Space Weather

RESEARCH ARTICLE
10.1029/2018SW002141

Nowcasting and Predicting the *Kp* Index Using Historical Values and Real-Time Observations

Yuri Y. Shprits[1,2,3], Ruggero Vasile[1], and Irina S. Zhelavskaya[1,2]

[1]GFZ German Research Centre for Geosciences, Potsdam, Germany, [2]Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany, [3]Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA, USA

Key Points:
- Short-term predictions should be made with solar wind data and long-term predictions with recurrence forecast
- The accuracy of predictions is much lower during disturbed geomagnetic conditions
- Rebalancing allows us to improve the accuracy of predictions during disturbed geomagnetic conditions

Supporting Information:
- Supporting Information S1
- Figure S1
- Figure S2
- Figure S3

Correspondence to:
Y. Shprits,
yshprits@gfz-potsdam.de

Citation:
Shprits, Y. Y., Vasile, R., & Zhelavskaya, I. S. (2019). Nowcasting and predicting the *Kp* index using historical values and real-time observations. *Space Weather*, *17*, 1219–1229. https://doi.org/10.1029/2018SW002141

Received 9 DEC 2018
Accepted 3 JUL 2019
Accepted article online 16 JUL 2019
Published online 5 AUG 2019

Abstract Current algorithms for the real-time prediction of the *Kp* index use a combination of models empirically driven by solar wind measurements at the L1 Lagrange point and historical values of the index. In this study, we explore the limitations of this approach, examining the forecast for short and long lead times using measurements at L1 and *Kp* time series as input to artificial neural networks. We explore the relative efficiency of the solar wind-based predictions, predictions based on recurrence, and predictions based on persistence. Our modeling results show that for short-term forecasts of approximately half a day, the addition of the historical values of *Kp* to the measured solar wind values provides a barely noticeable improvement. For a longer-term forecast of more than 2 days, predictions can be made using recurrence only, while solar wind measurements provide very little improvement for a forecast with long horizon times. We also examine predictions for disturbed and quiet geomagnetic activity conditions. Our results show that the paucity of historical measurements of the solar wind for high *Kp* results in a lower accuracy of predictions during disturbed conditions. Rebalancing of input data can help tailor the predictions for more disturbed conditions.

1. Introduction

Machine learning (ML) tools have become widely used in recent decades and are now extensively applied for a number of applications in the industry and academia. ML is a complex term that describes various methods and tools that allow computer systems to improve performance or "learn" from data and based on this "learning" process create systems targeted at various problems. The most common problems for ML are classification, clustering, and regression. Classification is a problem of ML when the output is restricted to a specific set of values. An example of classification may be the identification of different subsets of data or discrimination of photos according to the type of objects being photographed. Clustering is referred to as the problem where only input data are provided, and there are no preset or desired outputs, but the data need to be divided into a number of different subsets with common properties.

Regression is a process of estimating the relationship between input and output variables. The simplest method of regression is a linear regression but a number of nonlinear methods also exist and need to be used to describe complex dependences. In this study, we focus on the regression problem.

Regression algorithms often create nonlinear models based on training data that allow to predict the desired variables—such as in this study the *Kp* index—from the input data, such as in our case the historical solar wind data and historical observations of *Kp*. In this study, we focus on relating historical inputs with the current value of *Kp* and values of *Kp* in the future. The lead time for the prediction is usually referred to as "horizon", and the input variables as "features".

A number of ML methods have been recently applied for the prediction of the geomagnetic index *Kp*. The *Kp* index (Bartels, 1949; Mayaud, 1980) is one of the most commonly used measures of geomagnetic activity. Its long and continuous record makes this index most useful for climatological studies and for the comparison with historical values.

The analysis of the global geomagnetic disturbances for each individual station is complicated by the fact that variations during storms or substorms depend on the magnetic local time (MLT) and also show seasonal variability. To normalize the measurements, the first step in inferring local *K* indices is to subtract a quiet daily curve variation in the northward (*H*) and eastward (*D*) components of the magnetic field. There are a

number of standard practices that are used to retrieve such a curve (see, e.g., Mayaud, 1980). Each observatory can choose their own method to infer a quiet curve, but most of the official *Kp* observatories now use automated tools. The most popular method is the one from FMI in Finland. A detailed review of how *Kp* is derived is also provided in Takahashi et al. (2001).

To calculate *Kp*, first deviations from quiet daily curve are normalized by assigning them to one of the preset semilogarithmic bins to obtain a local *K* index. The 28 bins are chosen to run as (0, 0+, 1−, 1, 1+, … , 9−, 9). The average value of all local *K* values for all *Kp* stations is referred to as *Kp*. A detailed review of the derivation of the *Kp* index is given in Rostoker (1972).

The Space Weather Prediction Center of the National Oceanic and Atmospheric Administration (NOAA) and the U.K. Met Office use data products similar to *Kp* for operations, to define strong geomagnetic activity and to issue warnings to power grid operators. Satellite operators also often refer to it in order to identify anomalies that are caused by internal or surface charging (Hastings & Garrett, 2004).

The *Kp* index is also used for a wide range of scientific applications. For instance, in the ionospheric community, it is employed for parameterizing ionospheric ion outflow (Welling et al., 2015; Yau et al., 2011) and auroral particle precipitation (Emery et al., 2008). In magnetospheric physics, *Kp* is well correlated with a number of parameters such as cold-plasma density in the plasmasphere (Goldstein et al., 2014; Maynard & Chen, 1975; Pierrard et al., 2009), hot-plasma particle density (Denton et al., 2016; Korth et al., 1999), and the location of the plasmapause (Carpenter & Anderson, 1992). Some empirical models of the Earth's external magnetic field use *Kp* as their only input (Tsyganenko, 1989). *Kp* is also widely used to parameterize very low frequency (VLF) waves (Agapitov et al., 2015; Orlova et al., 2016; 2014; Shprits et al., 2007; Spasojevic et al., 2015) and ultra low frequency (ULF) wave amplitudes (Brautigam & Albert, 2000; Ozeke et al., 2014). For a detailed discussion on the use of *Kp* for identifying various phenomena in space, see Borovsky and Shprits (2017).

The availability of continuous observations of the solar wind from the upstream Lagrange point (L1) allowed for the development of tools for real-time nowcasting of the *Kp* index. Costello (1997) and Boberg et al. (2000) developed data-driven models based on artificial neural networks or simply neural networks that are now becoming increasingly popular for a variety of scientific and commercial applications (Bishop, 2006; Goodfellow et al., 2016). Wing et al. (2005) presented and tested different models that utilized both L1 solar wind and nowcasting of *Kp*, tested them for 1- and 4-hr predictions, and compared them to previously used methods, including neural networks and Nonlinear AutoRegressive Moving Average model with eXogenous inputs NARMAX (Balikhin et al., 2001; Boaghe et al., 2001). They also showed how the performance of the method can be improved by using the near-real-time values of the *Kp* index derived from near-real-time ground observations. The application of various ML tools has been the subject of a number of recent studies (Ayala Solares et al., 2016; Bala & Reiff, 2012; Ji et al., 2013; Tan et al., 2017; Wang et al., 2015; Wintoft et al., 2017).

While previous studies show that solar wind-driven models can provide accurate short-term predictions that can be further improved by using recent measurements and recurrence in *Kp* and solar wind, how much each of these methods can contribute to forecast accuracy and how prediction accuracy depends on the horizon time remain relatively unexplored. Understanding of the efficiency and limitations of the predictions based on different types of inputs can help develop independent algorithms for predictions. Combinations of such models, depending on the availability of data, may be used in an optimal way to produce an even more accurate forecast. As the probability distribution function is skewed toward smaller values, we also explore in this study how forecasts can be tailored to predict large values of *Kp* that are usually underrepresented in the available data sets.

The paper is organized as follows: In section 2, we describe the different data sources and processing procedures used to build the input data sets for our *Kp* forecast. In section 3, we describe the ML model used, the metric employed for evaluation, and the validation schemes. In section 4, we show the results of our study, that is, validated average errors for various forecast horizons, solar cycle dependence errors, and present methods to correct for a skewed distribution of the probability distribution function and tailor predictions to high *Kp* values.

2. Model Inputs

In this section, we describe the inputs and their combinations used to perform the short- and long-term *Kp* forecast as well as our data processing framework. These inputs are built out of two main sources of data:

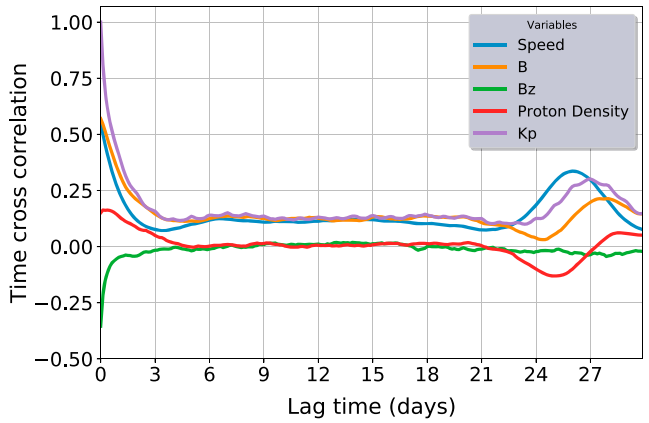

Figure 1. Linear cross correlation between *Kp* and 1-hr averaged solar wind variables and the interplanetary magnetic field (IMF), as well as the *Kp* time autocorrelation. The correlations decay fast in the first few hours, showing the lack of predictability of *Kp* by those variables for a large time horizon. A peak appears, though, at around 25-26-27 days, and it is caused by recurrent effects due to the solar rotation. The peak has its maximum at slightly different delay times for different variables.

the first source is the historical, definitive *Kp* index, provided by the World Data Center (http://wdc.org.ua/) and the German Research Centre for Geosciences (GFZ) (ftp://ftp.gfz-potsdam.de/pub/home/obs/kp-ap/wdc/). The second source of input is 1-min high-resolution propagated solar wind data from the OMNI website (ftp://spdf.gsfc.nasa.gov/pub/data/omni/high_res_omni/). Propagation of the solar wind data allows to predict the values of the solar wind at the bow shock with a lead time that is equal to the propagation time of the solar wind from L1 to the bow shock. For this study, we consider the period between 1 January 2000 and 31 December 2017, spanning almost two solar cycles, for which continuous solar wind observations are available. The physical variables used are the solar wind speed (V), the proton density (n), the vertical component of the magnetic field (B_z) in the GSM frame of reference, and the total magnetic field amplitude (B).

Due to the high cadence of the input solar wind variables, following Wintoft et al. (2017) and Tan et al. (2017), we extract statistical information out of these variables before using them as inputs. Specifically, we consider 3-hr windows of measurements for which we calculate average values as well as minimum and maximum values. We denote the average value of the variable x taken during the time interval $[-3, 0]$ by $\langle x_{[-3;0]} \rangle$. In a similar way, we define minimum and maximum values of a variable x as $\min\{x_{[-3;0]}\}$ and $\max\{x_{[-3;0]}\}$.

To study the dependence of *Kp* on the solar wind parameters and previous *Kp* values, we first consider the linear time-lagged correlation between *Kp* and different input variables. Figure 1 shows the time lag linear correlation between *Kp* and 1-hr averaged solar wind variables, as well as the *Kp* autocorrelation for a time lag of up to 30 days.

The correlation decreases very fast with increasing prediction lead time. A relatively big increase in the correlations occurs for horizons of approximately one solar rotation (i.e., 27 days). This correlation is particularly high during the declining phase of the solar cycle which is dominated by the recurrent activity due to high solar wind streams. Coronal mass ejection (CME)-induced storms cannot be accurately predicted from the consideration of the recurrence alone and may not significantly contribute to the correlation with previous solar cycle parameters.

Below, we describe various inputs for the prediction of *Kp* that will be considered in this study. We henceforth refer to these input variables as "data sets." We first consider two data sets that will be used as baseline predictions to estimate the relative efficiency of the predictions.

Persistence. In order to compare the performance of several model inputs, we introduce here the first baseline model, that is, the *Kp* Persistence model. This model assumes that the *Kp* forecast is given by the last available value of *Kp*

$$Kp(t_0 + h) = Kp(t_0 - 3) \tag{1}$$

where t_0 is the beginning of an interval of the *Kp* bin and h is the horizon time. For real-time operations, if $Kp(t_0 - 3)$ is not known, it must be substituted either with the previous value or with a nowcasted value obtained with another prediction system.

Kp Average. The second baseline model is the *Kp* Average model. This model estimates *Kp* at a later time by an average value of *Kp* over a long-term time interval for which the index is available, from 1932 until now

$$Kp(t_0 + h) = \langle Kp \rangle \tag{2}$$

SC Average. To account for the solar cycle dependence, we introduce a third baseline model, that is, the Solar Cycle Average (SC Average) model. Such a model could be used by an operator when model inputs, such as *Kp* time series and solar wind, are not available or reliable. This model is constructed by taking the *Kp* data set from 1932 to 2004 and calculating the average *Kp* for each day of the solar cycle counted from the epoch time, that is, the maximum of the sunspot number. Static models may also be used for planning, such as an

estimation of the orbital decay of future or ongoing missions. We do not perform any normalization of the solar cycle as the length of the solar cycles cannot be inferred ahead of time.

Solar Wind. This data set uses past solar wind measurements close to the current time t_0. The structure of the functional dependence that we will determine can be presented as

$$Kp(t_0+h) = f\left(\max\left(x^{(i)}_{[t_0-3(n+1);t_0-3*n]}\right), \min\left(x^{(i)}_{[t_0-3(n+1);t_0-3*n]}\right), \left\langle x^{(i)}_{[t_0-3(n+1);t_0-3*n]}\right\rangle \ldots\right) \quad (3)$$

where $n = 0, 1, 2$, $Kp(t_0+h)$ refers to the value of Kp during the time interval $[t_0, t_0+h]$ and the $x^{(i)}$ refer to different solar wind variables, that is, proton density, total speed, magnetic field amplitude, and magnetic field z component, taken within fixed 3-hr intervals prior to the current moment. In total, we therefore have 36 input features, 4 variables, 3 statistical functions (minimum, maximum, and average), and 3 time windows.

Kp Historical. This data set uses three values of Kp data close to the current time to forecast Kp at a later time.

$$Kp(t_0+h) = f\left(Kp(t_0-3), Kp(t_0-6), Kp(t_0-9)\right) \quad (4)$$

We exclude the dependence on $Kp(t_0)$, because in the context of real-time operations, this value may not be available or not accurate enough at the forecast time. Kp is a 3-hr index and in general cannot be calculated for the current 3-hr time interval, as the entire evolution of the fluctuation of the magnetic field is not available until the end of the current 3-hr interval. For this model we only consider three time intervals prior to the current time. The addition of earlier time periods did not result in a significant improvement of the forecast model.

Recurrence. This data set uses Kp data measured at approximately one and two solar rotations earlier in time with respect to the forecast time. We consider Kp during the 3 days one and two solar rotations prior to the current time. With this model we test to what extent the Kp index can be predicted using the information about Kp itself one and two solar rotations before. We use in total 54 input features defined as follows:

$$K_i = s(Kp_{[t_c-3*d, t_c+3*d]}) \quad (5)$$

where s is a statistical function (minimum, maximum, and average) taken over a window of time of different amplitudes ($d = 1, 2, 4$, i.e., 6, 12 and 24 hrs), while $t_c = 26, 27, 28, 53, 54, 55$ days. Then Kp is given by $f(K_i)$, where $i = 1, \ldots, 54$ is the index of the feature used. It should be noted that similar results can be obtained by using the recurrence of the solar wind and taking inputs for the previous two solar rotations. In this study, the Recurrence model uses Kp only and does not use solar wind inputs. Our sensitivity simulations showed that the addition of the solar wind recurrence to the Kp recurrence does not significantly improve the results of the modeling.

Full. The last data set that we consider is composed of all input variables that were used in the previous dynamic models, that is, solar wind and Kp data close to the current time and recurrence of solar wind and Kp data up to two solar rotations in the past. This model blends together the information contained in all the inputs and represents the most accurate forecast that can be achieved with these variables for a given empirical method. The operational disadvantage of such a model is that it can only be used when all of the inputs are available, which is unfortunately not always the case for real-time operations.

3. Methodology: ML and Model Validation

The processed data sets described in the previous section can now be used as input for data-driven models to provide an estimate for the Kp forecast at later times. Each data set is first normalized such that each input feature has values within the $[0, 1]$ interval. Normalization constants are saved to be used in the subsequent testing/operational phases.

In this study, we use feed-forward neural networks with one hidden layer. The addition of more layers does not improve the forecast but only leads to a longer training time. The number of neurons in the hidden layer is the only hyperparameter of the network that we explored systematically. This number, fixed to 20, was selected based on grid search hyper-parameter exploration, showing that a higher or lower number of neurons does not significantly change the results. Sigmoid functions have been used as activation functions for hidden layer neurons, while the final layer has only one neuron, which outputs the Kp forecast using

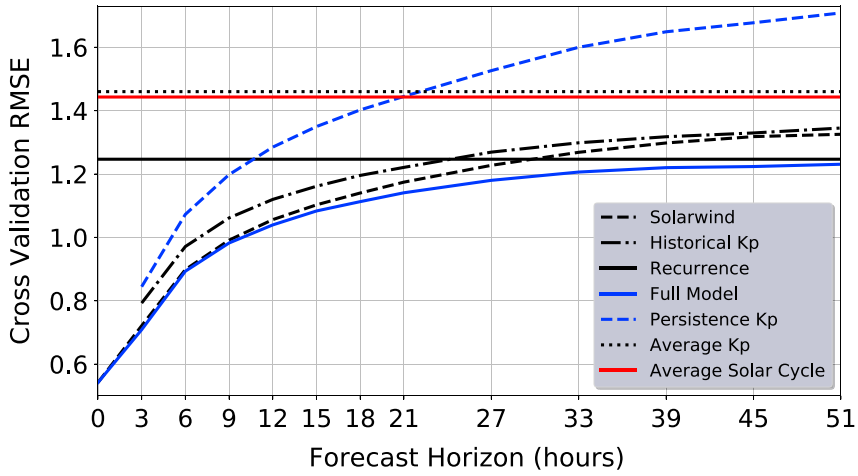

Figure 2. Cross-validation errors for all considered models as a function of the forecast horizon measured in hours. The solar wind-based model performs better for short-term forecasts, while the recurrence model dominates for long time horizons. The full model nicely interpolates between these two regimes, thus performing better than the other models in both cases. RMSE = root-mean-square error.

a linear activation function. The training phase is performed using the second-order Levenberg-Marquardt algorithm available from the ML Matlab Toolbox (https://de.mathworks.com/solutions/machine-learning.html), since it provides faster convergence and slightly better results than first-order methods. We use default Matlab parameters as a criterion for stopping the training. We have tested the sensitivity to the assumed maximum number of epochs and maximum validation failures, and the results are relatively insensitive as the minimum error is reached with only a few training cycles. To compare the performance of the models based on different inputs, we use the root-mean-square error (RMSE). The RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{N}} \tag{6}$$

where the sum is performed over the samples of a validation test set. The RMSE represents the square mean average of the residuals, where the residual is the difference between the predicted and observed values.

We use data from the period between the years 2000 and 2017. To avoid creating a bias for a particular phase of the solar cycle, the following validation scheme is used: First, we divide the data set into 11 yearly chunks selected between 2005 and 2016. We then perform a k-fold cross validation with $k = 11$. Each iteration of the cross validation utilizes one selected year for validation, and the rest of the data are used for training. The whole procedure, repeated 11 times, provides 11 training errors and 11 validation errors.

The forecast model performance is defined by the average of the 11 validation errors obtained as described above. This cross-validation RMSE is interpreted as the average performance error of the system for a given input data set and given forecast horizon. Moreover, it is built to be unbiased with respect to the solar cycle, since validation is performed on data from 11 consecutive years, the approximate length of a solar activity cycle.

4. Results

Figure 2 shows the RMSE error for the *Kp* forecast using the data sets described above. An increase in the lead time of prediction decreases the accuracy of all models except for the Recurrence model and the Average model, which remain constant with forecast horizon time. Figure 2 shows the accuracy of the forecast as a function of the prediction lead time. For the purpose of nowcasting, predictions driven by the solar wind provide the highest accuracy with an RMSE of ∼0.55 for the 0h horizon and ∼0.7 for the 3h horizon.

The short-term solar wind-based prediction is much more accurate than the forecast using the recurrence and persistence models or employing recent values of *Kp*. Surprisingly enough, the model based on the

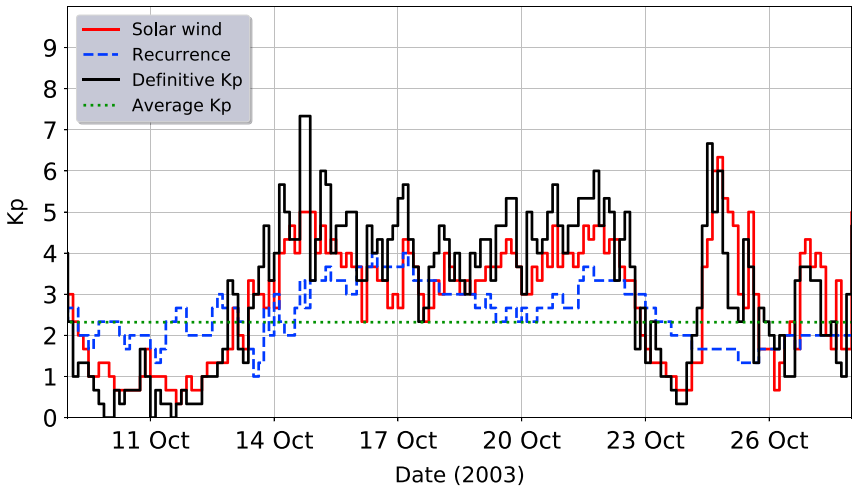

Figure 3. Example of a *Kp* forecast for a 3-hr horizon for the Solar Wind model (solid red line), the Recurrence model (dashed blue line), and the Average model (dotted green line). The definitive *Kp* is shown by the solid black line. RMSE = root-mean-square error.

combination of these methods leads to only a modest improvement of the results over the model driven only by the solar wind. The most likely explanation for such close performance of the models is that the prediction is already very accurate. The accuracy of the solar wind-based model significantly degrades with increasing forecast horizon and becomes comparable to the recurrence forecast for an approximately 1-day horizon time. For longer horizon times, recurrence provides a more accurate forecast than the forecast based on the solar wind and the forecast based on the average *Kp* over the entire data set of *Kp*. Interestingly enough, the SC Average model gives very similar results to a simple Average model. Most likely, the similarity between these two models is due to the fact that different solar cycles might have very different characteristics.

An example of a *Kp* forecast based on the Solar Wind, Recurrence, and Average data sets is shown in Figure 3. This example clearly illustrates that the solar wind-based model is capable of rather accurately reconstructing the values of the measured definitive *Kp* index. While the recurrence-based model is significantly less

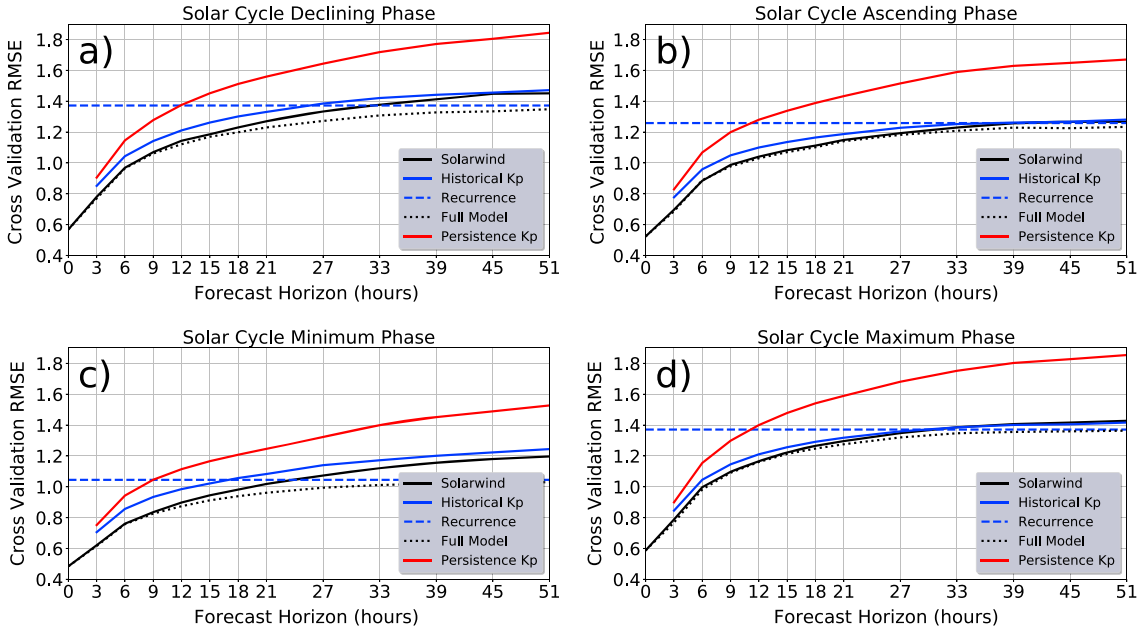

Figure 4. Performance of models using different input data sets, averaged during different phases of the solar cycle: (a) the declining phase 2002–2005 and 2017, (b) the ascending phase 2011–2013, (c) the solar minimum 2006–2010, and (d) solar maxima 2000–2001 and 2014–2016. RMSE = root-mean-square error.

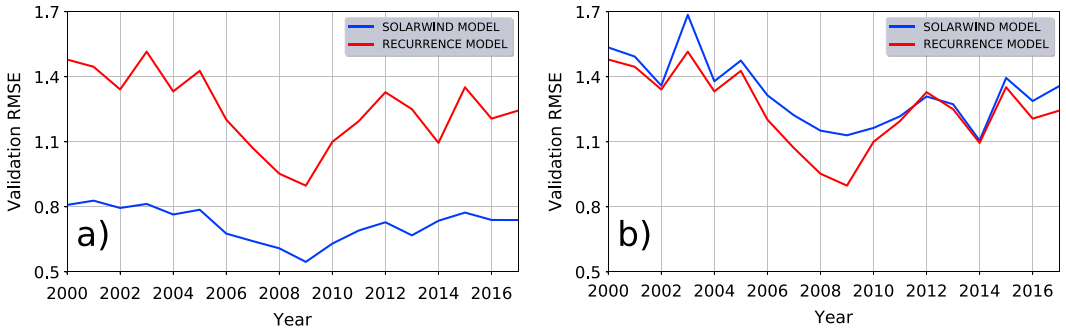

Figure 5. Variability of yearly validation errors for the Recurrence model (red line) and the Solar Wind model (blue line), shown for (a) horizon 0 and (b) horizon 48 hr. RMSE = root-mean-square error.

accurate than the solar wind-based model, it provides significant improvement over a simple model that continuously predicts a constant value of *Kp*.

4.1. Solar Cycle Dependence

In this section, we study the accuracy of the forecast as a function of the solar cycle phase. As currently continuous solar wind observations are provided only for a period of less than two solar cycles, we do not have sufficient data to check how the forecast accuracy changes from one solar cycle to another.

Figure 4 shows the results of the predictions for four different phases of the solar cycle. The highest accuracy is achieved during the solar minimum, while all the other three phases of the solar cycle show higher but rather similar values of RMSE. The small values of the RMSE during the minimum are related to small values of *Kp* observed during that part of the solar cycle. The fact that the declining phase does not have a lower RMSE for the Recurrence model is due to the fact that the declining phase has the highest average values of *Kp*, which leads to higher errors. The normalized RMSE (See supporting information Figure S2) shows lower values of RMSE for the Recurrence prediction during the declining and solar maximum phases of the solar cycle due to higher average values of *Kp*.

To compare the short- and long-term predictions during different phases of the solar cycle, Figure 5 shows the dependence of cross-validation errors for the Recurrence and Solar Wind models, during the time period between 2000 and 2017. The Solar Wind prediction clearly outperforms the Recurrence model for the zero horizon but performs worse than the Recurrence for the 2-day lead time prediction. The solar cycle dependence of the RMSE is also clear for these two models and given lead times.

4.2. *Kp* Distribution Rebalancing

Due to the skewness of the *Kp* distribution, low *Kp* values corresponding to quiet times significantly outnumber the high values of *Kp* during disturbed conditions. Depending on the application, stakeholders and scientists may be interested in forecasting the occurrence of only high values of *Kp*. Predictions of disturbed conditions may be the most important for predicting events which pose a threat of Geomagnetically Induced Currents (GIC) events, since warnings of GIC events are issued only for the strongest geomagnetic

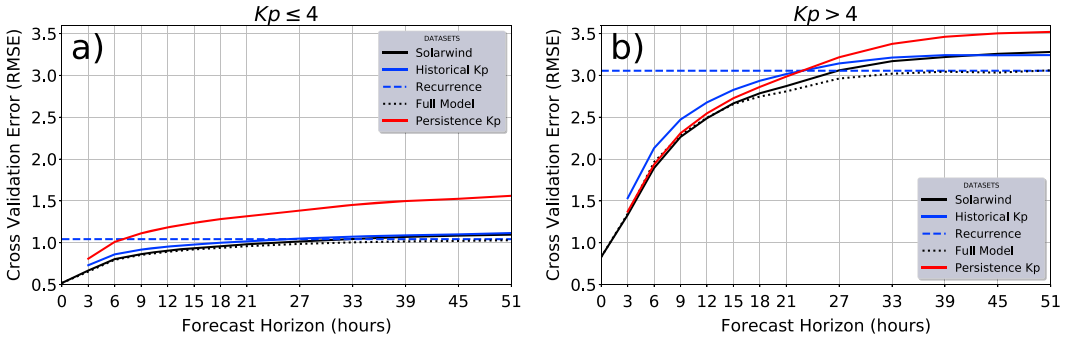

Figure 6. (a) Cross-validation RMSE for all available models as a function of forecast horizon for quiet and moderate conditions ($Kp \leq 4$) and (b) for disturbed conditions $Kp > 4$. RMSE = root-mean-square error.

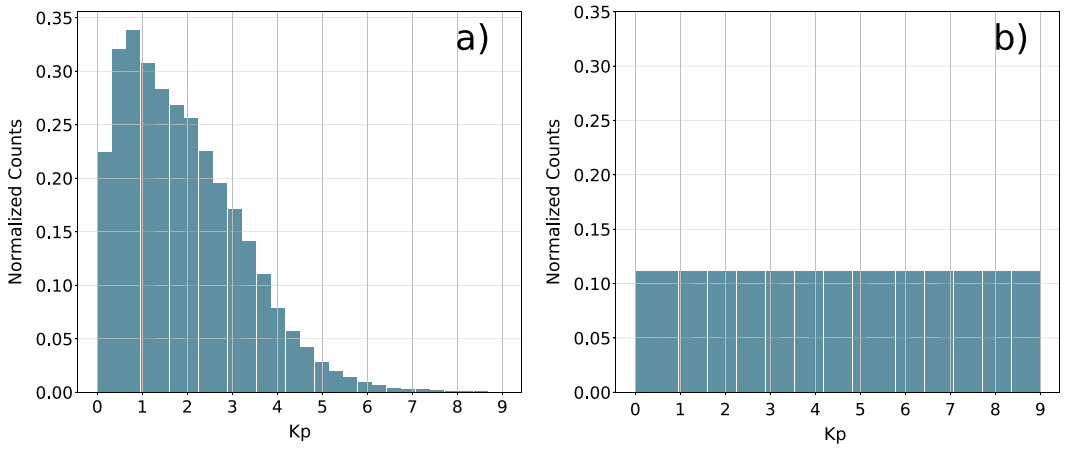

Figure 7. (a) Normalized *Kp* distribution in the period 2000–2017. (b) Effect of rebalancing over the *Kp* distribution.

conditions. For modeling of the ring current and radiation belts, high *Kp* values will play the most important role as most of the dangerous intensifications of the space environment conditions will occur during storms and substorms. Disturbed geomagnetic conditions may also play a dominant role for the orbital decay of satellites.

Figure 6 has a similar format to Figure 2, but the validation is performed for quiet ($Kp \leq 4$) and disturbed ($Kp > 4$) geomagnetic conditions. Data imbalance affects the performance, with high values of *Kp* being affected by much larger errors and in some cases leading to unreliable predictions.

A number of different attempts have been proposed previously to mitigate the imbalance in the *Kp* probability distribution function. For instance, Tan et al. (2017) built two different models for high and low activities, and a third model was used to choose which of the models to employ for forecasts. Another possible approach is to use a target-dependent cost function during the training of the neural network or any other ML algorithm. According to this method, larger errors are assigned to errors committed on high *Kp* estimates with respect to the ones made for low *Kp* values.

In this study, we use Random Oversampling to balance the *Kp* distribution. Randomly chosen high *Kp* values are added to the data set until the distribution becomes independent of *Kp*. The procedure is illustrated in Figure 7. For the skewed *Kp* distribution, we randomly add values for all bins, except for the most probable *Kp* bin. For each *Kp* bin resampling continues until the number of values in each bin becomes equal to the number of points in the bin with the highest occurrence in the original distribution.

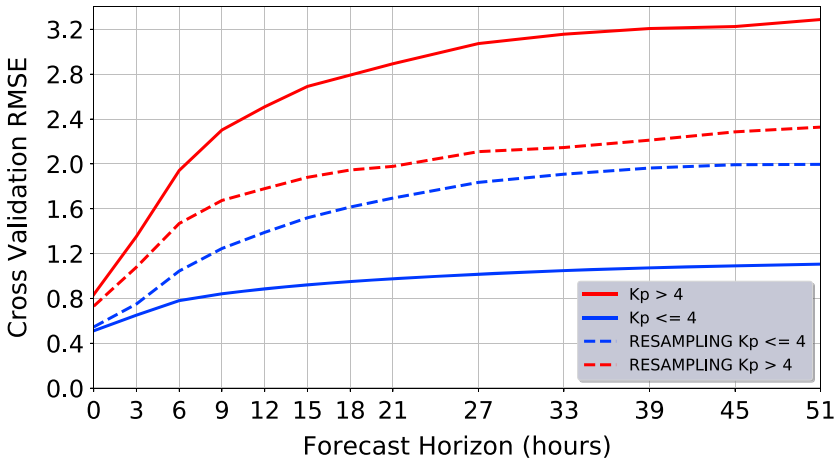

Figure 8. Comparison between Solar Wind models trained with balanced (dashed lines) and unbalanced (solid lines) data sets. The red lines show the performance for the disturbed condition data set ($Kp > 4$), while the blue lines illustrate this for quiet times ($Kp \leq 4$). RMSE = root-mean-square error.

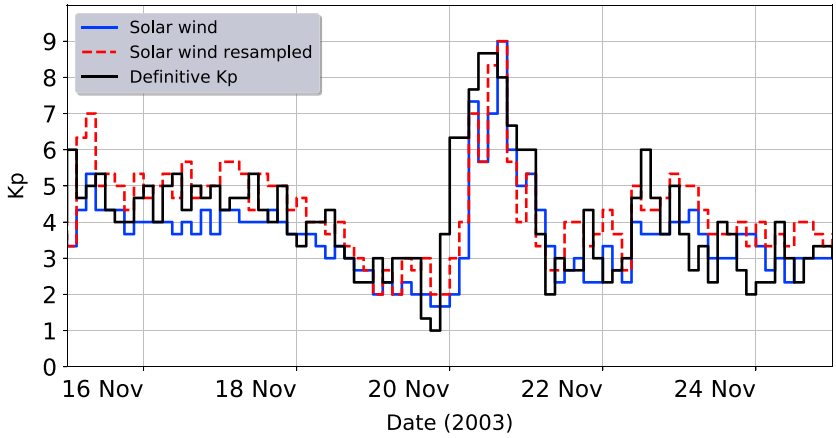

Figure 9. Comparison between the forecast results for a time horizon of 6 hr using the SOLAR WIND (solid blue line) and oversampled data (dashed red line). The definitive *Kp* index is provided as well (solid black line).

Figure 8 shows the results of the Resampled data sets for the Solar Wind model. We present predictions with the solar wind model using balanced data by the dashed lines and original data input by the solid lines. Balancing significantly increases the accuracy of predictions during disturbed geomagnetic conditions as the balanced network had a significantly higher number of high *Kp* points in the training data set. However, the rebalanced model performs worse on average as quiet geomagnetic conditions occur more often than disturbed ones, and the rebalanced network is more tailored for disturbed conditions. Figure 9 shows an example of the predictions of a storm in November 2003 using the original and rebalanced data sets. During disturbed conditions, the predictions made with the original data set systematically underestimate the definitive *Kp* values (see also supporting information), while the rebalanced predictions do not show such a strong bias. We have also tested a more sophisticated resampling method (Synthetic Minority Over-sampling Technique (Chawla et al., 2002), not shown in this manuscript), which gave a similar result.

5. Conclusions and Discussion

In this study, we perform a systematic analysis of the *Kp* prediction using empirical models for different types of inputs and prediction horizons. In particular, we study the performance of models based on recurrence, persistence, historical solar wind observations, and recent observations of *Kp*, as well as measurements of the solar wind. The results show that rather accurate predictions with RMSE less than 1 can be achieved only for horizon times of 6 to 20 hr depending on the phase of the solar cycle. The best performing model for short-term predictions is based on the solar wind parameters, and the addition of other input data provides only a very small improvement of the results.

For longer horizons of over approximately 2 days, recurrence, which is independent of horizon time, provides the best predictions. The accuracy is still rather low, but such a model is much more accurate than predictions based on an average value of *Kp* or solar wind conditions.

The predominance of low values of *Kp* in the probability distribution function presents another complication for the predictions of the periods of disturbed activity. Forecasts based on the original *Kp* data set systematically underestimate *Kp* during disturbed conditions, and methods more tailored for high *Kp* can be used for the forecasts that require the prediction of *Kp* above a certain level. We show that resampling can significantly improve the model results for disturbed conditions while only moderately decreasing the performance for quiet conditions.

Our results demonstrate that while predictions based on the in situ measurements of the solar wind and previously recorded values of *Kp* can provide reasonably accurate values for geomagnetic conditions, longer-term predictions require new input information or new approaches. Longer-term deterministic predictions such as 2- or 3-day predictions cannot be provided by models based on real-time input from only solar wind and *Kp*. Further improvements in long-term modeling should include global physics-based or empirical modeling driven by observations of the Sun, which can give longer-term predictions. Such

long-term predictions from global models of the heliosphere can be used to augment real-time data sets obtained from the L1 point or from ground-based observations. Another approach that can be used for the prediction of events above certain thresholds is to provide not just deterministic predictions but also a prediction of the probability of such events. Such an approach provides longer horizon predictions and can help stakeholders to estimate probabilities and risks and should therefore be the subject of future studies.

Acknowledgments
We acknowledge support by the European Unions Horizon 2020 research and innovation program under Grant agreement 776287 SWAMI, Geo.X and SFB 1294. We also thank Claudia Stolle, Juergen Matzka, Simon Wing, Peter Wintoft, David Jackson, and Sean Bruinsma for helpful discussions. We would like to also thank Dominika Boneberg for proofreading the manuscript. The authors used historical, definitive *Kp* index data provided by the World Data Center (http://wdc.org.ua/) and the German Research Centre for Geosciences (GFZ; ftp://ftp.gfz-potsdam.de/pub/home/obs/kp-ap/wdc/). The solar wind data are taken from the OMNI website (ftp://spdf.gsfc.nasa.gov/pub/data/omni/high_res_omni/).

References

Agapitov, O. V., Artemyev, A. V., Mourenas, D., Mozer, F. S., & Krasnoselskikh, V. (2015). Empirical model of lower band chorus wave distribution in the outer radiation belt. *Journal of Geophysical Research: Space Physics*, *120*, 10,425–10,442. https://doi.org/10.1002/2015JA021829

Ayala Solares, J. R., Wei, H. L., Boynton, R. J., Walker, S. N., & Billings, S. A. (2016). Modeling and prediction of global magnetic disturbance in near-Earth space: A case study for Kp index using NARX models. *Space Weather*, *14*, 899–916. https://doi.org/10.1002/2016SW001463

Bala, R., & Reiff, P. (2012). Improvements in short-term forecasting of geomagnetic activity. *Space Weather*, *10*, S06001. https://doi.org/10.1029/2012SW000779

Balikhin, M. A., Boaghe, O. M., Billings, S. A., & Alleyne, H. St. C. K. (2001). Terrestrial magnetosphere as a nonlinear resonator. *Geophysical Research Letters*, *28*, 1123–1126.

Bartels, J. (1949). The standardized index, Ks, and the planetary index, Kp. *IATME Bulletin*, *12b*, 97.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Berlin, Heidelberg: Springer-Verlag.

Boaghe, O. M., Balikhin, M. A., Billings, S. A., & Alleyne, H. (2001). Identification of nonlinear processes in the magnetospheric dynamics and forecasting of Dst index. *Geophysical Research Letters*, *106*, 30,047–30,066.

Boberg, F., Wintoft, P., & Lundstedt, H. (2000). Real time Kp predictions from solar wind data using neural networks. *Physics and Chemistry of the Earth*, *25*, 274–280.

Borovsky, J. E., & Shprits, Y. (2017). Is the Dst index sufficient to define all geospace storms? *Journal of Geophysical Research: Space Physics*, *122*, 11,543–11,547. https://doi.org/10.1002/2017JA024679

Brautigam, D. H., & Albert, J. M. (2000). Radial diffusion analysis of outer radiation belt electrons during the October 9, 1990, magnetic storm. *Journal of Geophysical Research*, *105*(A1), 291–309. https://doi.org/10.1029/1999JA900344

Carpenter, D. L., & Anderson, R. R. (1992). An ISEE/whistler model of equatorial electron density in the magnetosphere. *Journal of Geophysical Research*, *97*, 1097–1108.

Chawla, N. V., Bowyerk, K. W., Hall, L. O., & Kegelmeyer, P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*(2002), 321–357.

Costello, K. A. (1997). Moving the rice MSFM into a real-time forecast mode using solar wind driven forecast models (Ph.D. Dissertation), Rice Univ., Houston, Texas.

Denton, M. H., Henderson, M. G., Jordanova, V. K., Thomsen, M. F., Borovsky, J. E., Woodroffe, J., et al. (2016). An improved model of electron and ion fluxes at geosynchronous orbit based on upstream solar wind conditions. *Space Weather*, *14*, 511–523. https://doi.org/10.1002/2016SW001409

Emery, B. A., Coumans, V., Evans, D. S., Germany, G. A., Greer, M. S., Holeman, E., et al. (2008). Seasonal Kp, solar wind, and solar flux variations in long-term single-pass satellite estimates of electron and ion auroral hemispheric power. *Journal of Geophysical Research*, *113*, A06311. https://doi.org/10.1029/2007JA012866

Goldstein, J., De Pascuale, S., Kletzing, C., Kurth, W., Genestreti, K. J., Skoug, R. M., et al. (2014). Simulation of Van Allen Probes plasmapause encounters. *Journal of Geophysical Research: Space Physics*, *119*, 7464–7484. https://doi.org/10.1002/2014JA020252

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*.. Cambridge Massachussetts, London, England: the MIT Press.

Hastings, D., & Garrett, H. (2004). *Spacecraft-environment interactions*. Cambridge: Cambridge University Press.

Ji, E. Y., Moon, Y. J., Park, J., Lee, J. Y., & Lee, D. H. (2013). Comparison of neural network and support vector machine methods for *Kp* forecasting. *Journal of Geophysical Research: Space Physics*, *118*, 5109–5117. https://doi.org/10.1002/jgra.50500

Korth, H., Thomsen, M. F., Borovsky, J. E., & McComas, D. J. (1999). Plasma sheet access to geosynchronous orbit. *Journal of Geophysical Research*, *104*(A11), 25,047–25,061. https://doi.org/10.1029/1999JA900292

Mayaud, P. N. (1980). *Derivation, meaning, and use of geomagnetic indices*, Geophysical Monograph Series (pp. 22). Washington DC: American Geophysical Union. https://doi.org/10.1029/GM022

Maynard, N., & Chen, A. (1975). Isolated cold plasma regions: Observations and their relation to possible production mechanisms. *Journal of Geophysical Research*, *80*(7), 1009–1013. https://doi.org/10.1029/JA080i007p01009

Orlova, K., Shprits, Y., & Spasojevic, M. (2016). New global loss model of energetic and relativistic electrons based on Van Allen Probes measurements. *Journal of Geophysical Research: Space Physics*, *121*, 1308–1314. https://doi.org/10.1002/2015JA021878

Orlova, K., Spasojevic, M., & Shprits, Y. (2014). Activity-dependent global model of electron loss inside the plasmasphere. *Geophysical Research Letters*, *41*, 3744–3751. https://doi.org/10.1002/2014GL060100

Ozeke, L. G., Mann, I. R., Murphy, K. R., Rae, I. J., & Milling, D. K. (2014). Analytic expressions for ULF wave radiation belt radial diffusion coefficients. *Journal of Geophysical Research: Space Physics*, *119*, 1587–1605. https://doi.org/10.1002/2013JA019204

Pierrard, V., Goldstein, J., Andreé, N., Jordanova, V. K., Kotova, G. A., Lemaire, J. F., et al. (2009). Recent progress in physics-based models of the plasmasphere. *Space Science Reviews*, *145*(1–2), 193–229. https://doi.org/10.1007/s11214-008-9480-7

Rostoker, G. (1972). Geomagnetic indices. *Reviews of Geophysics*, *10*, 935–950.

Shprits, Y. Y., Meredith, N. P., & Thorne, R. M. (2007). Parameterization of radiation belt electron loss timescales due to interactions with chorus waves. *Geophysical Research Letters*, *34*, L11110. https://doi.org/10.1029/2006GL029050

Spasojevic, M., Shprits, Y. Y., & Orlova, K. (2015). Global empirical models of plasmaspheric hiss using Van Allen Probes. *Journal of Geophysical Research: Space Physics*, *120*, 10,370–10,383. https://doi.org/10.1002/2015JA021803

Takahashi, K., Toth, B. A., & Olson, J. V. (2001). An automated procedure for near-real-time *Kp* estimates. *Journal of Geophysical Research*, *106*(NoA10), 21,017–21,032.

Tan, Y., Hu, Q., Wang, Z., & Zhong, Q. (2017). Geomagnetic index Kp forecasting with LSTM. *Space Weather*, *16*, 406–416. https://doi.org/10.1002/2017SW001764

Tsyganenko, N. A. (1989). A magnetospheric magnetic field model with a warped tail current sheet. *Planetary and Space Science*, *37*(5), 1989.

Wang, J., Zhong, Q., Liu, S., Miao, J., Liu, F., Li, Z., & Tang, W. (2015). Statistical analysis and verification of 3-hourly geomagnetic activity probability predictions. *Space Weather*, *13*, 831–852. https://doi.org/10.1002/2015SW001251

Welling, D. T., Andre, M., Dandouras, I., Delcourt, D., Fazakerley, A., Fontaine, D., et al. (2015). The Earth: Plasma sources, losses, and transport processes. *Space Science Reviews*, *192*(1–4), 145–208. https://doi.org/10.1007/s11214-015-0187-2

Wing, S., Johnson, J. R., Jen, J., Meng, C.-I., Sibeck, D. G., Bechtold, K., et al. (2005). Kp forecast models. *Journal of Geophysical Research*, *110*, A04203. https://doi.org/10.1029/2004JA010500

Wintoft, P., Wik, M., Matzka, J., & Shprits, Y. (2017). Forecasting Kp from solar wind data: Input parameter study using 3-hour averages and 3-hour range values. *Journal of Space Weather and Space Climate*, *7*, 12. http://doi.org/10.1051/swsc/2017027

Yau, A. W., Peterson, W. K., & Abe, T. (2011). Influences of the ionosphere, thermosphere and magnetosphere on ion outflows. In W. Liu, & M. Fujimoto (Eds.), *The dynamic magnetosphere* (pp. 283–314). Dordrecht, Netherlands: Springer Science+Business Media. https://doi.org/10.1007/978-94-007-0501-2_16