SWSC

RESEARCH ARTICLE

Forecasting *Kp* from solar wind data: input parameter study using 3-hour averages and 3-hour range values

Peter Wintoft[1,*], Magnus Wik[1], Jürgen Matzka[2] and Yuri Shprits[2]

[1] Swedish Institute of Space Physics, Scheelevägen 17, 223 70 Lund, Sweden
[2] GFZ German Research Centre For Geosciences, Telegrafenberg, 14473 Potsdam, Germany

Abstract – We have developed neural network models that predict *Kp* from upstream solar wind data. We study the importance of various input parameters, starting with the magnetic component B_z, particle density n, and velocity V and then adding total field B and the B_y component. As we also notice a seasonal and UT variation in average *Kp* we include functions of day-of-year and UT. Finally, as *Kp* is a global representation of the maximum range of geomagnetic variation over 3-hour UT intervals we conclude that sudden changes in the solar wind can have a big effect on *Kp*, even though it is a 3-hour value. Therefore, 3-hour solar wind averages will not always appropriately represent the solar wind condition, and we introduce 3-hour maxima and minima values to some degree address this problem. We find that introducing total field B and 3-hour maxima and minima, derived from 1-minute solar wind data, have a great influence on the performance. Due to the low number of samples for high *Kp* values there can be considerable variation in predicted *Kp* for different networks with similar validation errors. We address this issue by using an ensemble of networks from which we use the median predicted *Kp*. The models (ensemble of networks) provide prediction lead times in the range 20–90 min given by the time it takes a solar wind structure to travel from L1 to Earth. Two models are implemented that can be run with real time data: (1) IRF-*Kp*-2017-h3 uses the 3-hour averages of the solar wind data and (2) IRF-*Kp*-2017 uses in addition to the averages, also the minima and maxima values. The IRF-*Kp*-2017 model has RMS error of 0.55 and linear correlation of 0.92 based on an independent test set with final *Kp* covering 2 years using ACE Level 2 data. The IRF-*Kp*-2017-h3 model has RMSE = 0.63 and correlation = 0.89. We also explore the errors when tested on another two-year period with real-time ACE data which gives RMSE = 0.59 for IRF-*Kp*-2017 and RMSE = 0.73 for IRF-*Kp*-2017-h3. The errors as function of *Kp* and for different years are also studied.

Keywords: *Kp* index / forecasting / neural network / verification

1 Introduction

The planetary *Kp* index (Bartels et al., 1939; Mayaud, 1980) is widely used as a general indicator of geomagnetic disturbances for mid-latitude regions, and as a general geomagnetic alert and hazard scale. The index is based on vector measurements of the geomagnetic field from 13 mid-latitude stations. The *Kp* index, or the equivalent *ap* index, is also used as input for numerous models due to the availability of historic values, near real-time data, and forecasts, but also due to the simplification it offers to characterise the magnetosphere with a single number. The latter is also supported by the interpretation of *Kp* as an indicator of magnetospheric convection (Thomsen, 2004). For example,

the *Kp* index is used as input to atmospheric density models (Bruinsma, 2015) used for satellite drag calculations; thermosphere/ionosphere models (TIE-GCM; Qian et al., 2014); plasmapause models (Heilig and Lühr, 2013); and auroral precipitating electron models (Zhang and Paxton, 2008). During the recent two decades the magnetospheric community has developed a number of robust parameterisations that are based on *Kp*. Parameterisations that are based on a single index significantly simplifies model calculations. Model parameters can be easily precomputed and do not require real time estimations as calculations for the parameterisations that are based on multiple parameters of the solar wind or geomagnetic indices. The first realistic empirical model of the magnetic field (Tsyganenko, 1989) is still often used to predict the dynamics of the magnetic fields when only one input variable needs to be used and magnetic field structure needs to be pre-computed. Simple prediction

*Corresponding author: `peter@lund.irf.se`

models of the ULF activity with *Kp* index (Brautigam and Albert, 2000) allowed Shprits et al. (2005, 2006) to perform long term simulations of the radiation belt with a simple model for radial diffusion. *Kp*-based parameterisations of VLF chorus and ELF/VLF hiss allowed for comprehensive and detailed long-term simulations of the dynamics of the radiation belts (Tu et al., 2013, and references therein). Recent studies of Orlova and Shprits (2014) and Orlova et al. (2016) further improved the accuracy of the *Kp* based models and provided parameterisations that can be used for quiet times and disturbed conditions. Introduction of more sophisticated models of the processes in the inner magnetosphere based on solar wind conditions and combination of other indexes may potentially lead to improvements as is the case for the global models of the magnetic field (Tsyganenko and Sitnov, 2005, 2007). There are also recent attempts to infer parameters of the magnetospheric processes by remote sensing or indirect measurements (Murphy et al., 2016). However, inaccuracies associated with indirect measurements, measurement errors, and incomplete coverage of measurements may in fact result in parameterisations that are less accurate than models based on the global *Kp* index. While future research should strive to provide improved parametrisations based on solar wind data and various indices, *Kp* based empirical models are likely to remain the workhorse of space physics and space weather models. *Kp* also remains the single most used geomagnetic index by satellite operators.

Thus, there has been a lot of effort to develop prediction models for the *Kp* index. Models have been developed that predict *Kp* from solar wind data at different temporal resolution and for different lead times using different algorithms, such as neural networks (Boberg et al., 2000; Wing et al., 2005; Bala and Reiff, 2012); support vector machines (Ji et al., 2013); probabilistic models (Wang et al., 2015); and models developed from system identification called non-linear autoregressive moving-average with external inputs (NARMAX and variants) (Ayala Solares et al., 2016). The different methods have in common that they derive the mathematical relations from the provided data and that they consist of non-linear functions with multi-dimensional inputs. The different models are very capable and thus it is difficult to decide on any specific methodology which is superior for the task. More important is the selection of representative datasets and input parameters, which is non-trivial as it should include estimates of the multi-dimensional probability distribution, but often one resorts to the one-dimensional case. The relevant inputs can be explored during the training of the models where either the individual physical plasma and magnetic field parameters are used, or some combination like the Boyle index (Bala and Reiff, 2012). Selecting the "best" model from verification measures is also not trivial, as different measures will lead to different choices. The measures-oriented approach may even lead to the selection of a non-optimal model (Murphy, 1993).

Several models have previously been implemented for real-time operation and have been running for many years (Boberg et al., 2000; Wing et al., 2005; Bala and Reiff, 2012). The three cited models are based on different types of neural networks. The inputs to the first model are 3-hour averages of solar wind density, velocity, and interplanetary magnetic field (IMF) component B_z in geocentric solar magnetospheric (GSM) coordinates. The second model uses 15 min averages of solar wind plasma and IMF to predict 15-minute linearly interpolated *Kp* values. The third model uses the hourly averages of the Boyle index, which is defined as $\Phi = 10^{-4} V^2 + 11.7B \sin^3 \theta/2$, with solar wind speed *V*, IMF magnitude *B*, and IMF clock angle θ, and with the oversampled *Kp* (the same *Kp* value is repeated 3 times) as target output.

For atmospheric/magnetospheric models driven by *Kp* the response time may be many hours or days implying that a *Kp* prediction model with short lead time is less useful. However, both observations and models may also show an immediate response to changes in *Kp* (Shprits et al., 2006) motivating the short lead time predictions.

We describe here our latest prediction models developed within the EU/H2020 project PROGRESS. The purpose of this work is to improve our existing *Kp* prediction model and to provide predicted *Kp* as input to radiation belt models that are developed within PROGRESS. The *Kp* prediction models are driven by solar wind plasma and magnetic field data. In this study we use only measured solar wind at L1, but in the future the goal is to drive the *Kp* prediction model with predicted solar wind.

2 Data

The data sets consist of ACE (Stone et al., 1998) 64-second resolution plasma data (McComas et al., 1998), 16-second resolution IMF data (Smith et al., 1998),[1] and the definitive 3-hour *Kp* values[2] (Mayaud, 1980). We resample the ACE data to 1-minute resolution. The ACE data are then propagated from the location of the spacecraft at L1 to Earth using the velocity in the GSE-*x* direction. This corresponds to the "flat delay", the simplest method out of four described in Mailyan et al. (2008). More specifically, we propagate the solar wind by shifting the timestamps. If a particular solar wind sample *i* at ACE has timestamp and data (t_i, x_i) the propagated sample will have

$$(s_i, x_i) = (t_i + \tau_i, x_i), \tag{1}$$

where $\tau_i = r_i/v_i$, *r* is the distance from ACE to 10 Earth radii $(10R_E)$ upstream from Earth and *v* is the speed. We use $10R_E$ as an approximate location of the bow shock, although it will vary by several R_E for different solar wind conditions (with typical timing errors less than a minute). We call τ the *propagation lead time*. Note that the s_i series will not have constant temporal cadence as both *r* and *v* changes with time. The observed propagation lead times τ are distributed bi-modally with peaks at 44 and 58 min, and with minimum and maximum at 20 and 90 min, respectively. Note that s_{i+1} can be earlier than s_i ($s_{i+1} < s_i$) if there is a large enough speed increase from one sample to the next, but this will have a minor effect as 3-hour values are created from the series. For each 3-hour *Kp* interval the corresponding intervals in the (*s*, *x*) solar wind series are identified and representative measures, e.g. 3-hour averages, are computed.

3 Models, time resolution and lead time

From an algorithmic perspective it can be advantageous to increase the temporal resolution by linear interpolation of the
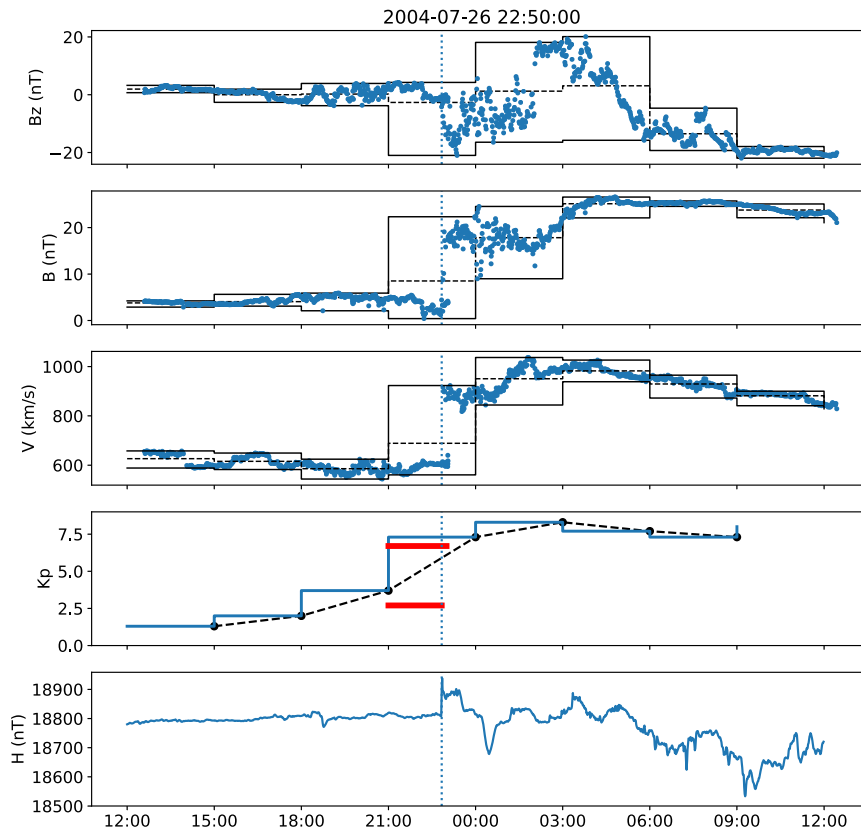
---

[1] http://www.srl.caltech.edu/ACE/ASC/level2/.
[2] http://www.gfz-potsdam.de/en/kp-index.

Fig. 1. An example showing the ACE solar wind data and *Kp*. The top two panels show the IMF B_z (GSM) and *B* 1-minute averages (blue curve), 3-hour averages (dashed line segments), and 3-hour minima and maxima (solid line segments). The middle panel shows the solar wind speed. The fourth panel shows final *Kp*. The solid blue line segments show the *Kp* values over their 3-hour intervals. Each dot shows the *Kp* value at the end of the 3-hour interval, with linearly interpolated values connecting the dots (dashed lines). The red horizontal lines show nowcast *Kp* determined from geomagnetic data up to 22:50UT and up to 23:00UT, respectively. The bottom panel shows the horizontal geomagnetic field at NGK with 1-minute resolution. The ACE 1-minute averages are the propagated values from L1 to Earth using the solar wind speed. The timestamp at the top indicates the location of the vertical dotted line.

3-hour *Kp* index in order to provide more timely predictions (Wing et al., 2005). However, it should be noted that *Kp* is calculated from the *K* values from the 13 *Kp* stations worldwide, where *K* represents the maximum range of geomagnetic variation with 1-minute resolution over a 3-hour UT interval. To obtain the geomagnetic variations, a quiet curve has to be estimated and subtracted from the recordings at the station. The variation is then projected along magnetic North (horizontal component), magnetic East and the vertical component and the maximum variation of these three components is translated into the corresponding *K* value (usually, the horizontal component shows the strongest variation). Therefore, it becomes unclear what an interpolated value represents.

Another approach to increase the temporal resolution is to use sub-3-hour-averages of the solar wind with *Kp* values repeated, e.g. in Bala and Reiff (2012) hourly solar wind averages are targeted against *Kp* values that are repeated 3 times for each 3-hour interval.

We take the approach of developing a prediction model in accordance with the original 3-hour definition of *Kp*. But relying on 3-hour averages of the solar wind (Boberg et al., 2000; Ayala Solares et al., 2016), to match the 3-hour *Kp*, may be problematic under certain conditions as there can be considerable change in the parameters that will be lost in the

average. However, other 3-hour transformations can be used instead of the average. In this work we use, in addition to 3-hour averages, the 3-hour minimum and maximum values of the solar wind data. Figure 1 shows IMF and speed propagated from L1 to Earth, the *Kp* values, and the horizontal geomagnetic field at Niemegk (NGK) geomagnetic observatory for an event on 26–27 July 2004. Each point in the solar wind plots is the measured value but shifted in time using observed velocity. If the velocity is missing the other parameters will also be missing as the lead-time cannot be determined. Note that when there is a sudden increase in velocity there may be multiple points for the same timestamp, but this does not affect the calculation of the minimum, average, or maximum values. Note also that the solar wind plots have the shifted time as timestamps, while the *Kp* and H plots show the true time. Thus, before 22:27UT the velocity is around 600 km/s giving a lead time of about 40 min with shifted timestamps extending to 23:02UT, then at 22:27UT the velocity jumps to above 900 km/s reducing the lead-time to about 25 min with shifted timestamp at 22:52UT thereby causing shifted time to go "backwards", clearly seen in the velocity plot. There is a storm sudden commencement (SSC) at 22:50UT (H component at NGK) which is caused by the "shock" in the solar wind, and it is close to the estimated time

of the propagated shock at 22:52UT. We use the term "shock" here without actually showing that it is a shock, we simply note that there is a sharp jump in speed. The SSC is then followed by the geomagnetic storm. As the SSC takes place in the 21:00–00:00UT interval there is a corresponding increase in *Kp* which reaches 7o. As the shock occurs in the later part of the 3-hour interval the solar wind averages do not capture the relevant physical properties, and can therefore not be used to predict the *Kp* = 7o. However, looking instead at the 3-hour maxima of the solar wind speed it can be associated with the *Kp* increase.

To explore the immediate effect on *Kp* from the sudden change in the horizontal geomagnetic field we compute a nowcast *Kp* with 10-minute cadence for the event in July 2004 over the interval 21UT through 23UT (Fig. 1). The reconstructed nowcast *Kp* is based on K values calculated using the FMI method (Menvielle et al., 1995) from definitive 1 minute data of the contributing observatories, which we obtained from INTERMAGNET. Computing *Kp* using geomagnetic data for the interval 21:00–22:50UT results in the value *Kp* = 3−. Extending the interval to 23:00UT gives *Kp* = 6+ which is close to the final *Kp* = 7o for that 3-hour interval. Thus, in the short 10-minute interval 22:50–23:00UT which contains the SSC most of the *Kp* increase has already taken place. A consequence of this is that we cannot expect, in the general case, that any further prediction lead time is possible than that given by the propagation lead time, when measured solar wind data is used.

4 Algorithms

In general terms, during model development and evaluation, the solar wind and *Kp* data are treated as series of values. The temporal alignment of the data are determined from each value's associated timestamp. As each value is generated by measurements over some time interval there will be some ambiguity in setting the timestamp. For the data selection and model development we define the timestamp as the time at the beginning of a measurement interval. For a real-time application it may be more convenient to put the timestamp at the end of the interval, in which case the timestamp of predicted *Kp* marks the end of the 3-hour interval.

The algorithm used is a feed-forward neural network with one hidden layer, similar to that used by Boberg et al. (2000). The mapping from input to output is captured by the function

$$y(\mathbf{x}) = a + \mathbf{v}g(\mathbf{b} + \mathbf{W}\mathbf{x}) = a + \sum_{i=1}^{m} v_i g\left(b_i + \sum_{j=1}^{n} w_{i,j} x_j\right) \quad (2)$$

where *a* is a scalar, **v** is the row matrix $[v_1,..., v_m]$, **b** is the column matrix $[b_1,..., b_m]^T$, **W** is the $m \times n$ matrix

$$\begin{pmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{m,1} & \cdots & w_{m,n} \end{pmatrix} \quad (3)$$

and **x** is the input column matrix $[x_1,..., x_n]^T$. The transfer function *g* is the bounded non-linear tanh function. The network can be viewed as being composed of several processing units organised in layers, where there are *n* input units, *m* hidden units, and one output unit.

The target output is the observed *Kp*, where *Kp* is treated as a continuous variable, although in reality it is a categorical variable with 28 different classes, $Kp \in \{0o, 0+, 1−, 1o,..., 9−, 9o\}$. The *Kp* categories are mapped to the variable $y \in \{0.0, 0.3, 0.7, 1.0,..., 8.7, 9.0\}$.

Finding the inputs x_j that are useful for *Kp* forecasts is part of the network development. The solar wind plasma and IMF acts on the magnetosphere through different processes that results in the temporal variation of the ground magnetic field from which *Kp* is derived. From the vast number of solar wind-magnetosphere studies the commonly used parameters for predictions are solar wind magnetic field component B_z, or both B_y and B_z, particle density *n*, and speed *V*, or expressed as the dynamic pressure $p \propto nV^2$ and the convective electric field component $E_y \approx −VB_z$. The dynamic pressure plays an important role in the creation of the global geomagnetic disturbances known as sudden impulses (SI) (Segarra et al., 2015), which becomes classified as a SSC if there is a subsequent storm, while the solar wind electric field E_y is generally considered to control the reconnection rate (Pulkkinen et al., 2007), although it has been suggested that it is not the solar wind electric field that controls the reconnection rate but it works due to a coincidence (Borovsky and Birn, 2014). As the neural network can approximate continuous functions arbitrarily well (Cybenko, 1989) the individual solar wind parameters can be provided as inputs from which the relations are found. In this work we study the magnetic magnitude $B = \sqrt{B_x^2 + B_y^2 + B_z^2}$, the B_y and B_z components, particle density *n*, and speed *V*.

Once the optimal network has been identified, all the parameters *a*, **v**, **b**, **W** are fixed to constant values and the output depends only on the input data. The parameters are called weights.

5 Results

5.1 Selection of datasets

The ACE solar wind data set used in this work extends from 1998 to 2015, during which there is also complete coverage of definitive *Kp* data. From a training and verification perspective it is desirable to have sub-sets with similar *Kp* statistics. In principle the sub-set selection can be made by treating the samples individually with the constraint that the statistics should be similar. However, as we also would like to study the performance for individual storms further constraints are required for the sub-set selection. We do this by performing the sub-set selection on 1-year chunks of data and studying the distribution. The full dataset is divided into three independent datasets which we call the training set, validation set, and test set. The three sets are composed of data for the years (1998, 1999, 2002–2005, 2007–2010, 2013–2015), (2000, 2006, 2012), and (2001, 2011), respectively. As the solar wind variables numerically have very different ranges they are normalised to lie in the range [−1, + 1], where the minimum and maximum values for each solar wind parameter get the values −1 and +1, respectively. The normalisation constants are determined from the training set.

5.2 Training procedure

A specific network is determined by the number of inputs and the number of hidden units, and by a certain realisation of random
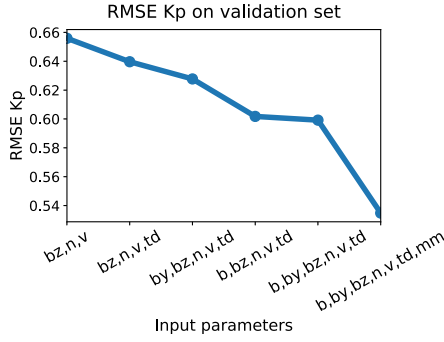
Fig. 2. The validation set RMS error as function of different combinations of solar wind inputs.

weights. During the training phase the training set is iteratively used to adapt the weights in order to minimise the summed squared errors (SSE) between the predicted Kp and the observed Kp. Given a large enough network this error will approach zero and the network would work as a table-lookup with very poor generalisation capabilities. Therefore, at each iteration in the training phase the SSE is also computed using the validation set. The number of iterations in the training phase is experimentally determined to continue well beyond the minimum in the validation SSE to make it plausible that the global minimum has been reach, although it cannot be guaranteed. The specific weights that lead to the minimum validation SSE defines the optimal network given the number of inputs and hidden units, and initial weights. Note that the test set has not been used during this phase but is left for the final evaluation.

5.3 Model inputs

The temporal evolution of Kp is determined by the solar wind driven processes and by internal magnetospheric processes, where the latter can be referred to the memory of the system. As the network in equation (Eq. (2)) has no internal memory the dynamics are instead coded into the inputs using past solar wind values. A dynamic network, like the Elman neural network (e.g. Lundstedt et al., 2002; Wintoft et al., 2015), could have been used instead, however, the manipulation of the time series becomes simpler when a non-dynamic network is used.

For each time shifted (Eq. (1)) solar wind parameter B, B_y, B_z, n, V, the 3-hour minimum, average, and maximum are calculated. Here we denote the magnitude of the vector magnetic field **B** with B. The timestamps t_i follow the cadence of Kp. The corresponding solar wind samples j with $t_i \leq s_j < t_i + 3$ hours are selected and the 3-hour values are computed. Thus

$$B_{\min}(t_i) = \min\{B_j, \forall j \quad \text{where } t_i \leq s_j < t_i + 3h\} \quad (4)$$

$$B_{avg}(t_i) = \langle\{B_j, \forall j \quad \text{where } t_i \leq s_j t_i + 3h\}\rangle \quad (5)$$

$$B_{\max}(t_i) = \max\{B_j, \forall j \quad \text{where } t_i \leq s_j < t_i + 3h\} \quad (6)$$

and similarly for B_y, B_z, n, and V. We can then collect any combination of solar wind input at time t_i into a vector **u** giving
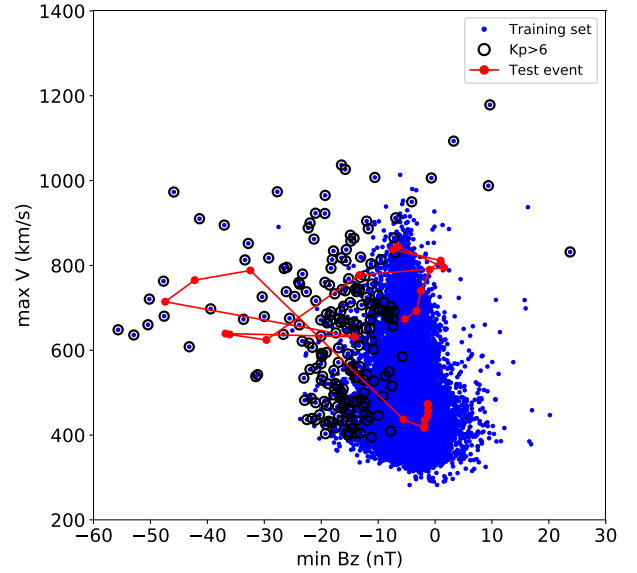

Fig. 3. Scatter plot of 3-hour $V_{\max}$ vs. 3-hour $B_{z,\min}$ for the training set (small blue dots) with samples marked with rings for $Kp > 6$. A storm event from the test set for the period 30 March 2001–1 April 2001 is shown as red dots connected with lines.

the series $\{\mathbf{u}(t_i)\} = \{\mathbf{u}_i\}$ and the corresponding target series $\{Kp_i\}$.

Using all the above solar wind parameters we first explore the required number of past solar wind inputs, i.e. the length of the time delay line. The input vector **x** for sample i (time t_i) is thus created as $\mathbf{x}_i = [\mathbf{u}_{i-N}, \ldots, \mathbf{u}_i]$ consisting of $N+1$ 3-hour segments. We train and validate a large number of networks with time delays extending from $N = 0$ to 8 and find that $N = 2$ is sufficient. We then fix the time delay line to $N = 2$ and study the effect of the various inputs. There is a large number of combinations that can be studied but we limit our search to the following six combinations: (B_z, n, V); (B_z, n, V, td); (B_y, B_z, n, V, td); (B, B_z, n, V, td); (B, B_y, B_z, n, V, td); and $(B, B_y, B_z, n, V, td, mm)$. The first is the same as used by Boberg et al. (2000). In the second case we include inputs indicated by the label td that corresponds to functions of local time and day of year according to

$$\sin 2\pi T/24, \cos 2\pi T/24, \sin 2\pi D/365, \cos 2\pi D/365 \quad (7)$$

where T and D are UT hour and day of year, respectively. We included this case as Kp show a similar dependence on local time and season as the Dst index (Cliver et al., 2001). The third case is similar to Bala and Reiff (2012) that uses (B, θ) which can be transformed to (B_y, B_z). In the fourth case we use the magnitude B instead of B_y, while in the fifth case we include both B and B_y. The final case with the parameter mm represents the inclusion of the minimum and maximum values of the solar wind parameters based on 1-minute data. Going through the training and validation procedure for the above combinations results in the RMS errors on the validation set as shown in Figure 2. We see that there is a decrease in error when time information (td) is added. There is further decrease when B_y is added, however, the decrease is much larger when total field B is used instead. Including both B and B_y makes a very little difference compared to using only B. Finally, including the maximum and minimum values makes a big improvement.
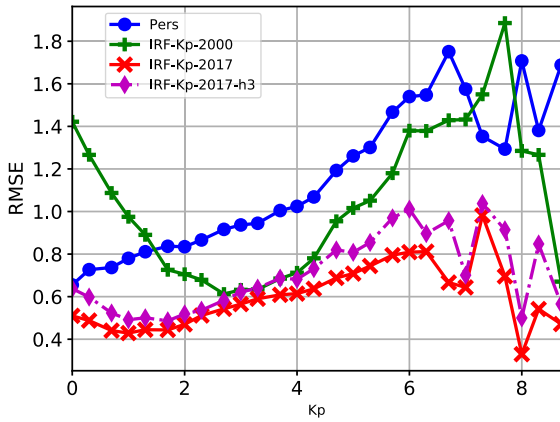
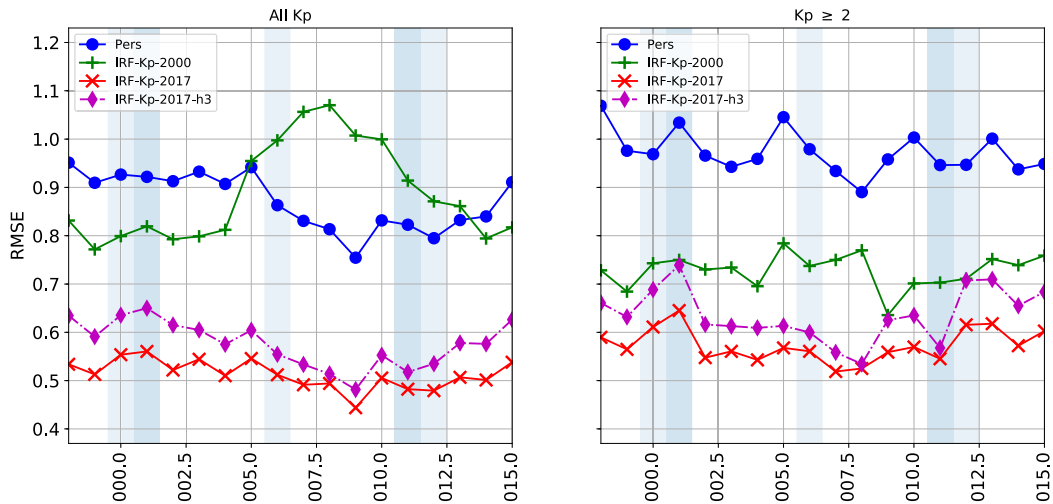**Fig. 4.** The RMS error (RMSE) as function of observed *Kp*.

Table 1. Measures and scores for predicted *Kp* vs. observed *Kp* using all data.

	BIAS	MAE	RMSE	CORR	MSESS:PERS
Pers	0.02	0.67	0.89	0.79	0.00
IRF-*Kp*-2000	0.39	0.72	0.88	0.83	0.02
IRF-*Kp*-2017-h3	0.00	0.46	0.59	0.91	0.56
IRF-*Kp*-2017	0.00	0.41	0.52	0.93	0.66

Fig. 5. The RMS error (RMSE) for each year. The validation set and test set are indicated with the light blue and dark blue shaded areas, respectively. The left plot is based on all data and the right plot only data for which *Kp* ≥ 2.

For each of the six input combinations a large number of networks are trained with different initial random weights and different number of hidden units n_h (3–15), and in Figure 2 the RMSE for the networks with the smallest validation errors are shown. Selecting a set of networks with inputs (*b*, *by*, *bz*, *n*, *v*, *td*, *mm*), all with validation RMSE close to the optimal, we notice that there can be considerable differences in predicted *Kp* from different networks at large *Kp* values. The reason for this is due to the following. The distribution of *Kp* falls off with increasing *Kp* and the input space is sparsely populated for large *Kp*. In Figure 3 we plot V_{max} vs. $B_{z,min}$ for the training set, but we could also selected any other pairs of parameters. There is a large number of data points around the central cluster of the distribution and for the extreme values the density is low. All training samples with *Kp* > 6 are also marked. During training the network tries to model this distribution and providing reliable generalisation becomes more difficult as the sample density goes down. An event from the test set (dots connected with lines) indicates the movement in the (V_{max}, $B_{z,min}$) space: It starts at B_z close to zero and small *V* and as the CME hits the Earth it moves into the low density region. Thus, even though different networks have similar validation RMSE the fitted functions in the low density regions may be quite different. In Figure 3 we only plot two variables, but in reality we have five variables making the distribution even sparser. To improve on the situation we use an ensemble of the 10 networks that have the lowest validation RMS errors and use the median predicted *Kp*, and also compute the variance of the predictions. It should be noted that the variance is very small in the densely populated region while it grows for the more extreme events.

In the following we define two models that have slightly different inputs. In both cases we use an ensemble of neural networks with all solar wind parameters and the UT-DOY functions. The difference is in whether the maximum and minimum values derived from 1-minute data are used or not. In the IRF-*Kp*-2017 model we include the max/min values, while the IRF-*Kp*-2017-h3 model only relies on the averages.

5.4 Verification of models

The models developed here are compared against each other, and against the Boberg et al. (2000) model that is currently running at RWC-Sweden[3]. The simple persistence model, predicted *Kp* is equal to the previous *Kp*, is also

---

[3] http://lund.irf.se/forecast/kp/.

Table 2. Measures and scores for observed *Kp* and predicted *Kp* using test data.

	BIAS	MAE	RMSE	CORR	MSESS:PERS
Pers	0.01	0.67	0.90	0.78	0.00
IRF-*Kp*-2000	0.31	0.68	0.84	0.83	0.14
IRF-*Kp*-2017-h3	−0.04	0.48	0.63	0.89	0.52
IRF-*Kp*-2017	−0.04	0.42	0.55	0.92	0.63

Table 3. RMS error before onset (pre), after onset (post), and the average of the two. There is in total 368 events. Numbers inparenthesis are computed from the test set for which there are 45 events.

	pre	post	average
Pers	1.09 (1.22)	2.76 (2.96)	1.92 (2.09)
IRF-*Kp*-2000	1.00 (0.85)	1.47 (1.77)	1.24 (1.31)
IRF-*Kp*-2017-h3	0.79 (0.81)	1.27 (1.73)	1.03 (1.27)
IRF-*Kp*-2017	0.70 (0.60)	0.92 (1.10)	0.81 (0.85)

Table 4. RMSE and CORR for predicted *Kp* using ACE Level 2 data (L2) and ACE real-time data (RT) as inputs for the period 1 April 2011 to 1 March 2013. Coverage indicates whether samples corresponding to timestamps of the L2 or RT set have been used in computing RMSE and CORR. Median indicates whether the 5-minute median filter to *n* and *V* has been applied.

	Model	Input	Coverage	Median	RMSE	CORR
1	IRF-*Kp*-2017	L2	L2	False	0.49	0.92
2	IRF-*Kp*-2017	RT	L2	True	0.51	0.91
3	IRF-*Kp*-2017	RT	RT	True	0.59	0.89
4	IRF-*Kp*-2017	RT	RT	False	0.65	0.86
5	IRF-*Kp*-2017-h3	L2	L2	False	0.54	0.91
6	IRF-*Kp*-2017-h3	RT	L2	True	0.56	0.90
7	IRF-*Kp*-2017-h3	RT	RT	True	0.73	0.85
8	IRF-*Kp*-2017-h3	RT	RT	False	0.75	0.84

included in the comparison, where the observed *Kp* are the final *Kp*. We apply 5 statistical measures on the complete dataset covering the years 1998 and onwards (Tab. 1). The measures are: the BIAS, or mean error; mean absolute error (MAE); root mean square error (RMSE); linear correlation (CORR); and the mean squared error (MSE) skill score $1 - \mathrm{MSE_{model}}/\mathrm{MSE_{persistence}}$ with persistence as the reference. We do not consider persistence a useful model in itself, however, on many scores it performs very well due to autocorrelations in the series.

From Table 1 it is seen that the Boberg et al. (2000) (IRF-*Kp*-2000) performs poorer than the persistence model with respect to BIAS and MAE, while the RMSE are similar, and the CORR are slightly better. Both IRF-*Kp*-2017 models show significant improvements on all scores compared to the IRF-*Kp*-2000 model.

The statistics in Table 1 are computed on all available data, thus including all three sets: training set, validation set, and test set. We therefore also compute the measures using only the test set (Tab. 2) and the result is similar.

We further explore the model performance over the range of *Kp* values. In Figure 4 the RMSE is plotted as function of *Kp*. The IRF-*Kp*-2000 model has small errors in a quite narrow interval of *Kp* values. The performance is quite poor below 1+, which is also reflected by the positive BIAS in Tables 1 and 2. The IRF-*Kp*-2017 models show a much more consistent behaviour over the full *Kp* range, and at all points they are better than the persistence model.

It is also interesting to study the performance as function of time. In Figure 5 the RMSE is computed on each year from 1998 to 2015, where we have also indicated the years

belonging to the training set (white background), validation set (light blue), and test set (darker blue), respectively. A striking feature is the huge increase in RMSE for IRF-*Kp*-2000 centered on year 2008 (left plot). However, that is an effect of the poor performance for *Kp* < 2 and those years are dominated by low *Kp*. Including only data with *Kp* ≥ 2 (right plot) removes the peak. The IRF-*Kp*-2017 model performs consistently well over all years. Note also that the RMSE for the validation and test sets are similar to that of the training set.

Finally, we study the errors around instances when there is an increase in observed *Kp* from one 3-hour interval to the next. We identify all events when $Kp(t) - Kp(t-3\mathrm{h}) > 2$ and compute the RMSE for each model for the pre-event 3-hour interval, and the post-event 3-hour interval. The results are summarised in Table 3. Here we clearly see that the persistence performs poorly, as expected, and that the IRF-*Kp*-2017 models performs best. When the min/max values are not used (IRF-*Kp*-2017-h3) the predictions will be temporally smoother and therefore miss some of the *Kp* increases.

In Bala and Reiff (2014) several different *Kp* prediction models were validated with different statistical measures for data over the period April 2011 to February 2013. The models are driven by solar wind using various combinations of parameters with lead times from 1 to 4 h. The evaluation of their 1-hour forecast model resulted in RMSE = 0.83, CORR = 0.77 (Tab. 1 in Bala and Reiff, 2014). In our Table 2 (test data) we obtain RMSE = 0.55 and CORR = 0.92 for the IRF-*Kp*-2017 model, and more specifically for the same years (2011–2013) we see that RMSE varies between 0.48 to 0.51 (Fig. 5, left plot) and between 0.52 to 0.58 without min/max inputs
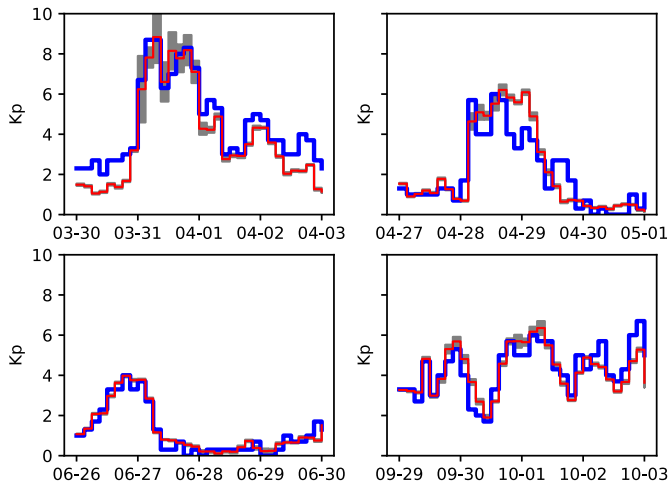
Fig. 6. Four events from the test set (2001) covering 4 days each. Observed *Kp* (blue thick line), predicted *Kp* from the IRF-*Kp*-2017 model (red thin line), and the σ (grey regions) are shown.

(IRF-*Kp*-2017-h3). However, the Bala and Reiff (2014) model was evaluated on ACE real-time data. We therefore run our models for the same time period (April 2011 to February 2013) using both ACE Level 2 data (L2) and ACE real time data (RT) which results in the statistics shown in Table 4. The RMSE using L2 data is consistent with the RMSE in Figure 5, but we see that there is an increase in errors for both models when RT data is used. This is an effect of that the RT data are preliminary and contain larger measurement errors. The largest differences are seen between the L2 and RT plasma density, the correlation is only 0.88 while the other solar wind parameters have correlations between 0.94 to above 0.99. There are also several instances when the L2 density is below $20\,\mathrm{cm}^{-3}$ while the RT density lies between $50\text{–}150\,\mathrm{cm}^{-3}$. Single spikes in the RT data can be removed with a 5-minute median filter and has a significant effect on the results (Median column). Rows 3 and 4 show that the median filtering has a larger effect on the IRF-*Kp*-2017 model as compared to the IRF-*Kp*-2017-h3 model (rows 7 and 8). Another issue is that there are fewer available records for the L2 density data than for the RT density data, thus comparing the prediction statistics directly between the L2 and RT driven models is not useful. Therefore, when the models are run with RT data we compute the statistics for the samples in the full RT set, but also for the smaller set that overlaps with the timestamps in the L2 set (coverage column). Thus, the statistic can be compared between rows 1 and 2, and 5 and 6.

5.5 Dependence on distance from line-of-sight to spacecraft

The ACE spacecraft orbit around the L1 location can bring it far from the Sun-Earth line-of-sight (LOS). The distance is $R = \sqrt{y^2 + z^2}$ in GSE coordinates and can reach about 300 Mm corresponding to an angle of $12°$ as seen from Earth. Thus, we would like to study if the LOS distance can have an effect on the prediction accuracy. A possible dependency of the errors on the line-of-sight distance should affect both models therefore we only study the most accurate model. The

prediction errors from the IRF-*Kp*-2017 model are binned as function of *R* and for each bin the RMSE and maximum error are computed. The bins are chosen so that each contain 5% of the data (approx. 1800 samples each). However, no systematic variation as function of *R* can be seen.

5.6 Case studies

We select four events from the test set for year 2001, each extending over four days, and run the IRF-*Kp*-2017 model using the ACE Level 2 one-minute data. The results are shown in Figure 6. The events have different characteristics with the event in March–April 2001 being the strongest with observed *Kp* = 9− (blue thick curve). In the pre-storm period *Kp* varies between 2o to 3+ while the predicted *Kp* (red thin curve) varies between 1 and 1.5. At midnight *Kp* jumps to 7− with the prediction at ≈6, followed by another jump to 9− where predicted *Kp* first reaches 7.5 and then comes close to 9−. As this is an ensemble prediction we also compute the standard deviations (σ) which are shown with grey bars. For low *Kp* values the σ is very small and increases with increasing *Kp*. It should be noted that the σ value does not provide any information on how accurate the prediction is, it instead reflects the model uncertainty due to the low density sample space from which the models were derived (Fig. 3). In regions in the input space with high sample density, usually corresponding to low *Kp* values, all the models in the ensemble provide very similar predictions and thereby small σ values. The April–May event is also a clear storm with a large increase in *Kp* but weaker than the March–April event. In June there is a weak more gradual storm that is followed by very quite conditions that the model captures well. The September–October event shows more extended activity at medium levels.

It is interesting to study how the models operate when continuously fed with solar wind data at 1-minute cadence. The same event as in Figure 6 (top, left) is shown in Figure 7 but in more detail around the event onset. The prediction models are run with 1-minute cadence by sliding the 3-hour filter on the 1-minute solar wind data. As mentioned before, the timestamps identifying each 3-hour interval can be defined to be any time within the interval, but once defined needs to be consistently applied. For the purpose of Figure 7 we define the timestamps to mark the end of each 3-hour interval. At the time indicated by label 1 in the solar wind data (top panel) the last 3-hour solar wind input extends three hours back in time from that point. The IRF-*Kp*-2017 model provides a predicted *Kp* and predicted lead time τ giving the point 1 in the *Kp* plot (third panel). At point 2 the shock has passed and at that time the predicted *Kp* and τ gives the point 2 in the *Kp* plot. The red curve in the *Kp* plot traces each prediction for each minute, and the timestamps of the predictions will not be sorted in increasing order as the lead-time varies. As we put the timestamps at the end of the 3-hour input intervals it also means that the predicted *Kp* marks the end of each 3-hour interval, and the 3-hour intervals have been indicated by the red horizontal bars ending at each point. Clearly, we could have chosen the timestamps to mark the start of each interval on the inputs and then the points in the *Kp* plot would be shifted left by 3 h, but the red horizontal bars would be unchanged.
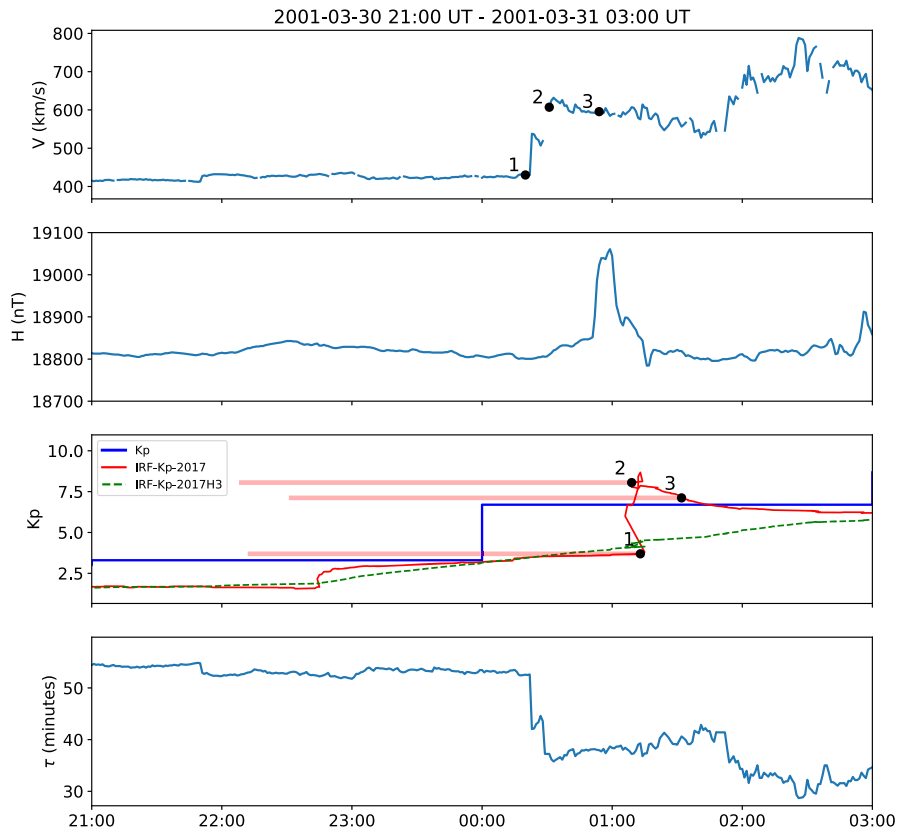
Fig. 7. Predicted *Kp* at 1-minute cadence. From top to bottom: solar wind velocity at spacecraft location; horizontal geomagnetic field at NGK; final and predicted Kp; prediction lead time τ in minutes. In the third panel, the red solid curve is predicted *Kp* from IRF-*Kp*-2017, and green dashed curve using IRF-*Kp*-2017-h3. The labels 1, 2, and 3 (first and third panels) mark three moments with observed inputs and corresponding predictions. The three horizontal red lines indicates the corresponding 3-hour intervals that the forecasts represent.

The shock in the solar wind arrives at the spacecraft just after 00:20UT on the 31st (label 1, top panel). There is an SSC observed at NGK with peak at 01:00UT (second panel). The observed *Kp* is 3+ in the 3-hour interval preceding the interval with the SSC and then increases to *K* = 7o (blue step curve, third panel). At 00:20UT the solar wind speed is 420 km/s (label 1), giving a prediction lead time of τ ≈ 55 min, thus providing the predicted *Kp* = 3.7 as shown by label 1 in the *Kp* plot. During the shock passage the lead time decreases to τ ≈ 40 min and predicted *Kp* increases to 8.0 (label 2). Note that the last pre-shock predicted *Kp* (1) is further into the future than the first predicted post-shock *Kp* (2). Naturally, in a real-time setting only the latest predicted *Kp* is used. Also note that the SSC indicates that the shock arrives earlier than predicted, which can be seen in that the peak in H (second panel) comes before label 2 in the *Kp* plot. However, the prediction still gives a warning of the event although the lead time is shorter than 40 min. For comparison, the predictions from the IRF-*Kp*-2017-h3 model (without min/max values) are also shown (green dashed). The model performs similar but smoother and with some lag.

6 Discussion and conclusions

Kp prediction models driven by measured solar wind inevitably provide short lead times, around 20–90 min, even though *Kp* has a resolution of 3 h. The short lead time is determined by events with sudden changes in the solar wind, like shocks, that may cause sudden changes in the geomagnetic field to which *Kp* is sensitive as it measures the range of variability. Models that rely on 3-hour averages (Boberg et al., 2000), or interpolated values at higher cadence than 3 h (Wing et al., 2005), will not be able to capture that characteristic of the *Kp* index. However, the impact of this effect will depend on the time within the fixed 3-hour interval (00:00–03:00, 03:00–06:00, ...) at which substantial changes in the solar wind, like shocks, occur. In real-time operation (where the 3-hour window glides) the effect will always become more apparent. We find that the only available prediction lead time, when using measured solar wind, is that given by the propagation lead time from the spacecraft to Earth. With the single case shown here it is evident that the physics underlying the increase in *Kp* can be directly linked to the SSC and precursor solar wind disturbance.

The improvement seen in the IRF-*Kp*-2017 and IRF-*Kp*-2017-h3 models over the IRF-*Kp*-2000 has several causes. In Table 11 the RMSE is 33% lower for IRF-*Kp*-2017-h3 compared to IRF-*Kp*-2000, and 41% lower for IRF-*Kp*-2017. The IRF-*Kp*-2000 model includes only the solar wind inputs B_z, *n*, *V*. Figure 2 shows the decrease of RMSE when further inputs are added, reducing the error with 9% with all inputs excluding the min/max inputs, and with 19% when the min/max inputs are included. Further, the 2000 model was trained on a smaller data set (about 10,000 samples) compared to the

2017 models (about 25,000 samples). Finally, the 2017 models consist of an ensemble of neural networks that further improve the accuracy. It should also be noted that a large fraction of the RMSE decrease is caused by the improvement for $Kp < 2$, and when only considering $Kp \geq 2$ the improvement over IRF-Kp-2000 is 12% and 21% for IRF-Kp-2017-h3 and IRF-Kp-2017, respectively (Fig. 5).

To extend the prediction lead time, but still using measured solar wind at L1, one can make forward projections based on the recent solar wind history reaching lead times up to 3–4 h (Boberg et al., 2000; Wing et al., 2005; Bala and Reiff, 2012). However, for events with solar wind disturbances causing strong SSC the onset of the Kp storm will be missed.

Much effort is also put into the prediction of the solar wind using numerical models that are driven by solar observations. The current operational model ENLIL (Odstrcil, 2003) predicts the solar wind speed and density, but not magnetic fields.[4] Within PROGRESS models are developed that will predict the solar wind plasma and vector magnetic fields at L1 using GONG solar magnetograms. At this stage only quasi-static features like coronal streamers and coronal holes are considered. The model consist of the AWSoM model (van der Holst et al., 2014) up to about 30 solar radii coupled with the Lare3d/SWIFT model (Arber et al., 2001; Arber, 2016) out to L1. As both plasma and vector magnetic fields will be predicted they can be used as inputs to the IRF-Kp-2017-h3 model to extend the prediction lead time to a couple of days. This will be tested within the PROGRESS project using predictions from the AWSoM/SWIFT model. However, current solar wind predictions are still quite crude and it will be interesting to see how close to 3-hour averages they will come.

From measured solar wind data at L1 the prediction lead time is determined by the time it takes the solar wind disturbance to reach the Earth's bow shock and magnetopause. Our approach is to use only the solar wind speed and the spacecraft-Earth distance to compute the lead time, the "flat delay". This is a robust approach that can be used on real-time solar wind data, but there are also several sources of errors. The "flat delay" assumes a planar structure orthogonal to the propagation direction along the spacecraft-Earth direction. Based on a set of 198 events in the solar wind, with a clear change in magnetic field direction within a minute, it was found that the "constrained minimum variance" method (Mailyan et al., 2008) achieved the best result for propagating discontinuities between two solar wind spacecraft. However, for both the "flat delay" and the "constrained minimum variance" methods about 90% of the events had timing errors less than 10 min, although the latter method has a larger fraction with errors close to 0 min. But, the method can only be applied when a clear discontinuity can be identified and data points in a time interval around the discontinuity must be used, decreasing further the lead time. Another issue is that the dynamical evolution from L1 to Earth is not considered, and in cases of shocks (Viñas and Scudder, 1986) there may be additional timing errors when using the bulk speed instead of the shock speed.

The two models presented here (IRF-Kp-2017 and IRF-Kp-2017-h3) react differently on sudden changes in the solar wind. The IRF-Kp-2017 model responds promptly as can be seen in Figure 7, while the IRF-Kp-2017-h3 shows a delay. Models relying on hourly averages (Bala and Reiff, 2012) will perform somewhere in between. Other averages could also be considered, but when an average is used it will affect the lead time. It is clear that the IRF-Kp-2017 model shows a greater degree of variability in its predictions as it responds to sudden changes in the solar wind. It should be noted that when the models are driven at 1-minute cadence the inputs still have 3-hour resolution and targets the 3-hour Kp giving 3-hour Kp predictions every minute. The performance measures in Tables 1 and 2 are based on one prediction per 3-hour interval, but not necessarily the best prediction, just the prediction that happen to coincide with the official 3-hour interval. As there exist no 1-minute Kp it is not possible to compute statistics for the model driven by 1-minute data. However, as is seen from the specific event in Figure 7 the IRF-Kp-2017 predictions are on average closer to the observed Kp as compared to IRF-Kp-2017-h3 predictions. It would be interesting in a future study to construct a series with high-resolution Kp against which the IRF-Kp-2017 predictions can be compared.

It is interesting to see that the magnitude of the solar wind magnetic field B has a large influence (Fig. 2) and that the importance of B_y is very small when B is used. In the work by Borovsky and Birn (2014) it is argued that it is not the electric field E, and thereby (B_y, B_z), that is at work in solar wind-magnetosphere coupling. Instead, when they derive the reconnection rate they find a relation containing the magnitude B. Our results seems to be in line with this.

In Table 4 it is seen that the prediction accuracy drops when using ACE RT data instead of ACE L2 data, in particular it is the RT density data that have the largest errors. In some cases the errors are due to short spikes (1–2 min duration) which can be efficiently removed using a 5-minute median filter, but which also reduces the prediction lead time with 2 min. For the IRF-Kp-2017 model the RMSE drops from 0.65 to 0.59 when the median filter is applied on the RT plasma data. However, to compare with the L2 predictions the RT samples are chosen from the same timestamps when L2 data exist and now the statistics are quite close. But, in real-time operation the additional information provided in the L2 set cannot be used, therefore, it is more reasonable to assume an RMS error of 0.59 when the median filter is applied, whereas the evaluation on the L2 set gives a theoretical limit of the model.

As noted above there are differences between the ACE RT and ACE L2 data that will degrade the accuracy of the Kp predictions. It is also seen that in the RT data set, and only using records with plasma status flag equal to zero, a large number of plasma density measurements have been removed after the processing to the science level L2 set for the two years analysed. It is difficult to detect these erroneous measurements in a real-time situation, although the 5-minute median filter removes all short lived spikes. In situations when real-time data is missing, e.g. for the ACE spacecraft during strong proton events, it is clear that the Kp predictions will not work, however, when there are incorrect measurements the models will produce Kp predictions although the accuracy will be lower, see rows 3 and 7 in Table 4.

---

[4] http://www.swpc.noaa.gov/products/wsa-enlil-solar-wind-prediction.

In real-time operation both IRF-*Kp*-2017 and IRF-*Kp*-2017-h3 can be driven with a cadence determined by the solar wind input data. The real-time ACE or DSCOVR data are provided with a 1-minute cadence and hence the predicted *Kp* can be provided once per minute. Due to the construction of the prediction algorithm the predicted *Kp* is still valid over 3-hour intervals but sliding with the input cadence. For example, a prediction given at 13:00UT with lead-time $\tau = 55$ min will provide a *Kp* prediction valid for the interval [10:55UT, 13:55UT]. From a practical perspective the published predicted *Kp* can be given for the [12:00UT, 15:00UT] official *Kp* interval. The next minute this interval will be shifted by 1-minute plus any change in lead-time. This means that published prediction for the last 3-hour *Kp* interval will possibly change until all predicted *Kp* values belong to the next interval.

The new models (IRF-*Kp*-2017 and IRF-*Kp*-2017-h3) will be implemented for real time operation using the DSCOVR solar wind data. As the IRF-*Kp*-2000 forecast and GFZ nowcast *Kp* are already implemented comparisons between the different implementations can be made in real-time operation. The first estimate of *Kp* is available 100 min into the 3-hour interval and then continuously updated for many hours after the corresponding time interval as the later data will affect the determination of the quiet curve. In the coming years it will be interesting to track the prediction accuracy using DSCOVR and to compare with ACE predictions. The prediction lead time will vary between 20–90 min and the predictions can be issued once per minute.

References

Arber T. 2016. Propagation of the solar wind from the Sun to L1. In: 13th European Space Weather Week, Oostende, Belgium.

Arber T, Longbottom AW, Gerrard CL, Milne AM. 2001. A staggered grid, Lagrangian-Eulerian remap code for 3-D MHD simulations. *J Comput Phys* **171** (1): 151–181.

Ayala Solares JR, Wei H-L, Boynton RJ, Walker SN, Billings SA. 2016. Modelling and prediction of global magnetic disturbance in near-Earth space: a case study for *Kp* index using NARX models. *Space Weather* **14** (10): 899–916.

Bala R, Reiff P. 2012. Improvements in short-term forecasting of geomagnetic activity. *Space Weather* **10**: S06001. DOI: 10.1029/2012SW000779.

Bala R, Reiff P. 2014. Validating the Rice neural network and the Wing *Kp* real-time models. *Space Weather* **12**: 417–425. DOI: 10.1002/2014SW001075.

Bartels J, Heck NH, Johnston F. 1939. The three-hour-range index measuring geomagnetic activity. *J Geophys Res* **44**: 411–454. DOI: 10.1029/TE044i004p00411.

Boberg F, Wintoft P, Lundstedt H. 2000. Real time *Kp* predictions from solar wind data using neural networks. *Phys Chem Earth C: Solar Terr Planet Sci* **25**: 275–280.

Borovsky JE, Birn J. 2014. The solar wind electric field does not control the dayside reconnection rate. *J Geophys Res: Space Phys* **119**: 751–760. DOI: 10.1002/2013JA019193.

Brautigam D, Albert J. 2000. Radial diffusion analysis of outer radiation belt electrons during the October 9, 1990, magnetic storm. *J Geophys Res: Space Phys* **105** (A1): 291–309.

Bruinsma S 2015. The DTM-2013 thermosphere model. *J Space Weather Space Clim* **5**: A1. DOI: 10.1051/swsc/2015001.

Chollet F. 2017. Keras: deep learning for Python. URL: https://github.com/fchollet/keras.

Cliver EW, Kamide Y, Ling AG, Yokoyama N. 2001. Semiannual variation of the geomagnetic Dst index: evidence for a dominant nonstorm component. *J Geophys Res* **106**(A10): 21297–21304.

Cybenko G. 1989. Approximation by superposition of a sigmoidal function. Math Control *Signals Syst* **2**: 303–314.

Heilig B, Lühr H. 2013. New plasmapause model derived from CHAMP field-aligned current signatures. *Ann Geophys* **31**: 529–539. DOI: 10.5194/angeo-31-529-2013.

Hunter JD 2007. Matplotlib: a 2D graphics environment. *Comput Sci Eng* **9**: 90–95.

Ji EY, Moon YJ, Park J, Lee JY, Lee DH. 2013. Comparison of neural network and support vector machine methods for *Kp* forecasting. *J Geophys Res* **118**: 5109–5117. DOI: 10.1002/jgra.50500.

Lundstedt H, Gleisner H, Wintoft P. 2002. Operational forecasts of the geomagnetic Dst index. *Geophys Res Lett* **29**(24): 34-1–34-4. DOI: 10.1029/2002GL016151.

Mailyan B, Munteanu C, Haaland S. 2008. What is the best method to calculate the solar wind propagation delay? *Ann Geophys* **26**: 2383–2394.

Mayaud PN. 1980. Derivation, meaning, and use of geomagnetic indices. *Geophysical monograph*, Vol. **22**. American Geophysical Union.

McComas DJ, Bame SJ, Barker P, Feldman WC, Phillips JL, Riley P, Griffee JW. 1998. Solar wind electron proton alpha monitor (SWEPAM) for the Advanced Composition Explorer. *Space Sci Rev* **86**: 563–612

McKinney W. 2010. Data structures for statistical computing in Python. In van der Walt S, Millman J, eds.Proceedings of the 9th Python in Science Conference, pp. 51–56.

Menvielle M, Papitashvili N, Häkkinen L, Sucksdorff C. 1995. Computer production of K indices: review and comparison of methods. *Geophys J Int* **123**: 866–886.

Murphy AH. 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Am Meteorol Soc* **8**: 281–293.

Murphy K, Mann I, Rae I, Sibeck D, Watt C. 2016. Accurately characterizing the importance of wave-particle interactions in radiation belt dynamics: the pitfalls of statistical wave representations. *J Geophys Res: Space Phys* **121** (8): 7895–7899.

Odstrcil D. 2003. Modeling 3-D solar wind structure. *Adv Space Res* **32**: 497–506. DOI: 10.1016/S0273-1177(03)00332-6.

Orlova K, Shprits Y. 2014. Model of lifetimes of the outer radiation belt electrons in a realistic magnetic field using realistic chorus wave parameters. *J Geophys Res: Space Phys* **119** (2): 770–780.

Orlova K, Shprits Y, Spasojevic M. 2016. New global loss model of energetic and relativistic electrons based on Van Allen Probes measurements. *J Geophys Res: Space Phys* **121** (2): 1308–1314.

Pulkkinen TI, Palmroth M, Tanskanen EI, Ganushkina NY, Shukhtina MA, Dmitrieva NP. 2007. Solar wind-magnetosphere coupling: a review of recent results. *J Atmos Solar-Terr Phys* **69**: 256.

Qian L, Burns AG, Emery BA, Foster B, Lu G, Maute A, Richmond AD, Roble RG, Solomon SC, Wang W. 2014. The NCAR TIE-GCM. In Huba J, Schunk R, Khazanov G, eds. Modeling the ionosphere-thermosphere system. Geophysical monograph series, Vol. **201**. John Wiley & Sons Ltd., Chap. 7. DOI: 10.1002/9781118704417.ch7.

Segarra A, Nosé M, Curto JJ, Araki T. 2015. Multipoint observation of the response of the magnetosphere and ionosphere related to the sudden impulse event on 19 November 2007. *Space Weather Space Climate* **5**: A13. DOI: 10.1051/swsc/2015016.

Shprits Y, Thorne R, Friedel R, Reeves G, Fennell J, Baker D, Kanekal S. 2006. Outward radial diffusion driven by losses at magnetopause. *J Geophys Res: Space Phys* **111**(A11): A11214.

Shprits Y, Thorne R, Reeves G, Friedel R. 2005. Radial diffusion modeling with empirical lifetimes: comparison with CRRES observations. *Ann Geophys* **23**: 1467–1471.

Smith CW, L'Heureux J, Ness NF, Acuna MH, Burlaga LF, Scheifele J. 1998. The ACE magnetic fields experiment. *Space Sci Rev* **86**: 611.

Stone E, Frandsen A, Mewaldt R, Christian E, Margolies D, Ormes J, Snow F. 1998. The advanced composition explorer. *Space Sci Rev* **86**: 1–22. DOI: 10.1023/A:1005082526237.

Thomsen MF. 2004. Why *Kp* is such a good measure of magnetospheric convection. *Space Weather* **2**: S11004. DOI: 10.1029/2004SW000089.

Tsyganenko N. 1989. A magnetospheric magnetic field model with a warped tail current sheet. *Planet Space Sci* **37** (1): 5–20.

Tsyganenko N, Sitnov M. 2005. Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *J Geophys Res: Space Phys* **110**(A3): A03208.

Tsyganenko N, Sitnov M. 2007. Magnetospheric configurations from a high-resolution data-based magnetic field model. *J Geophys Res: Space Phys* **112**(A6): A06225.

Tu W, Cunningham G, Chen Y, Henderson M, Camporeale E, Reeves G. 2013. Modeling radiation belt electron dynamics during GEM challenge intervals with the DREAM3D diffusion model. *J Geophys Res: Space Phys* **118** (10): 6197–6211.

van der Holst B, Sokolov IV, Meng X, Jin M, Manchester WB, Toth G, Gombosi TI. 2014. Alfven Wave Solar Model (AWSoM): coronal heating. *Astrophys J* **728**(2): 81. DOI:10.1088/0004-637X/782/2/81.

Viñas AF, Scudder JD. 1986. Fast and optimal solution to the "Rankine-Hugoniot Problem". *J Geophys Res* **91**(A1): 39–58.

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. DOI: 10.1002/2015SW001251.

Wing S, Johnson JR, Jen J, Meng CI, Sibeck DG, Bechtold K, Freeman J, Costello K, Balikhin M, Takashi K. 2005. Kp forecast models. *J Geophys Res* **110**: A04203. DOI: 10.1029/2004JA010500.

Wintoft P, Wik M, Viljanen A. 2015. Solar wind driven empirical forecast models of the time derivative of the ground magnetic field. *J Space Weather Space Clim* **5**: A7. DOI: 10.1051/swsc/2015008.

Zhang Y, Paxton LJ. 2008. An empirical Kp-dependent global auroral model based on TIMED/GUVI FUV data. *J Atmos Solar-Terr Phys* **70** (8–9): 1231–1242. DOI: 10.1016/j.jastp.2008.03.008.