PRediction Of Geospace Radiation Environment and Solar wind parameterS

Work Package 3
Forecast of the evolution of geomagnetic indices

Deliverable 3.1
Survey of existing operational models forecasting Kp, Dst, and AE

Peter Wintoft, Magnus Wik, Simon Walker, Michael Balikhin, Vitaly Yatsenko, Natalia Ganushkina
February 29, 2016

Document Change Record

Issue	Date	Author	Details
v1	Feb. 16, 2015	P. Wintoft	Initial draft
v2	Mar. 27, 2015	P. Wintoft, S. Walker	Updated draft
v3	April 7, 2015	P. Wintoft, S. Walker, M. Wik, N. Ganushkina	Updated draft
v4	April 10, 2015	P. Wintoft, S. Walker	Final version
v5	Feb 29, 2016	P. Wintoft	Minor updates according to comments by reviewer.

Contents

Summary

The overall aim of WP 3 concerns improvement and new development of models based on data driven modelling, such as CNN and NARMAX. Existing models for Dst and Kp will be analysed and verified with the aim of finding weaknesses and to suggest improvements. Solar wind and geomagnetic indices shall also be analysed in order to develop models for the identification of features, such as (but not limited to) shocks, sudden commencements, and substorms. Such categorisation will aid the model development and verification, and can also serve as alternative approach to models providing numerical input-output mapping. In addition to the development of Dst and Kp models new models will be developed to forecast AE. The models will be implemented for real-time operation at IRF and data and plots will be provided on a web server.

This deliverable is targeted to identify existing operational Kp, Dst, and AE forecast models. The models are analysed regarding their respective requirements and benefits considering, e.g. inputs, latency, lead time, and resources. A subset of the identified models that are available to the team will be verified in D3.3

1   Introduction

The coupling from the interplanetary medium through the magnetopause into the magnetosphere is non-linear with respect to the driving plasma and magnetic fields. This has been know for a long time, e.g. the energy input through reconnection is to a first approximation related to a ramp function of the upstream electrical field. Further, the dynamic evolution of the magnetospheric state is both driven by the external interplanetary medium and internal processes. Various aspects of the magnetospheric state can to some extent be described with geomagnetic indices, as originally devised. However, indices have come into use as space weather indicators for disturbed or hazardous conditions, and as inputs to various empirical and physical models of the magnetosphere and ionosphere. This report tries to summarise the current empirical models, with emphasis on operational models, for the coupling from the interplanetary medium as measured at L1 to the geomagnetic indices Kp, Dst, and AE.

Geomagnetic indices, and indices in general, "aims at giving summarised information in a continuous way concerning a more or less complex phenomenon which varies with time" (Mayaud 1980). The different geomagnetic indices are constructed in ways to provide summary information of different geophysical processes. The indices studied here are the ring current index Dst, global range index Kp, and the auroral electrojet index AE. These indices are commonly used to characterise different aspects of space weather and for indication of various degree of disturbances on technological systems. The indices are also used as inputs to empirical or physics based models that describes some space weather regime.

The aim here is not to study the processes that are captured in the indices, but to study ways of predicting the indices. The work is further constrained to only study models that are driven by upstream interplanetary plasma and magnetic filed data, and possibly the indices themselves. The interplanetary data and indices are described in more detail

in technical note D3.2.

The indices are used as inputs to numerical geophysical models. In PROGRESS the Inner Magnetosphere Particle Transport and Acceleration Model (IMPTAM) (Ganushkina et al. 2001, 2005, 2006, 2012) for low energy electrons in the inner magnetosphere (Ganushkina, Liemohn, Amariutei & Pitchford 2013, Ganushkina, Amariutei, Shpritz & Liemohn 2013) now operating it online at http://imptam.fmi.fi. IMPTAM computes the distribution of low energy electrons (10-150 keV), which is critically important for radiation belt dynamics. This seed population is further accelerated to MeV energies by various processes. The electron flux at these energies is largely determined by convective and substorm-associated electric fields and varies significantly with geomagnetic activity driven by the solar wind. Inward electron transport includes also radial diffusion and excites plasma wave instabilities that give rise to local electron acceleration and electron precipitation into the atmosphere. It should be noted that the electron flux at these energies is important for surface charging. At present, the model is driven by the following parameters provided in real time:

- 1 minute resolution data of solar wind number density, total plasma bulk velocity and solar wind dynamic pressure (used to calculate the electric and magnetic fields within the magnetosphere, the resulting particle motion, and to define the boundary conditions in the plasmasheet - the source of the magnetospheric electrons);

- 1 minute resolution data of the Interplanetary Magnetic Field (IMF) comprising of three components in GSM coordinates (used in models as previous bullet);

- hourly values of the Dst index (used in the magnetospheric magnetic field models);

- 3-hour Kp index (used in empirical parameterizations for electron lifetimes for representation of electron losses);

- 1 minute AE index (used to determine timings to launch electromagnetic pulses in the magnetosphere to reproduce substorm changes for additional electron acceleration).

IMPTAM is run only in near-real time. If the driving parameters are predicted, the model can provide corresponding forecast of low energy electrons.

2   Overview of existing models

2.1   Solar wind data

For the model development the solar wind data originates either from the OMNI dataset or directly from individual spacecraft (WIND, ACE, etc.). The solar wind data in the OMNI set consist of data from several different spacecraft that have been advected to a location close to Earth. When original spacecraft data are used for development the spacecraft–Earth propagation delay will be present.

In real time operation all data come from the ACE spacecraft around the L1 location. Depending on which type of data that have been used for development the propagation delay needs to be handled.

2.2   Forecasting Kp

We have identified 3 models predicting Kp. The first two models have been implemented for real time operation and are described in more detail in Section 3.

The Boberg et al. (2000) model is a linear combination of two feed-forward neural networks with tapped delay line inputs, with the two networks specialising on quiet time and storm time periods, respectively. The inputs are 3-hour averages of solar wind density, speed, and IMF B_z. These models have been developed using the OMNIweb data set from the period 1976-1996. The storm level network has been trained on events that contain at least one value with $Kp \geq 5_-$ and the quiet level model on the complement dataset. The combined model consists of a weighted average of the outputs from the two models, with weights determined from the errors of the respective model as function of forecast Kp value. The models were developed to predict the Kp value for the following 3-hour interval based on solar wind data at Earth.

The Wing et al. (2005) models are feed-forward multilayered neural network with feed-back in the hidden layer, also known as an Elman recurrent neural network (Elman 1990). Three models have been developed based on linearly interpolated Kp values from the original 3 hour resolution to 15 minute resolution, and 15 minute running hourly average solar wind. The models predict Kp with approximately 1 to 4 hours lead time.

The Bala & Reiff (2012) models consist of a neural networks to predict Kp with lead times of 1, 3, and 6 hours. The inputs are solar wind speed, magnetic field magnitude, and magnetic field clock angle, i.e. the angle of the magnetic field in the $y - z$ plane using the OMNI set. The solar wind data are first transformed using the Boyle index

$$\Phi = 10^{-4}v^2 + 11.7B\sin^3\frac{\theta}{2} \tag{1}$$

and forming a vector of delayed inputs. The 1-hour forecast uses hourly averages of the Boyle index covering the last 8 hours and forecast a 1-hour "oversampled" Kp. It is not clear if the "oversampling" is linear interpolation or some other procedure. The 3-hour forecast uses 3-hour averages covering the last 21 hours. Te lengths of the delay lines were optimised through training and testing.

2.3   Forecasting Dst

There are an overwhelming number of papers that address the forecasting of Dst. This report does not provide a detailed description of each model, but tries instead to summarise the main points. Some of the models have been implemented for real time operation and are further described in Section 3.

The first class of models are empirical relations that map solar wind to Dst and originates from the first-order differential equation suggested by Burton et al. (1975)

$$\frac{dDst^*}{dt} = Q - \lambda Dst^*, \tag{2}$$

where Q is the injection rate controlled by the solar wind electric field and Dst^* is the pressure corrected Dst

$$Dst^* = Dst - b\sqrt{p} + c, \tag{3}$$

and $p \propto nv^2$ is the dynamic pressure. The constants λ, b, and c are determined from data. Variants of this model has later been suggested affecting both the injection rate (Fenrich & Luhmann 1998) and the decay term (O'Brien & McPherron 2000).

The model of Temerin & Li (2006) also belongs to this class, although this model consists of a very complex function including some 100 coefficients that are determined empirically. The details of the model are given in Section 3.

Another empirical model aims at predicting the minimum Dst instead of the detailed evolution (Wang, Shen, Wang & Ye 2003). For the study hourly averaged ACE solar wind data were used, and data gaps filled in with WIND data. The ACE-Earth propagation time were not considered. The focus was on intense storms at levels $Dst < -50$ nT and $Dst < -100$ nT. The empirical relation $Dst_{\min} = -19.01 - 8.43(-VB_z)^{1.09}(\Delta t)^{0.3}$ nT where Δt is the duration of the driving event, although it is not clear if this parameter can be unambiguously determined.

Another class of models are based on moving average (MA) filters or autoregressive moving average (ARMA) filters, or similar. The Vassiliadis et al. (1999) model is a nonlinear second-order ARMA filter that is driven by 5-minute averaged solar wind data. It was stated that the hourly resolution of Dst is too coarse to capture essential dynamics. This led Vassiliadis et al. (1999) to construct their own 5 minute resolution Dst index based on data from the same geomagnetic stations and smoothing the result with a 25 minute running average. The inputs to the model are the solar wind measured by the ISEE 3 spacecraft around L1. The solar wind were ballistically propagated to the upstream magnetopause location and applying a 25-minute moving average. They also consider the approximate 25-minute delay in the Dst response identified by Burton et al. (1975). The model is driven by the solar wind pressure and the solar wind electric field VB_s, where $B_s = -B_z$ when $B_z < 0$ and 0 otherwise. However, the pressure correction terms also have a dependence on the level of solar wind activity.

The modelling with the use of neural network in space physics was probably introduced by Koons & Gorney (1991) for a model of geosynchronous electron flux and shortly followed by Lundstedt (1992). The neural network approach for Dst prediction has then been used in a large number of papers. Initially static networks (i.e. no recurrent feedback) with delayed inputs were developed (Lundstedt & Wintoft 1994, Gleisner et al. 1996). To better model the dynamic evolution of Dst, especially for the recovery phase, network with feed-back connections were used (Lundstedt et al. 2001, Watanabe et al. 2002, Pallocchia et al. 2006). These models are all driven by the solar wind plasma density, velocity and magnetic field components. The model by Pallocchia et al. (2006) also investigated the use of a single, magnetic field driver, motivated by the fact that the ACE plasma instrument outage during proton events. A large set of different coupling functions were also investigated by Wu & Lundstedt (1997). The Boyle index has also been used as input to a neural network, similar to the Kp model, forecasting Dst with lead times of 1 and 3 hours (Bala & Reiff 2012).

The nonlinear autoregressive moving average model with exogenous inputs (NARMAX) was applied to Dst by Boaghe et al. (2001). The model use hourly averages of solar wind data ballistically propagated from L1 to Earth and make 1-hour forecast of Dst. The inputs are VB_s, with the history of the past 10 hours, and past values of observed Dst going back 3 hours. Another approach, also using a NARMAX model, used

wavelet filtered VB_s as inputs (Wei et al. 2004). In this case the past 2 hours of VB_s and the past 4 hours of Dst were used. Another NARMAX model developed by Zhu et al. (2007) also included the solar wind pressure term and extended the prediction lead time to 5 hours. The NARMAX model by Boynton et al. (2011) suggests another coupling function according to $p^{1/2}V^{4/3}B\sin^6(\theta/2)$. An 11-year long series of hourly average data from the OMNI set were used.

Another type of model is a relevance vector machine (RVM) for one-hour forecast of Dst (Andriyas & Andriyas 2015). The RVM is a general regression technique based on Bayesian statistics. Hourly averages of solar wind data from the ACE and WIND spacecraft were used for the model development and evaluation. The solar wind magnetic field components B_y and B_z, plasma speed and flow direction (latitude and longitude), ion density, alpha to proton ration, temperature, dynamic pressure, plasma beta, and daily F10.7 flux were used as inputs. The forecast is Dst one hour ahead using the measurements at spacecraft location, i.e. no explicit propagation were performed and thus part of the lead time is due to the propagation delay.

Six different models are compared on 63 intense storms in the work by Ji et al. (2012), where intense is defined as having an event with at least one hour with $Dst < -100$ nT. In the study, all models were driven by hourly averaged ACE solar wind data and not considering the ACE-Earth propagation. They used four metrics for the comparison: linear correlation; RMSE; difference between observed minimum Dst and predicted minimum Dst; difference between time of observed minimum and time of predicted minimum. The result was that the Temerin & Li (2006) had the overall best performance, and that the same model was the best model for range $-200 < Dst < -100$ nT, while three models (Wang, Chao & Lin 2003, Temerin & Li 2006, Boynton et al. 2011) where comparable for $Dst < -200$ nT.

2.4   Forecasting AE

The prediction of $AE(>0)$, or $AU(>0)$ and $AL(<0)$, where $AE = AU - AL = AU + |AL|$, have been addressed in a number of papers, but currently only one model is operational.

The Goertz et al. (1993) model consists of three first-order differential equations, one for AU and two for AL, with 6 free parameters that are quantitatively determined from observed solar wind, AL, and AU data. The model is derived from basic physical reasoning. The input is the east-west convective electrical field $E_y = -VB_z$ modulated to be 0 for northward B_z, i.e. in principle B_s is used. At the time, only solar wind ISEE 3 data with 1-minute resolution IMF and 1-hour resolution velocity were available.

Hernandez et al. (1993) used two methods, a linear filter and a neural network, to make single step predictions of AL based on either solar wind only or solar wind and past values of AL. The temporal resolution was varied between 2.5 to 30 minutes. The solar wind input was the eastward electric field derived from VB_s.

Vassiliadis et al. (1995) used a non-linear filtering approach based on a set of locally linear MA filters or ARMA filters to predict AL. Ttwo variants were studied: using solar wind only (non-linear MA filter), and solar wind and past AL (non-linear ARMA filter). The filter coefficients were optimised for single step predictions using 2.5 minute resolution

data, and using past predicted AL the model was iterated to provide 4 hour prediction, assuming known solar wind data.

Gleisner & Lundstedt (1997) used nonlinear time-delayed feed-forward neural networks to predict AE. The input data were 5-min averaged solar wind data, V, n, and the IMF components B_y and B_z, from the spacecraft IMP 8 during 1973-1974. The neural networks consisted of one hidden layer, with 8 hidden nodes, and the input data were given as samples with a time window of 100 minutes. A comparison between individual solar wind variables and various coupling functions, such as e.g. VB_s, was also performed as well as a comparison to linear neural networks. There is no mention of lead time, but it is our understanding that they used time shifted AE data to the magnetopause as in Gleisner & Lundstedt (2001).

Gleisner & Lundstedt (2001) again used a neural network, but this time an Elman recurrent network. The solar wind data set is similar to the previous study, with V, n, and B_z, as well as as the AE but with a time resolution of 2.5 minutes. In this study, the AE data were time shifted, because of the solar wind travel time to the magnetopause from IMP 8. The network has 4 hidden nodes (and therefore 4 context nodes). Due to the dynamical behaviour of the recurrent network, no time delayed input data were needed, and therefore only 3 input nodes were used. A comparison was also made with the previous study.

Another neural network using time delayed Boyle index as inputs has been developed to forecast hourly average AE with 1 and 3 hours lead time (Bala & Reiff 2012).

Amariutei & Ganushkina (2012) used an ARMAX based neural network model to forecast the AL index. Input data consists of VB_s with 1 minute resolution, from the OMNI dataset for the whole year 1998, i.e. propagated to Earth. In this study, the emphasis was on accurate forecasts of substorm onsets, therefore the input data were smoothed using a moving average of 7 minutes. We assume that a higher temporal resolution is desirable but that the 7 minute average makes a good compromise between predictability and resolution. The exogenous inputs represent the past prediction errors. The network consists of two hidden layers, the first with seven nodes and a nonlinear activation function, and the second with three nodes and a linear activation function. Forecast lead times between 10 and 60 minutes were tested.

The AE model by Luo et al. (2013) is similar to the Temerin & Li (2006) model and consists of an empirical relation with a large number of terms. The model uses 10-minute averages of solar wind data and interpolated daily 10.7 flux, and predicts AU, AL, and AE with 10 minute lead time. The AL model builds in the model in Li et al. (2007) which also introduces an empirical time shift of about 7 minutes that is added to the lead time. The model is further described in Section 3.3.1.

3   Operational forecast models

A survey of currently available operational models that forecast the indices using L1 solar wind data has been carried out. The following models have been identified for each of the three indices. A short description of each model is provided and references are given.

3.1 Kp models

Two providers of Kp forecasts have been identified.

3.1.1 Swedish Institute of Space Physics

Two different models with different lead times are operational at IRF-Lund (Boberg et al. 2000) (Figure 1). The models use solar wind plasma and IMF as inputs. The models are feed-forward neural networks with tapped delay line inputs.

The inputs are 3-hour averages of solar wind density, speed, and IMF B_z. The models have been developed based on the OMNI dataset and data covering the years 1976–1996 were selected. The OMNI set consists of hourly averaged solar wind data that have been advected from the spacecraft location to just upstream of Earth.

Each model consists of two sub-models specialised for the prediction of quiet periods and storm periods, respectively. The storm level model has been trained on events that contain at least one value with $Kp \geq 5_-$ and the quiet level model on the complement dataset. The combined model consists of a weighted average of the outputs from the two models, with weights determined from the errors of the respective model as function of forecast Kp value.

The models were developed to predict the Kp value for the following 3-hour interval based on solar wind data at Earth. The models implemented for real-time operation do not specifically handle the 10–80 minutes additional lead time given by the ACE location at L1. The forecasts are always provided for the standard Kp intervals 00, 03, ..., 21. The models are run with a cadence of 1 hour which means that the forecast lead time varies between 0 and 2 hours based on the Kp timestamps, or 3 and 5 hours as Kp extends over three hours, and that the forecast Kp is only final for the 3-hour forecast. Another model is also implemented that makes a now-cast of the current Kp value.

3.1.2 Space Weather Prediction Center

Three different Kp forecasting models, with different inputs and lead times, are implemented at SWPC (Wing et al. 2005) (Figure 2). Each model is a feed-forward multilayered neural network with feed-back in the hidden layer, also known as an Elman recurrent neural network (Elman 1990).

The models have been developed based on linearly interpolated Kp values from the original 3 hour resolution to 15 minute resolution. The interpolation is unphysical and does not add any new information, however, it may be motivated from the aspect of having a more responsive system providing forecast with a higher cadence.

For the model development, the solar wind parameters are ballistically propagated from L1 to Earth using measured solar wind velocity. Then, the parameters are resampled to hourly running averages with a 15 minute cadence. As the solar wind input is propagated to Earth, the prediction lead times defined in the models below should be interpreted as the sum of L1–Earth propagation time and possibly a lead time due to magnetospheric processes or in statistical terms.

One model predicts Kp "approximately 1 hour ahead" using solar wind plasma, IMF, and nowcast Kp as inputs. The actual lead time is determined from the solar wind
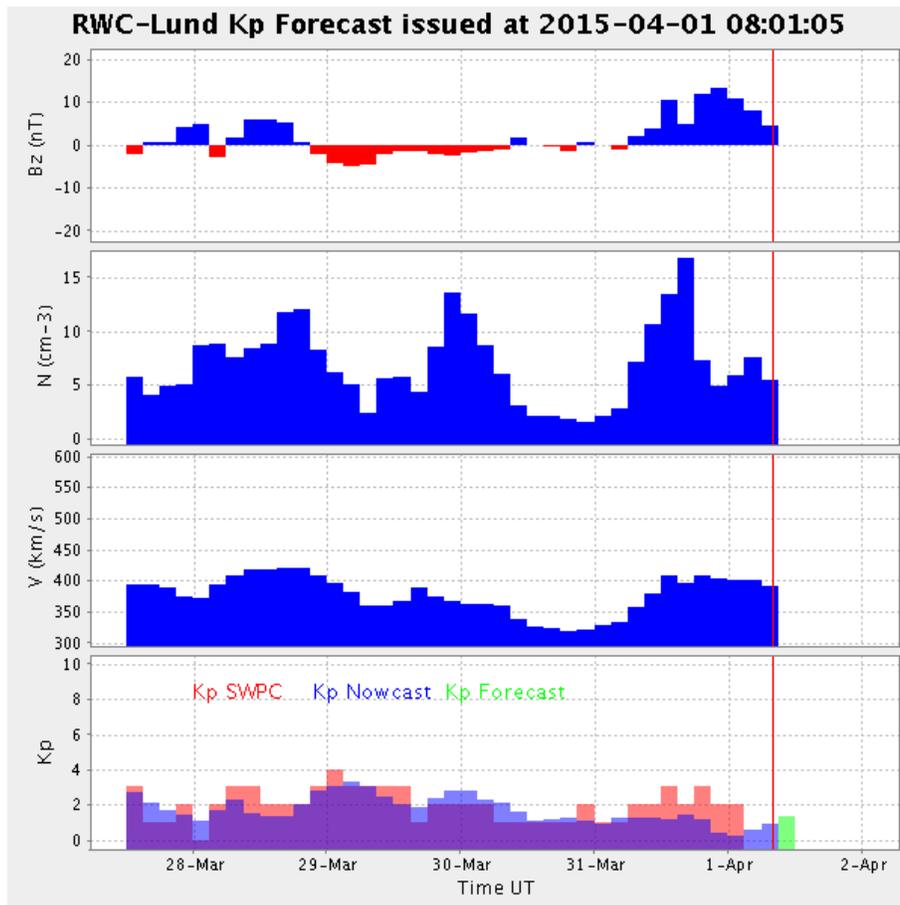
Figure 1: IRF-Lund Kp forecast at 2015-04-01 08:49 UT.

propagation time, thus the "1 hour" should be interpreted as anything between 10 minutes and 80 minutes. The nowcast Kp is derived from a few geomagnetic stations that provide near-realtime magnetic measurements (Takashi et al. 2001).

The next model is identical to the above except that the prediction lead time is approximately 4 hours, which should be interpreted as solar wind travel time plus 3 hours, thus in the range 3 hours and 10 minutes to 4 hours and 20 minutes. The stated performance of this model can in part be attributed to magnetospheric processes and in part to statistical correlations. This is quite visible in Figure 5(h) in Wing et al. (2005) where the forecast Kp lags behind observed Kp.

Finally, the last model use only solar wind as inputs, which is motivated by the possible lack of near-realtime Kp data. The model provides "1 hour ahead" forecast which we again interpret as 10-80 minutes lead time. From an evaluation point of view, this type of model also has the advantage that there is no autocorrelation feeding back as the target quantity is not used as input.
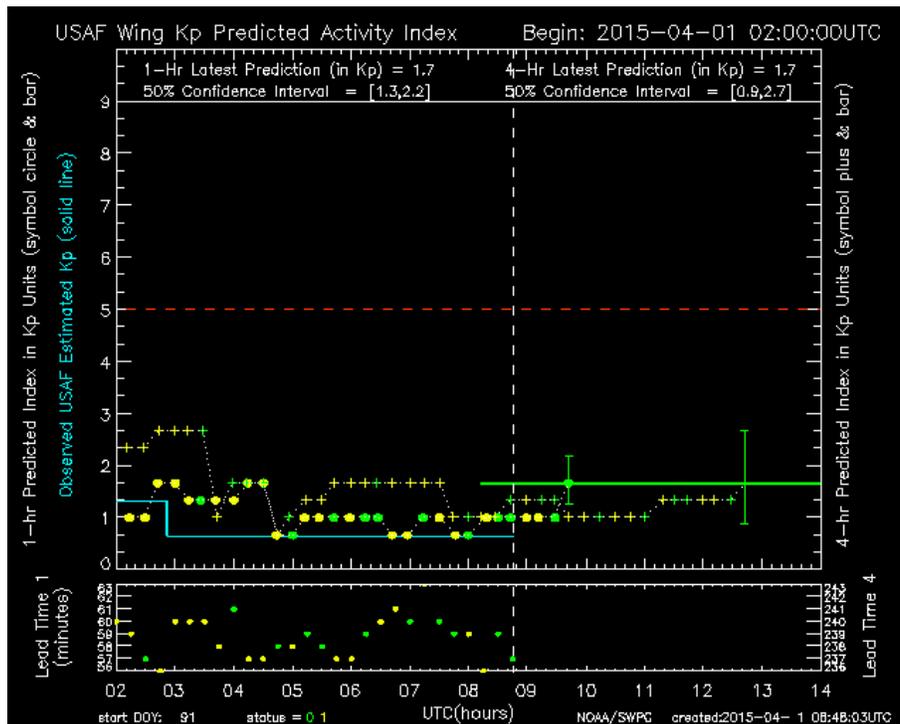

Figure 2: SWPC Kp forecast at 2015-04-01 08:49 UT.

3.2 Dst models

Four providers of Dst forecasts have been identified.

3.2.1   Swedish Institute of Space Physics

The IRF-Lund *Dst* forecast model consists of a feed-forward neural network with internal feed-back, an Elman network, driven by solar wind data (Lundstedt et al. 2001) (Figure 3).

The inputs are 1-hour averages of solar wind density, speed, and IMF B_z. For the model development the hourly average data from the OMNI data set were used, again noting that the in this set the solar wind data have been advected to Earth. The target output is *Dst*, also from the OMNI set, for the next hour.

In real-time operation the hourly averaged solar wind data are propagated to Earth using the solar wind speed and the average L1–Earth distance of $1.5 \cdot 10^6$ km, which provides additional lead time between 10 to 80 minutes. The models are run with a cadence of 10 minutes.
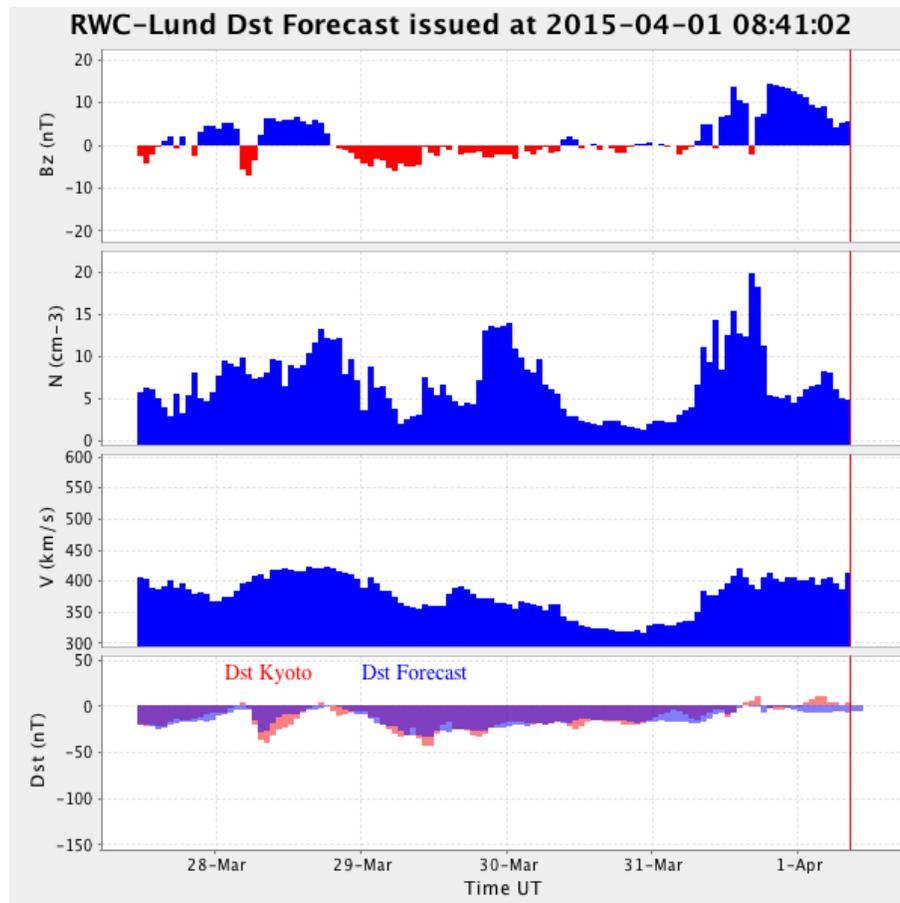

Figure 3: IRF-Lund *Dst* forecast at 2015-04-01 08:49 UT.

3.2.2   Laboratory for Atmospheric and Space Physics, USA

The *Dst* forecast model implemented at LASP consists of an empirically determined difference equation driven by solar wind data (Temerin & Li 2006).

The inputs are 10-minute average solar wind density, velocity component along Sun-Earth line, IMF B_y and B_z, and total magnetic field. Data come from the WIND and ACE spacecraft.

The model is a difference equation with a sum of 6 terms, each term consisting of non-linear expressions with estimated parameters. There are in total more than 100 free parameters and they are basically found through manual trial-and-error. The RMS error between model Dst and target Dst is minimised by individually changing each parameter by a small amount. However, it is not clear how the initial set of parameters were found.

In real-time operation 10-minute averages of ACE solar wind data are ballistically propagated to Earth using a constant ACE-Earth distance of $238.6R_E = 1.519 \cdot 10^6$ km, and the interpolated to 10-minute timestamps. The model is typically run three steps into future, i.e. 30 minutes. This suggests that the model has a prediction lead time of 40 to 110 minutes.
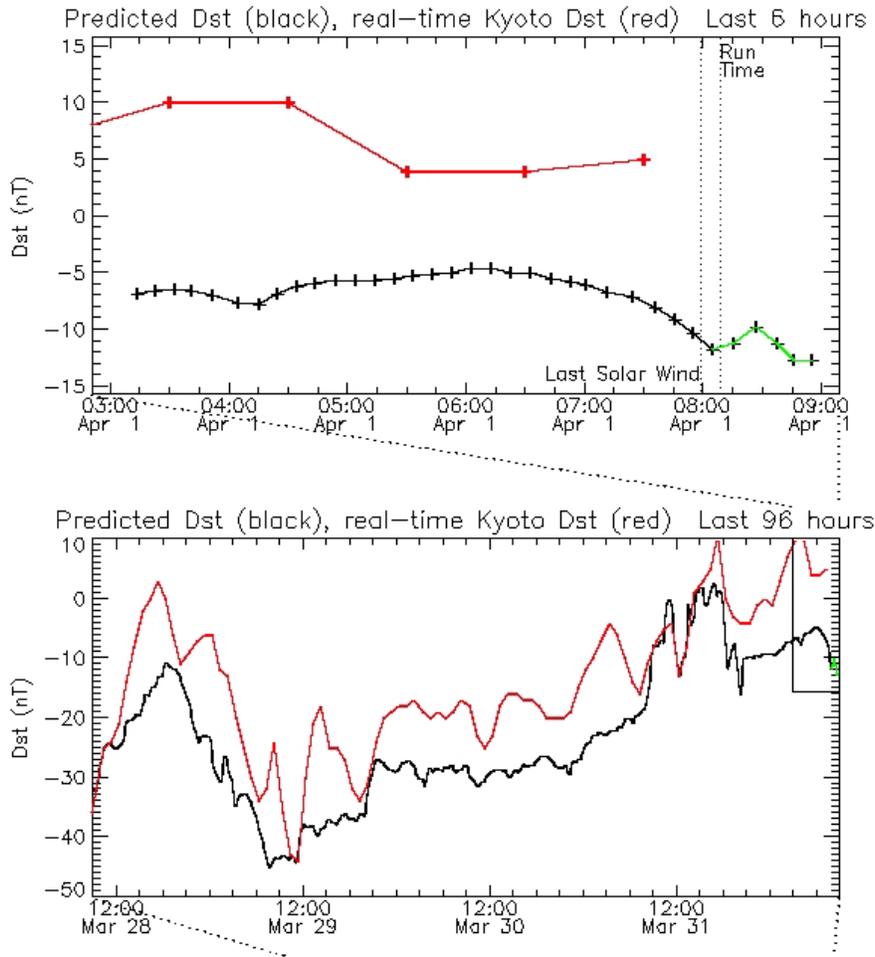

Figure 4: LASP Dst forecast at 2015-04-01 08:49 UT.

3.2.3 National Institute of Information and Communications Technology, Japan

The NICT operational model consist of an Elman neural network driven by ACE solar wind data (Watanabe et al. 2002).

Three different models have been tested with different inputs and training sets and the optimal model that was selected for operation have the following inputs: density, speed, IMF **B** vector and B magnitude and predicts two hours ahead. The two hours assumes that the ACE-Earth travel time is always one hour and that the magnetospheric response is one hour.

The implemented model is run with a cadence of one hour using 1-hour average ACE solar wind data.
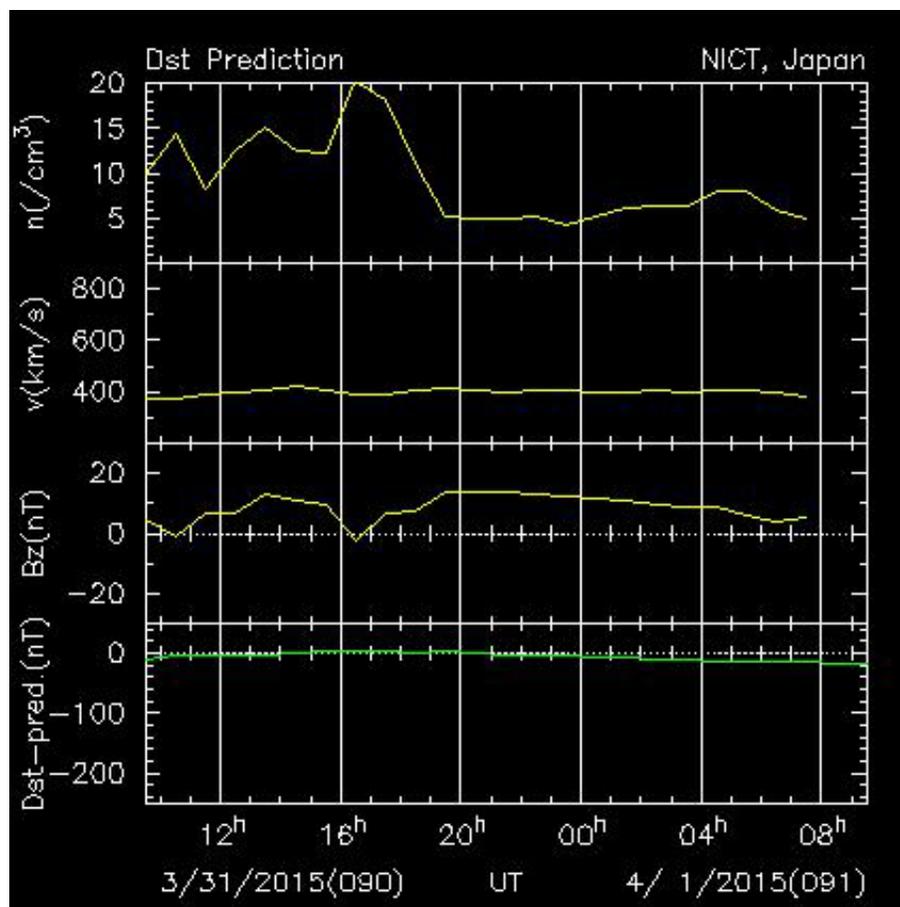

Figure 5: NICT *Dst* forecast at 2015-04-01 08:49 UT.

3.2.4 Moscow State University

The MSU model consist of a neural network with tapped delay line. The model is not described in one single paper but it is described at the web page and partly in Dolenko et al. (2014).

The inputs are hourly averages of ACE solar wind plasma density, speed, temperature, IMF B vector and magnitude B, and past values of Kyoto quick-look Dst.

The operational model runs with a cadence of 30 minutes a few minutes after the half hour or the full hour. The half-hour forecast is preliminary and is later replaced by the full-hour forecast. The stated forecast lead time is 0.5 to 1.5 hours and there is no handling of the ACE–Earth propagation.
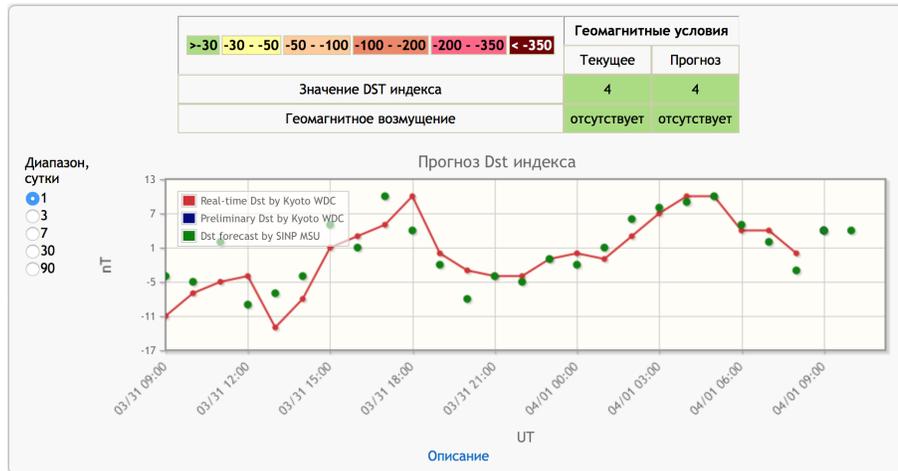

Figure 6: SINP Dst forecast at 2015-04-01 08:49 UT.

3.3 AE models

Only one model currently exists.

3.3.1   Laboratory for Atmospheric and Space Physics

The AE model consists of an empirically derived set of equations (Luo et al. 2013) that builds on a previously developed model for AL (Li et al. 2007) and is similar to the Dst model (Temerin & Li 2006).

The inputs are 10-minute averages of the solar wind density, speed, IMF **B** vector and magnitude B, and 10-minute interpolated values from the daily F10.7 index. The solar wind data come from the WIND and ACE spacecraft. For model development the solar wind data were propagated from the spacecraft location and then interpolated to 10 minute averages.

There are two models consisting of difference equations, each with a sum of 5 terms for AL and AU respectively. The indices are also 10-minute averages and the forecast is made for the next time stamp, i.e. a 10-minute forecast. The outputs from the two models are then combined into $AE = AU - AL$. After comparison with the measured ALindex, a time delay of 7.30 minutes were included to the model output. There is about 100 free parameters in each model that are empirically found through a manual procedure.
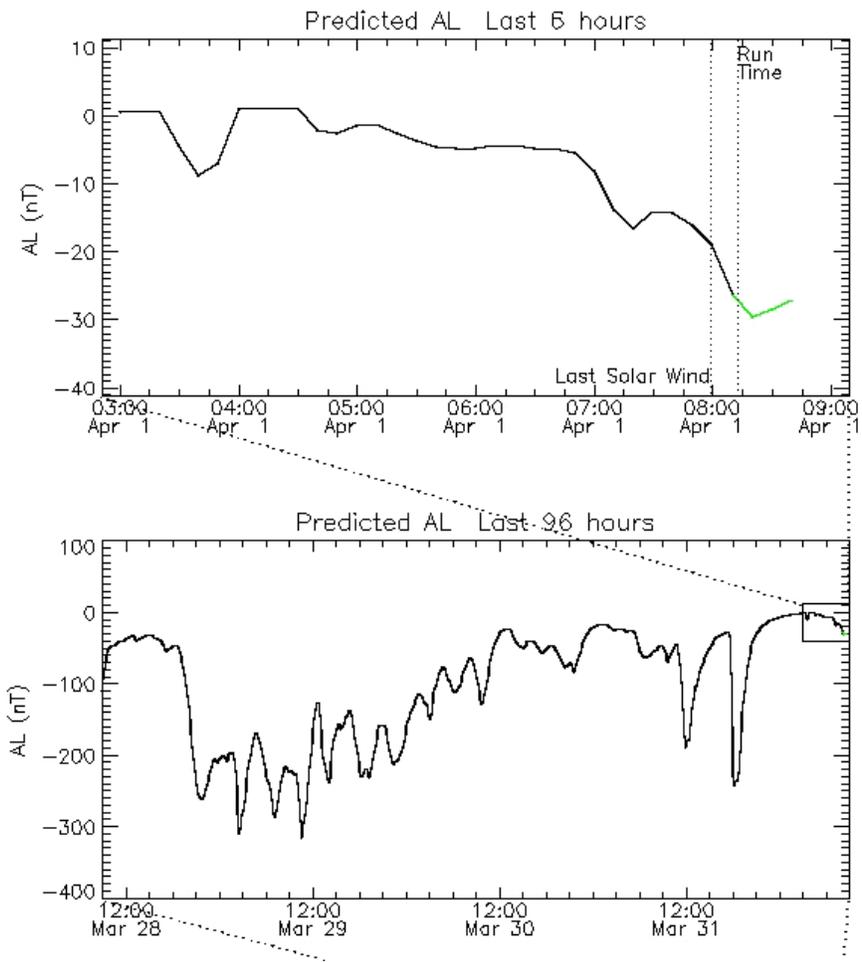
Figure 7: LASP *AE* forecast at 2015-04-01 08:49 UT.

4 Discussion

4.1 Verification and model selection

From the survey carried out in this report it is not possible to select the "best" model, or models, based on the published metrics that have been used for verification. There are several reasons for this. Firstly, there are a large number of metrics that can be used, several of them overlapping in what they measure, and the interpretation is not always clear. The report by Wintoft et al. (2011) contains an overview of metrics and approaches applied in space weather. There is also no theoretical guidance on how to select appropriate metrics, although the series of papers by Murphy & Winkler (1987), Murphy (1988), Murphy et al. (1989), Murphy (1993, 1996), Murphy & Wilks (1998) provides a deeper understanding of the problem. Secondly, it has to be very clearly defined what a model tries to forecast and develop a metric that captures that aspect. E.g., the *Dst* index goes typically through three phases during a storm: pressure related initial phase, reconnection related main phase, and the recovery phase. Each phase has different dynamics with different statistical distributions. A metric containing all phases may provide good scoring for a model that mainly captures the recovery phase although other models may be better at the main phase. The Geospace Environment Modeling (GEM) challenge for *Dst* (Rastätter et al. 2013) compares empirical models and MHD models using four different metrics. The models were run to predict *Dst* for for different time intervals, each containing storms at different levels. Depending on choice of metric, the models were ranked different, illustrating both the problem of interpretation of metric and that different models may perform better for certain phenomena. Another issue is the choice of events which may bias the model selection.

4.2 Sampling and timestamps

All datasets used here, whether it is solar wind data or indices, are time series data, i.e. each value has an associated time stamp. At the same time, each value has been determined over some time interval Δt and there can be an ambiguity in whether the time stamp represents e.g. the beginning, the centre, or the end of the time interval. However, in most cases the time stamps mark the beginning of each interval, and this will also be used here. This is important to keep in mind for both real time monitoring and forecasts. E.g., even if the measurement or derivation of some quantity can be made in real time, the timestamps will lag between Δt and $2\Delta t$ from real time. We define this as the sampling lag

$$\tau_s \in [\Delta t, 2\Delta t]. \tag{4}$$

The three indices capture phenomena at different time scales. The official *Kp* index has a temporal resolution of three hours and is defined for the intervals 00-03, 03-06, 06-09, 09-12, 12-15, 15-18, 18-21, 21-24 UT. For a time series of *Kp* values each value is associated with timestamps 0, 3, 6, 9, 12, 15, 18, and 21 according to the definition above, and thus has a sampling lag of $\tau_s \in [3, 6]$ hours.

The official Dst index has a resolution of $\Delta t = 1$ hour with timestamps at 00, 01, ..., 23 UT and thus a sampling lag of $\tau_s \in [1, 2]$ hours.

The official AE index has a resolution of $\Delta t = 1$ minute with timestamps at 00:00, 00:01, ..., 23:59 UT and thus a sampling lag of $\tau_s \in [1, 2]$ minutes.

Naturally, the sampling lag also affect the interpretation of timestamps of the input data. Thus, a model that provides forecasts in real time will have input data that lags behind real time with the sampling lag, if no other lags are present, and this needs to be explained in a real time service.

A forecast model provides an estimate of the target data with a lead time of τ_l. If the input and target sample intervals are equally long (Δt), and if the forecasts are issued at the timestamps of the observations, then the timestamp of the forecast will be $\tau_l - \Delta t$ ahead of real time, and the forecast quantity will be valid for the interval $[\tau_l - \Delta t, \tau_l]$. If the forecasts are issued at other times the valid time interval will be $[\tau_l - \tau_s, \tau_l - \tau_s + \Delta t]$. If the input sample interval Δt_{in} and output sample interval Δt_{out} are not equal then the forecast interval becomes $[\tau_l - \tau_{s,\text{in}}, \tau_l - \tau_{s,\text{in}} + \Delta t_{\text{out}}]$. There may also be further lags introduced if the forecast quantity is forced to certain timestamps, like e.g. the three-hour timestamps 00, 03, ..., 21 of Kp.

Another issue is that any filtering applied to the input data will increase Δt_{in} and affect the forecast lead time, something that may be missed during model development and testing.

4.3   Prediction lead time

Plasma and magnetic field data measured at L1 offers a forecast lead time between 10 to 80 minutes for a range of speeds between 300 and 2000 km/s before the disturbance reach the magnetosphere. Using minimum and maximum distances to ACE of $1.4 \cdot 10^6$ to $1.6 \cdot 10^6$ km and magnetopause location between 10 to 20 R_E gives the travel time as function of speed shown in Figure 8. It is often quoted in the literature that there is about a one-hour lead time from the L1 location. However, as the figure shows, this is only true for speeds, in the best case, less than about 500 km/s, but more typically around 400 km/s.
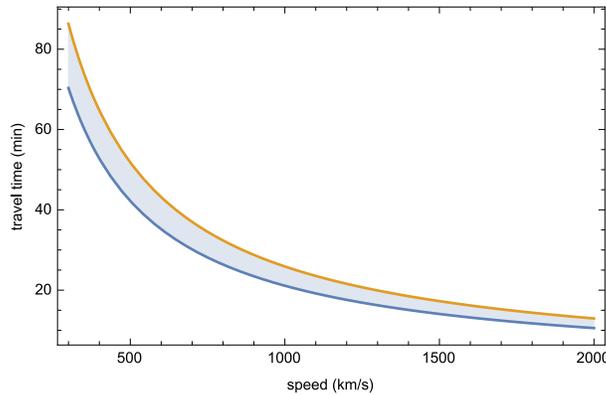

Figure 8: ACE–magnetopause travel time as function of speed.

As the indices are global there is no local time effect that can be exploited for additional lead time. When the interplanetary disturbance reach the magnetopause all indices will react immediately. However, the indices respond differently.

The Kp index is a weighted measure of the range of variation of the horizontal component of the magnetic field (K index) collected from several globally distributed stations. The physics generating the disturbance is not revealed in the Kp index as different phenomena like sudden impulses, pulsations, and substorms are lumped together.

The Dst index is more targeted to a specific magnetospheric system than the Kp index, although it has contributions from several different phenomena. The start of a Dst storm usually starts with an increase during the initial phase associated with magnetopause current caused by the direct compression from the solar wind. Then there is a short delay for the ring current to build up causing the negative main phase. Then, if no more disturbances arrive in the solar wind, the ring current decays with a gradual increase of Dst to zero-level over approximately 12–24 hours. The different phenomena are associated with different prediction lead times.

The AE index has the highest temporal resolution and monitors activity with contributions mainly from auroral latitudes. Similar to Kp, AE is a range index formed by the difference between AU and AL, and thus is non-negative. The temporal resolution enables the identification of sudden impulses, pulsations, and substorms. As before, the sudden impulses follow immediately after the solar wind disturbance hits the magnetopause while there can be a shorter delay before the substorm sets in. As soon as the substorm is over AE falls to low values.

Generally, higher levels of geomagnetic activity are associated with high solar wind speed. Figure 9 shows the estimated CDF for 3-hour maximum solar wind speed simultaneous with $Kp \geq 7$. About 50% of the samples are associated with speeds greater than 600 km/s, indicating propagation lead time of 40 minutes or less.
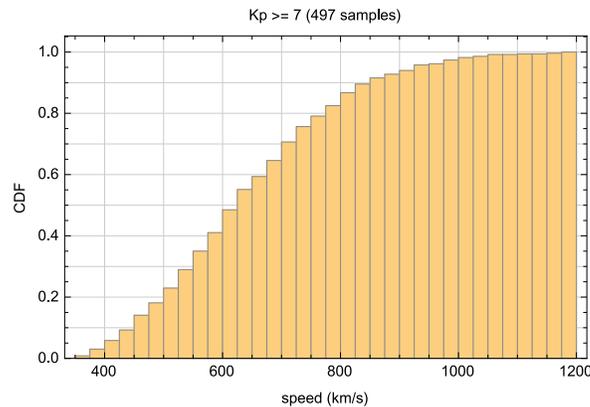

Figure 9: Estimated cumulative density function (CDF) as function of solar wind speed for $Kp \geq 7$.

5   Conclusions

There are two broad categories of models, models based on simplified physical reasoning and models based on data feature extraction algorithms. Naturally, the division is not clear-cut and there is a mix of both approaches in both categories. The first approach provides guidance to the parameter selection of the second approach, while the second approach provides the possibility of relations that more exactly describe the solar wind–magnetosphere mapping.

The models can be further categorised into two classes, depending on whether they use time-lagged target parameter as input or not. As the L1–Earth forecast models only provide short-term forecasts, in the best case a few hours, it is important to properly evaluate the effect of the autocorrelation in the target time series when the models are validated. A subset of the identified models that are available to the team will be verified in D3.3.

The actual prediction lead times for published models are not always clearly defined. Also, the lead time needs to be physically motivated and understood. From solar wind measurements around L1, a lead time of 10-80 minutes is possible due to the propagation. Another few minutes will be present from the bow shock–magnetopause propagation. Then the magnetospheric substorm and storm processes will add another ≈ 30 minutes. There are also local time effects that can provide additional lead time, however, as the considered indices are global that effect can not be exploited. Longer lead times are not physically comprehensible but may work under the assumption that nothing changes in the solar wind.

All models that rely on both ACE plasma and magnetic field data will fail during solar storms associated with high proton fluxes as the plasma instrument becomes inoperational. Several of the largest geomagnetic storms occur during those events. It will be interesting to see how the DSCOVR spacecraft will operate during those conditions.

Several of the models provide "preliminary" forecasts, which means that the forecast value for a certain timestamp in the future may change when the timestamp moves to the past. This means that plots with observed and forecast data will look better than they actually are.

A crucial point is how the models will behave during extreme events. One problem is of course related to the ACE plasma instrument outage during proton events, which hopefully will be solved with the DSCOVR spacecraft. However, even under the assumption that the solar wind data will be correct the models may fail due to the fact that they may be operating in a domain they were not developed for.

References

Amariutei, O. A. & Ganushkina, N. Y. (2012), 'On the prediction of the auroral westward electrojet index', *Annales Geophysicae* **30**, 841–847.

Andriyas, T. & Andriyas, S. (2015), 'Relevance vector machines as a tool for forecasting geomagnetic storms during years 1996-2007', *Journal of Atmospheric and Solar–Terrestrial Physics* pp. 10–20.

Bala, R. & Reiff, P. (2012), 'Improvements in short-term forecasting of geomagnetic activity', *Space Weather* **10**, S06001.

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', *Journal of Geophysical Research* **106**(A12), 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, Part C: Solar, Terrestrial & Planetary Science* **25**, 275–280.

Boynton, R. J., Balikhin, M. A., Billings, S. A., Sharma, A. S. & Amariutei, O. A. (2011), 'Data derived narmax dst model', *Annales Geophysicae* **29**, 965–971.

Burton, R. K., McPherron, R. L. & Russell, C. T. (1975), 'An empirical relationship between interplanetary conditions and dst', *Journal of Geophysical Research* **80**(31), 4204–4214.

Dolenko, S. A., Myagkova, I. N., Shiroky, V. R. & Persiantsev, I. G. (2014), Objective discrimination of geomagnetic disturbances and prediction of dst index by artificial neural networks, *in* 'Proceedings of the 10th Intl Conf. "Problems of geocosmos"', pp. 270–275.

Elman, J. L. (1990), 'Finding structure in time', *Cognitive Science* **14**, 179–211.

Fenrich, F. R. & Luhmann, J. G. (1998), 'Geomagnetic response to magnetic clouds of different polarity', *Geophysical Research Letters* **25**, 2999–3002.

Ganushkina, N. Y., Amariutei, O. A., Shpritz, Y. Y. & Liemohn, M. W. (2013), 'Transport of the plasma sheet electrons to the geostationary distances', *Journal of Geophysical Research* **118**.

Ganushkina, N. Y., Liemohn, M. W., Amariutei, O. A. & Pitchford, D. (2013), 'Low energy electrons (5-50 kev) in the inner magnetosphere', *Journal of Geophysical Research* **119**, 246–259.

Ganushkina, N. Y., Liemohn, M. W. & Pulkkinen, T. I. (2012), 'Storm-time ring current: model-dependent results', *Annales Geophysicae* **30**, 177–202.

Ganushkina, N. Y., Pulkkinen, T. I., Bashkirov, V. F., Baker, D. N. & Li, X. (2001), 'Formation of intense nose structures', *Geophysical Research Letters* **28**(3), 491–494.

Ganushkina, N. Y., Pulkkinen, T. I. & Fritz, T. (2005), 'Role of substorm-associated impulsive electric fields in the ring current development during storms', *Annales Geophysicae* **23**, 579–591.

Ganushkina, N. Y., Pulkkinen, T. I., Liemohn, M. W. & Milillo, A. (2006), 'Evolution of the proton ring current energy distribution during april 21-25, 2001 storm', *Journal of Geophysical Research* **111**, A11S08.

Gleisner, H. & Lundstedt, H. (1997), 'Response of the auroral electrojets to the solar wind modeled with neural networks', *Journal of Geophysical Research* **102**(A7), 14,269–14,278.

Gleisner, H. & Lundstedt, H. (2001), 'Auroral electrojet prediction with dynamical neural networks', *Journal of Geophysical Research* **106**(A11), 24,541–24,549.

Gleisner, H., Lundstedt, H. & Wintoft, P. (1996), 'Predicting geomagnetic storms from solar-wind data using time-delay neural networks', *Annales Geophysicae* **14**(7), 679–686.
URL: *http://www.ann-geophys.net/14/679/1999/*

Goertz, C. K., Shan, L. H. & Smith, R. A. (1993), 'Prediction of geomagnetic activity', *Journal of Geophysical Research* **98**(A5), 7673–7684.

Hernandez, J. V., Tajima, T. & Horton, W. (1993), 'Neural net forecasting for geomagnetic activity', *Geophysical Research Letters* **20**, 2707–2710.

Ji, E. Y., Moon, Y. J., Gopalswamy, N. & Lee, D. H. (2012), 'Comparison of dst forecast models for intense geomagnetic storms', *Journal of Geophysical Research* **117**, A03209.

Koons, H. C. & Gorney, D. J. (1991), 'A neural network model of relativistic electron flux at geosynchronous orbit', *Journal of Geophysical Research* **96**, 5549–5556.

Li, X., Oh, K. S. & Temerin, M. (2007), 'Prediction of the al index using solar wind parameters', *Journal of Geophysical Research* **112**, A06224.

Lundstedt, H. (1992), 'Neural networks and predictions of solar-terrestrial effects', *Planetary and Space Science* **40**(4), 457–464.

Lundstedt, H., Gleisner, H. & Wintoft, P. (2001), 'Operational forecasts of the geomagnetic *Dst* index', *Geophysical Research Letters* **106**(A6), 1–4.

Lundstedt, H. & Wintoft, P. (1994), 'Prediction of geomagnetic storms from solar wind data with the use of a neural network', *Annales Geophysicae* **12**, 19–24.

Luo, B., Li, X., Temerin, M. & Liu, S. (2013), 'Prediction of the au, al, and ae indices using solar wind parameters', *Journal of Geophysical Research* **118**, 7683–7694.

Mayaud, P. N. (1980), *Derivation, meaning, and use of geomagnetic indices*, Vol. 22 of *Geophysical monograph*, American Geophysical Union.

Murphy, A. H. (1988), 'Skill scores based on the mean square error and their relationships to the correlation coefficient', *Monthly Weather Review* **116**, 2417–2424.

Murphy, A. H. (1993), 'What is a good forecast? An essay on the nature of goodness in weather forecasting', *American Meteorological Society* **8**, 281–293.

Murphy, A. H. (1996), 'The finley affair: A signal event in the history of forecast verification', *Weather and Forecasting* **11**(1), 3–20.

Murphy, A. H., Brown, B. G. & Chen, Y.-S. (1989), 'Diagnostic verification of temperature forecasts', *American Meteorological Society* **4**, 485–501.

Murphy, A. H. & Wilks, D. S. (1998), 'A case study of the use of statistical models in forecast verification: Precipitation probability forecasts', *Weather and Forecasting* **13**(3), 795–810.
URL: *http://dx.doi.org/10.1175/1520-0434(1998)013%3C0795:ACSOTU%3E2.0.CO;2*

Murphy, A. H. & Winkler, R. L. (1987), 'A general framework for forecast verification', *American Meteorological Society* **115**, 1330–1338.

O'Brien, T. P. & McPherron, R. L. (2000), 'Forecasting the ring current dst in real time', *Journal of Atmospheric and Solar-Terrestrial Physics* **62**, 1295–1299.

Pallocchia, G., Amata, E., Consolini, G., Marcucci, M. F. & Bertello, I. (2006), 'Geomagnetic Dst index forecast based on IMF data only', *Annales Geophysicae* **24**, 989–999.

Rastätter, L., Kuznetsova, M., Glocer, A., Welling, D., Meng, X., Raeder, J., Wiltberger, M., Jordanova, V. K., Yu, Y., Zaharia, S., Weigel, R. S., Sazykin, S., Boynton, R., Wei, H., Eccles, V., Horton, W., Mays, M. L. & Gannon, J. (2013), 'Geospace environment modeling 2008–2009 challenge: Dst index', *Space Weather* **11**, 187–205.

Takashi, K., Toth, B. A. & Olson, J. V. (2001), 'An automated procedure for near-real-time kp estimates', *Journal of Geophysical Research* **106**(A10), 21017–21032.

Temerin, M. & Li, X. (2006), 'Dst model for 1995–2002', *Journal of Geophysical Research* **111**, A04221.

Vassiliadis, D., Klimas, A. J., Baker, D. N. & Roberts, D. A. (1995), 'A description of the solar wind-magnetosphere coupling based on nonlinear filters', *Journal of Geophysical Research* **100**(A3), 3495–3512.

Vassiliadis, D., Klimas, A. J., Valdivia, J. A. & Baker, D. N. (1999), 'The dst geomagnetic response as function of storm phase and amplitude and the solar wind electric field', *Journal of Geophysical Research* **104**(A11), 24,957–24,976.

Wang, C. B., Chao, J. K. & Lin, C. H. (2003), 'Influence of the solar wind dynamic pressure on the decay and injection of the ring current', *Journal of Geophysical Research* **108**(A9), 1341.

Wang, Y., Shen, C. L., Wang, S. & Ye, P. Z. (2003), 'An empirical formula relating the geomagnetic storm's intensity to the interplanetary parameters: $-\overline{VB_z}$ and δt', *Journal of Geophysical Research* **30**(20), 2039.

Watanabe, S., Sagawa, E., Ohtaka, K. & Shimazu, H. (2002), 'Prediction of the dst index from solar wind parameters by a neural network method', *Earth Planets Space* **54**, 1263–1275.

Wei, H. L., Billings, S. A. & Balikhin, M. (2004), 'Prediction of the dst index using multiresolution wavelet models', *Journal of Geophysical Research* **109**, A07212.

Wing, S., Johnson, J. R., Jen, J., Meng, C. I., Sibeck, D. G., Bechtold, K., Freeman, J., Costello, K., Balikhin, M. & Takashi, K. (2005), 'Kp forecast models', *Journal of Geophysical Research* **110**, A04203.

Wintoft, P., Buresova, D., Bushell, A., Heynderickx, D., Núñez, M., Perrone, L., Qahwaji, R., Schmutz, W., Thomson, A. W. P., Tsagouri, I. & Viljanen, A. (2011), Metrics for space weather models, Technical report, COST ES0803.

Wu, J. G. & Lundstedt, H. (1997), 'Neural network modeling of solar wind-magnetosphere interaction', *Journal of Geophysical Research* **102**, 14,457–14,466.

Zhu, D., Billings, S. A., Balikhin, M. A., Wing, S. & Alleyne, H. (2007), 'Multi-input data derived dst model', *Journal of Geophysical Research* **112**, A06205.