grid based and able to generate rainfall fields of any size. The generator works on an hourly time step. The principal functionality can be best described on the basis of figure 63. The settings are described at the right hand side of the figure. First the overall duration is specified. Then the volume for this duration is set by means of the return period. This information is the result of the meteorological catchment analysis. A break can separate the events that are generated, according to the setting chosen. Then parameters of the distributions follow: type, skewness, maximum value and noise.

The generator is able to produce both advective and convective rainfall fields. This is set with a simple radio button.

Rainfall events are generated according to the following principles resulting from a detailed analysis of the meteorological characteristics of historical flood events:



Setting of duration

Definition of volume by means of the return period

Definition of pause in between the generated events

Type of distribution

Skewness of event

Definition of max. intensity

Noise of the sample

Radius of convective field

Direction of advective field

Speed of advective field

Radiobutton: advective - convective

figure 63: PAI-OFF rainfall generator GUI.

1) The rainfall distribution in the basin is dependent on the height above mean sea level. Therefore the generated precipitation P(x,y,t) is a function of height.

$$P(x,y,t)=[1+hf_2 * hf_1(x,y)] * P_K(x,y,t) \quad \text{(equation 50)}$$

with: $P_K(x,y,t)$ = local precipitation from DWD (1997)

hf_2 =height factor for specific events with: $(0 \le hf_2 \le 2)$

$hf_1(x,y)$ =local height differentiation according to:

$$hf_1(x,y)=[H(x,y)-500]/1000 \quad \textbf{(equation 51)}$$

where: $H(x,y)$ =mean height [m] of cell (x,y)

2) Convective storms are generated in the form of ellipses. The axes of these rainfall fields are defined by:

direction: 0: Main axis is aligned east-west

1,57: Main axis is aligned south-north

eccentricity: direction = 0 yields:

eccentricity < 1:main axis in south-north direction

eccentricity = 1:circle

eccentricity > 1:main axis in east-west direction

radius:= ½ * diameter of the secondary axis

A single time step of a convective rainstorm is exemplarily shown in figure 64.

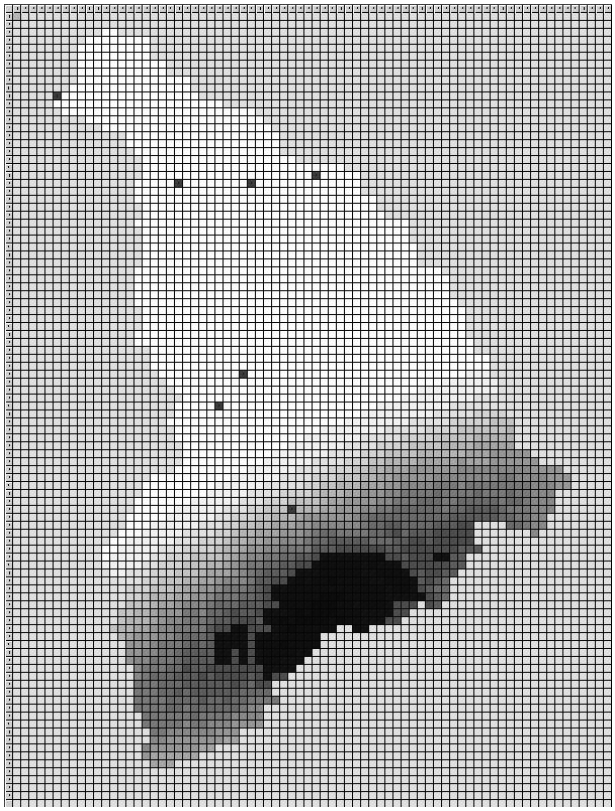


figure 64: Convective rainstorm over the southern part of the considered catchment.

APPENDIX 3

The features used in the test application:

In the following denotation of the polynomial *G*, the degree of state features and hydrologic response features is separated by a "+". The following combinations are permuted with equation 17.

G 1+0 $P_{10n} = S_i \,|_{i=1}^{I}$

G 0+3 $P_{01n} = \Delta F_j \,|_{j=1}^{J}$

$\cup \quad P_{02n} = \Delta F_j^2 \,|_{j=1}^{J}$

$\cup \quad P_{03n} = \Delta F_j^3 \,|_{j=1}^{J}$

G 1+3 $P_{11n} = S_i \Delta F_j \,|_{i=0}^{I} |_{j=0}^{J}$

$\cup \quad P_{12n} = S_i \Delta F_j^2 \,|_{i=1}^{I} |_{j=1}^{J}$

$\cup \quad P_{13n} = S_i \Delta F_j^3 \,|_{i=1}^{I} |_{j=1}^{J}$.

G 2+3 $P_{21n} = S_i S_k \Delta F_j \,|_{i=0}^{I} |_{k=i}^{I} |_{j=0}^{J}$

$\cup \quad P_{22n} = S_i S_k \Delta F_j^2 \,|_{i=1}^{I} |_{k=i}^{I} |_{j=1}^{J}$

$\cup \quad P_{23n} = S_i S_k \Delta F_j^3 \,|_{i=1}^{I} |_{k=i}^{I} |_{j=1}^{J}$

G 3+3 $P_{31n} = S_i S_k S_m \Delta F_j \,|_{i=0}^{I} |_{k=i}^{I} |_{m=k}^{I} |_{j=0}^{J}$

$\cup \quad P_{32n} = S_i S_k S_m \Delta F_j^2 \,|_{i=1}^{I} |_{k=i}^{I} |_{m=k}^{I} |_{j=1}^{J}$

$\cup \quad P_{33n} = S_i S_k S_m \Delta F_j^3 \,|_{i=1}^{I} |_{k=i}^{I} |_{m=k}^{I} |_{j=1}^{J}$

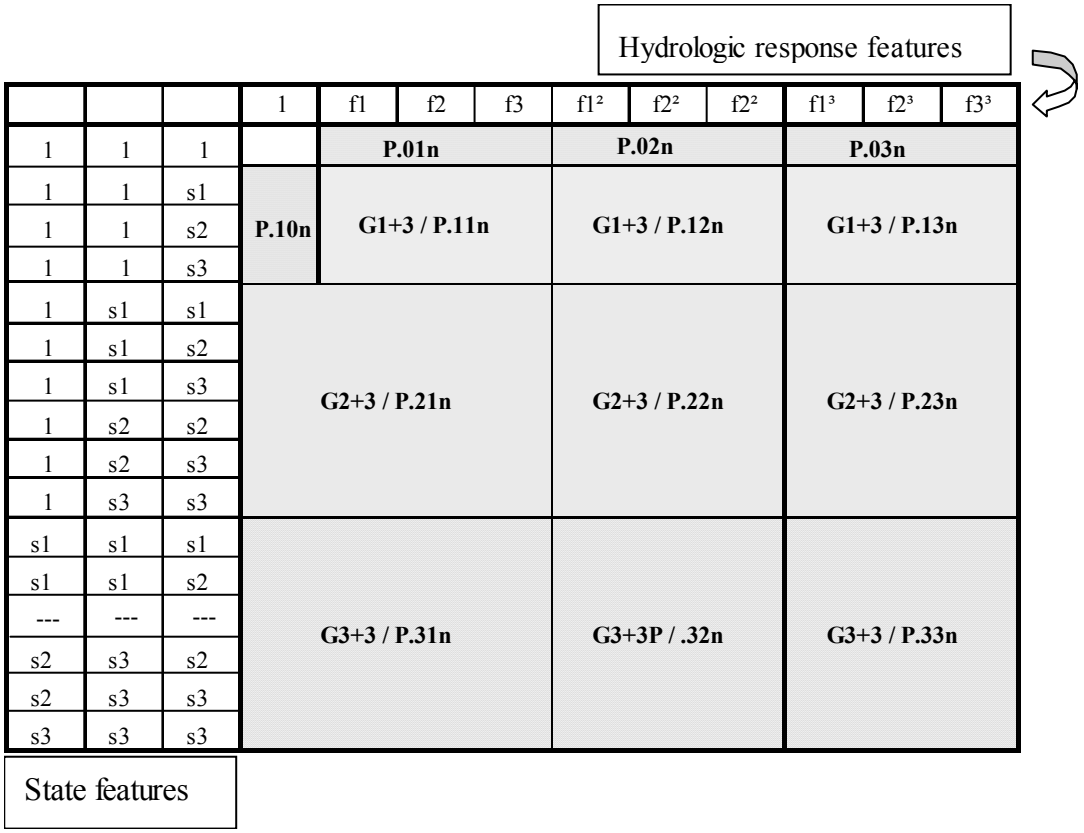


figure 65: Tableau of the possible perturbation of features according to equation14.

APPENDIX 4

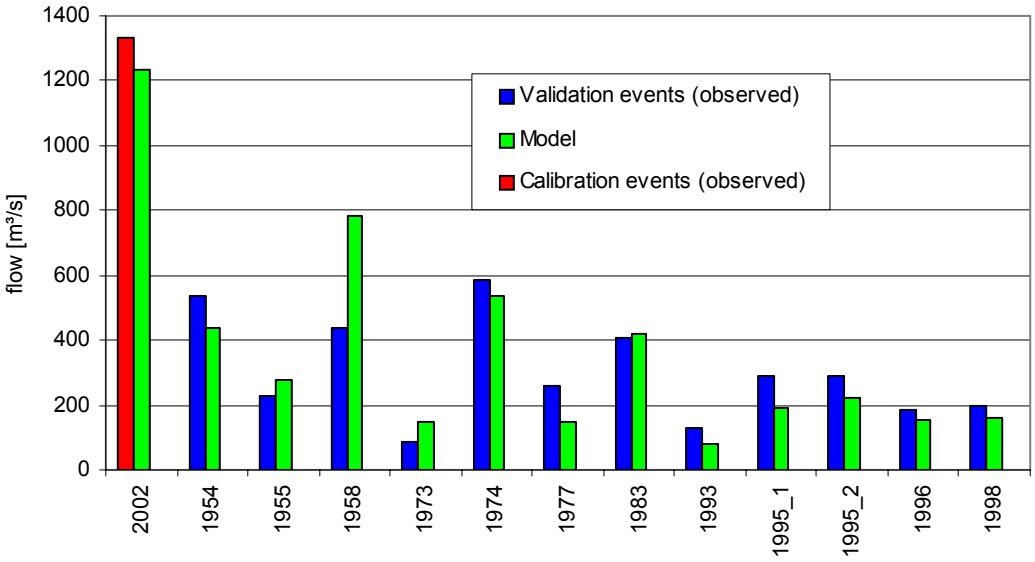


figure 66: Calibration and validation performance of the singular parameter set for large events.