Document reference: IRM/GALOCAD/OUT320-1	Deliverable : OUT320-1	Date: 19-April-2008 Version: 2
Contract ref : GJU/06/2423/CTR/GALOCAD	Written by : I. Kutiev, B. Andonov, P. Muhtarov	Verified by : S. Stankov

GALOCAD

Development of a Galileo Local Component for nowcasting and forecasting of atmospheric disturbances affecting the integrity of high precision Galileo applications.

WP 320 Technical Report :

"Hybrid model for forecasting Dourbes K index"

Geophysical Institute, BAS, Acad. G. Bonchev str., bl. 3, Sofia 1113, Bulgaria

LIST OF ABBREVIATIONS

DRC	Diode Rectifier Circuit
GI/BAS	The Geophysical Institute of the Bulgarian Academy of Sciences
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HDK	Hybrid Dourbes K model
IMF	Interplanetary Magnetic Field
K	K index, 3-hour estimate, from the Dourbes station
Kd	K index value, hourly interpolated
Kdf	HDK model value
Kdfp	HDK forecast value
Ksw	MAK model value of K based on solar wind data
KU	K index unit
MA	The Muhtarov-Andonov model based on Bz data only
MAK	The Muhtarov-Andonov-Kutiev model based on solar wind data
NOAA	National Oceanic and Atmospheric Administration
RMI	Royal Meteorological Institute
RMS	Root Mean Square
RTK	Real Time Kinematics
TEC	Total Electron Content
TECU	TEC Unit (10^{16} electrons/m^2)
TAD	Travelling Atmospheric Disturbance
TID	Travelling Ionospheric Disturbance
WP	Work Package

TABLE OF CONTENT

1. Introduction

The geomagnetic activity is the primary cause of ionospheric disturbances. There are several ways in which the increased geomagnetic activity can affect large scale ionospheric structures: through changes in thermospheric composition, ion drag by atmospheric winds, electrodynamic drifts, and large scale travelling atmospheric disturbances (TADs). Geomagnetic activity characterizes the degree of disturbance of the Earth magnetic field and the level of this disturbance is quantified by a number of indices, each of which characterizes the origin and the time scale of its variations. The *Dst* index, for example, represents magnetic disturbances caused by the magnetospheric ring current with 1-hour sampling rate, while the *sym-H* index quantifies the same disturbances with 1-minute sampling rate. The auroral activity is represented by the *AE* index and its components: *AU* representing the eastward electrojet (in the evening sector) and *AL* the westward (morning) electrojet. Magnetic disturbances at midlatitudes are caused mainly by the ring current and auroral electrojets, although other magnetospheric currents and telluric currents on the Earth surface can also contribute. The midlatitude magnetic disturbances are quantified by a number of indices *A* and *K*, depending on the method of deriving them. Each of the abovementioned indices is derived from a specific number of magnetic stations and represents the global (planetary) geomagnetic activity. It is well accepted that the planetary 3-hour *K* (*Kp*) and the daily *A* (*Ap*) indices are the best representatives of the large-scale midlatitude ionospheric disturbances (Thomsen, 2004; and the references therein).

The small-scale ionospheric disturbances affecting GNSS applications may have another, non-geomagnetic origin. They can be caused by TADs, atmospheric tides, acoustic waves, etc. As shown by Warnant et al. (2006a, 2006b), these ionospheric disturbances, and particularly those related to the RTK ionospheric intensity, can be observed both during and outside geomagnetic storms. During storms, however, the RTK intensity is much higher than during geomagnetically quiet periods. Nebdi et al. (2004) have shown that excessive RTK intensity (>19 I_E units) is observed in 83% of the cases when Kp>7, and in 100% of the cases at Kp >8. It means that the ionospheric disturbances generated during high geomagnetic activity (i.e. for high Kp values) tend to stronger affecting the GPS positioning accuracy. In this sense, the Kp index specification and prediction is of vital importance for the GNSS users.

The 3-hour Kp index, being a good representation of the large-scale ionospheric disturbances known as ionospheric storms, is too inaccurate when representing disturbances that may potentially degrade the GPS positioning accuracy. The small-scale disturbances are localized phenomena and the local K index derived from the nearest magnetic station better fits these disturbances than the planetary Kp. Additionally, the 3-hour time-scale is much larger for the shorter characteristic time of small-scale ionospheric variations. Taking into account these general constrains, we use further the 1-hour K index derived at the Dourbes magnetic observatory. The local K index derivation from magnetogram records is specific for each magnetic station because it depends on the station's latitude and because the deviations from the quiet Sq variations need to be calibrated with those of the reference magnetic station of Niemegk (Bartels et al., 1939;

Mayaud, 1980). Della-Rose et al. (1999) have shown that the K index obtained by the same methodology but with different time windows (from 15 min to 5 hours) provide considerably different values. This fact raise the question of whether the 1-hour K values differ from the conventional 3-hour K index values and to what extent. To avoid this uncertainty, we plan to calculate K values in a 3-hour window, but sliding this window with 1-hour step. The K value at each step is assigned to the center of window. We do not know whether this method has been used before in geomagnetic practice, but it is widely used in ionospheric studies. The sliding window has the advantage to smoothes out small-scale variations and increases the sampling rate.

The numerous models providing forecast of geomagnetic indices can be divided in two main groups. The first group includes models based on extrapolation of past data and delivering forecast for a few hours ahead. However, the extrapolation technique cannot take into account the sharp increase of geomagnetic activity at the onset of geomagnetic storms and the forecast is therefore limited to periods of relatively quiet conditions. The second group of models is based on solar wind parameters which are used as predictors of the solar wind – magnetosphere interactions.

The geomagnetic indices reflect the complex non-linear transfer of energy deposited by the solar wind into the magnetosphere with a subsequent transfer of energy to the polar ionosphere. Some models (Costello, 1997; Boberg et al., 2000; Balikhin et al., 2001; Boaghe et al., 2001; Wing et al., 2005) use neural network (NN) technique to forecast the Kp, AE and Dst indices with a lead time of 30-60 minutes which is the time needed for the solar wind stream to pass the distance between the Lagrangian (L1) point and the Earth's magnetosphere. Other models use empirical relations between the solar wind and the geomagnetic activity parameters which are based on some well-known physical analogues. In a further description, we will consider this type of analogue models. Hones (1979) and Klimas et al. (1992) have used the dripping faucet analogue to describe the plasmoid formation in the magnetosphere tail, the main transposer of solar wind energy into substorm activity. Baker et al. (1990) and Vassiliadis et al. (1993) have used an electric LRC circuit of a damped linear oscillator to represent the return of a "perturbed" magnetosphere to its "quiet" state. Muhtarov and Andonov (2000), denoted further as MA, developed a model relating Kp to the Interplanetary Magnetic Field (IMF) Bz component by using an electric diode rectifier circuit (DRC) analogue. It is well accepted that Bz is the main driver of geomagnetic activity and of the Kp variations in particular (Kamide et al., 1998). The circuit of the MA model is similar to that of Vassilliadis et al. (1993) in which the inductance part was replaced by a half-wave diode rectifier. MA used Bz as an input voltage, while the output voltage was defined as a "modified" Bz (Bzm), having positive variations only. The hourly values of the new quantity, Bzm, were correlated with the hourly-interpolated Kp values to obtain the model parameters.

The negative turn of Bz causes an increase of Kp, known as "driven" response (Klimas et al., 1991). The increase of B_z or turning positive is not followed by an immediate and proportional decrease of Kp. The changes of Kp appear more gradual and delayed. MA found that the cross-correlation between Bz and Kp had a maximum at a time lag of about 2 hours. This means that Kp best correlates with Bz from the previous 2 hours. In

order to improve the dependence of Kp on Bz, MA introduced a new function of Bz (denoted as Bzm), which is positive and contains a delayed reaction to Bz changes. To do this, MA used an analogy with another inertial system, which involves “loading” and “unloading” processes with different time constants.

The electrical circuit shown on Fig.1a represents such a system. The circuit includes a half-wave diode rectifier *D*, a smoothing capacitor *C* and two resistors R_1 and R_2. If assuming that the input voltage U_1 is a step-like function formed by discrete values at arbitrary moments of time and $R_2 >> R_1$ the output voltage U_2 within the time-step [ti, ti+1] is given by the well-known relationship:

$$U_2(t_{[i+1]}) = \begin{cases} (U_2(t_i) - U_1(t_i))\exp\left(-\dfrac{t_{[i+1]} - t_i}{T_1}\right) + U_1(t_i); & U_2(t_i) < U_1(t_i) \\ U_2(t_i)\exp\left(-\dfrac{t_{[i+1]} - t_i}{T_2}\right); & U_2(t_i) > U_1(t_i) \end{cases} \quad (1)$$

where $T_1 = R_1 C$ and $T_2 = R_2 C$.

The expressions (1) have recurrent feedback: the voltage U_2 obtained from the previous step $[t_{i-1}, t_i]$ is placed on the right-hand side of the equations for obtaining U_2 in the next step $[t_i, t_{i+1}]$. The first expression in (1) represents a process of loading the capacitor *C* with a time constant T_1, while the second expression represents the unloading process with a time constant T_2. The loading takes place while U_1 is higher than U_2 and the diode is open. If U_1 becomes lower than U_2, the diode is closed and the capacitor starts discharging through the resistor R_2.

The abovedescribed process is schematically presented on Fig.1b. The input voltage U_1 is represented as a simple sinusoid curve (the thin line) and the output voltage U_2 curve (the solid line). The loading takes place when $U_1 > U_2$. The output voltage U_2 accepts now only positive values, gradually decreasing when U_1 is lower. The time constants T_1 and T_2 shown in Fig.1b are arbitrary and are used just to demonstrate the functioning of the circuit but do not relate to the values considered later in the paper. Making use of the analogy with this electrical scheme, MA defined Bzm through equation (1) with a replacement of U_1 value with Bz and the U_2 value with Bzm.

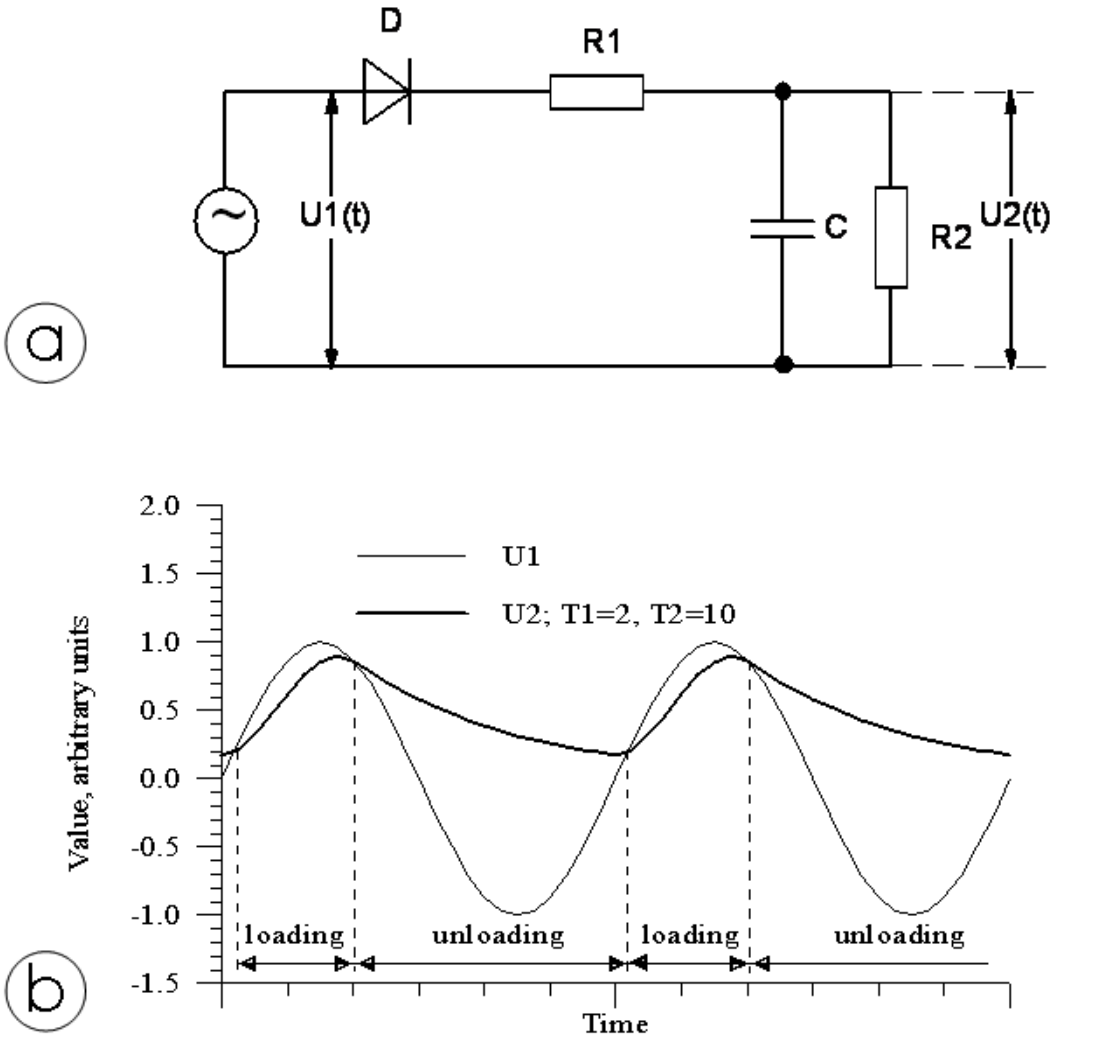

Figure 1: (a) Equivalent electric circuit DRC, giving the delayed reaction of the output voltage U2 to the input voltage U1. D is a half-wave diode rectifer, C is a capacitor, R_1 and R_2 are resistors;(b) a sample of a sinusoidal input voltage U_1 and the resulting output voltage U_2. Loading and unloading phases are marked with horizontal bars. T_1 and T_2 are assured arbitrary to demonstrate the functioning of the circuit and do not relate to the values considered in the paper.

By using 27 years of IMF data (1973-1999), MA estimated that Bzm can raise the correlation between Bz and Kp from -0.4 to 0.7. Later, Andonov et al. (2004) further denoted as MAK, improved the model, adding dependences on solar wind dynamic pressure and velocity. The new model, named MAK, is part of the present HDK model. A short description of the MAK model is given below.

2. MAK model

MAK calculated the cross-correlation between Kp and the available solar wind parameters. By ranking the magnitude of the cross-correlation, they included in the model, besides Bzm as in MA model, the higher correlated with Kp solar wind velocity V and dynamic pressure P. In this way MAK obtained the average dependence of Kp on those parameters. Based on the form of these dependences, the MAK model has then obtained in the following form:

$$\mathrm{Ksw} = a_o + a_1\mathrm{Bzm} + a_2\mathrm{P} + a_3\mathrm{V} + a_4\mathrm{Bzm}^2 + a_5\mathrm{P}^2 \qquad (2)$$

Ksw now represents the model value of (2) which is presumably very close to Kp. The coefficients a_i are obtained by fitting the expression on the right side of (2) to the ground-based Kp values. This time the solar wind parameters were taken from the ACE (Advanced Composition Explorer) space platform for the years 1998-2004. ACE data are more homogenous than the data from the IMP satellites and assure better accuracy in the modeling. Because the ACE data base consist of hourly values of Bz, V and P, the 3-hour Kp values for the same period were interpolated to obtain the hourly values. To perform the fitting of expression (2), MAK first obtained the time constants T1 and T2 in Bzm in the following way. The right hand side of (2) was repeatedly fitted to Kp values from the whole database by using the grid of pairs (T_1, T_2). The corresponding RMS (root mean square) deviations of model from data were calculated for each pair and then compared. MAK defined the model time constants as those having the minimum RMS. Using ACE database in period 1998-2004, MAK found T_1=3 and T_2=7 hours. Therefore, the time constants of the delayed reaction of Kp to Bz forcing were obtained as average over the whole database.

The model error (alternatively, the model accuracy) is an important characteristic of the model. The average root mean square deviation of the model Ksw values from the observed Kp values (i.e. the overall model error) was estimated at 0.63 KU (K unit). It is known that the conventional Kp is defined in the range from 0 to 9, having 3 grades around each unit (for example, 2^-, 2^0, 2^+). In order to digitize these symbols, we assign the value of 1.66 to "2^-", 2.0 to "2^0", and 2.33 to "2^+". Similar assignment (-0.33, 0, +0.33) is made to each Kp unit. In Fig.2 a comparison is made between the MA model error (diamonds), the MAK model error calculated from the 3-hour Kp values (triangles), and the MAK model error calculated with all hourly Kp values (circles). The error was calculated separately for each 1 K unit of Kp. This presentation clearly shows that the model error is lowest around the point of Kp=2 where the occurrence probability has a maximum. At this point, the three estimates are closest to each other. It is evident that the MAK model with hourly Kp performs best. Even for extreme cases of Kp=8, the model error still remains below 1 K unit. Notice that the overall error is not a simple averaging of curves in Fig.2 but depends also on the number of values in each Kp unit bin, which has a maximum around Kp=2. The model error estimated by using the 3-hour Kp data set is 0.72 KU, or about 14% higher than that estimated from the interpolated hourly Kp set. The model error of MA is 0.96 KU.

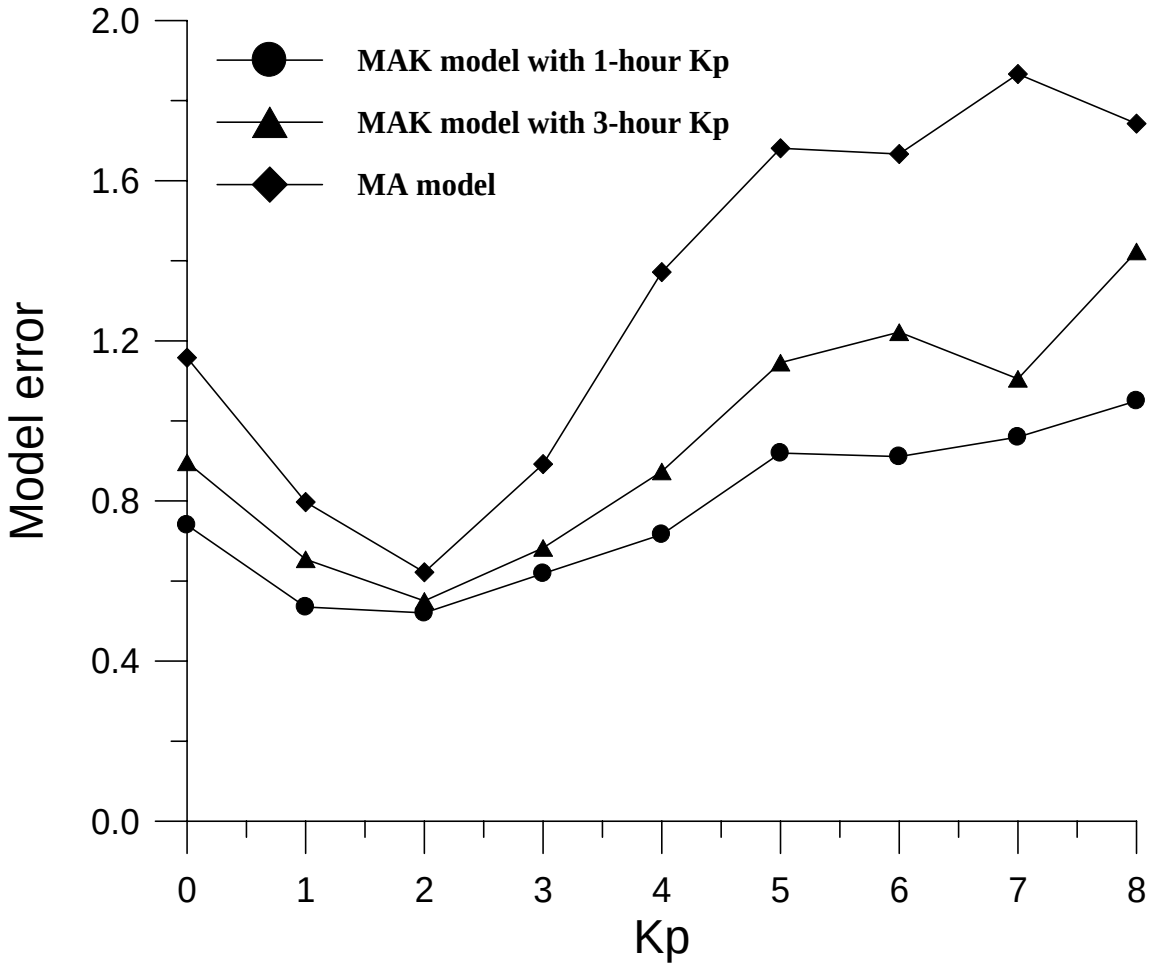

Figure 2: Model errors of: MA (diamonds), MAK with 3-hour Kp (triangles) and MAK with 1-hour interpolated Kp (circles) calculated for each unit-wide bin of the Kp magnitude.

It is important to compare the MAK model performance with some of the internationally recognized models, for example the Costello forecast model (Costello, 1997) that was implemented in the operational forecast service at the Space Environment Center of NOAA, Boulder, Colorado. A comparison is shown in Fig.3. Although the Costello model is based on the Neural Network (NN) technique and the MAK model employs another, linear regression method (2), both model errors are of similar magnitude. Figure 4 shows another comparison: between MAK and Lund Observatory Kp forecasting model (Boberg et al., 2000), also based on the NN technique. The absolute values of the MAK model error, shown in Fig. 2 and Fig. 3 are quite different, which is due to the different databases used for the comparison. For the comparison with Costello model, only few months of data in 2003 were available. The limited database resulted in much larger RMSE than it is shown in other figures. In Fig.4, comparison is made again on limited data by using the 3-hour Kp only. Therefore, these model errors should be compared separately and not in the context of further error estimates which are based on much larger database.

Geomagnetic activity models, like MAK, based on solar wind measurements onboard ACE, can predict sudden changes some hours before the Earth magnetic field reacts. It takes about half an hour for the solar wind to pass the distance between L1 libration point, where ACE is places, and magnetosphere (4.6 x10^6 km) and two more hours for development of substorms and intensification of the ring current. Therefore, we can accept that solar wind based models are capable to predict changes in the K index at least 3 hours in advance. Apart from this important advantage, these models have certain disadvantages. Usually, the Kp changes are not proportional to the IMF Bz negative deviations. MAK time constants, derived on statistical basis, cannot properly determine Kp rate of change in all individual cases.

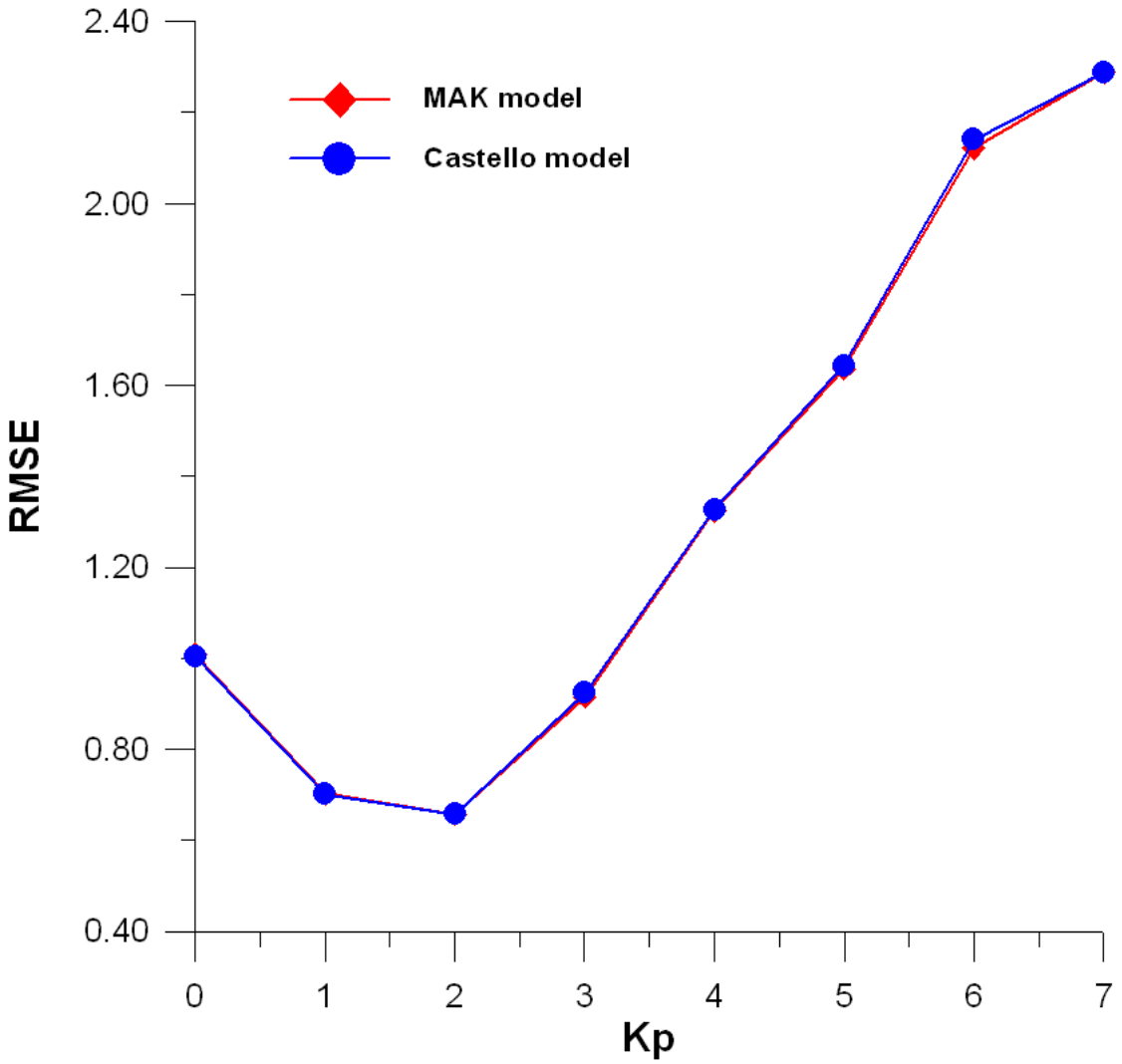

Figure 3: Comparison between MAK and Castello model errors (RMSE)

As it was pointed out above, forecasting models based on weighted extrapolation of past K values can make reasonable predictions for a much longer period of time (longer lead time) than the solar wind based models can possibly do.

Theoretically, reasonable predictions can be made for a time period up to the value of the time constant of the autocorrelation function. Since the K index time constant is 18-20 hours, predictions based on the weighted extrapolation can be made for nearly one day ahead provided that sudden changes in the solar wind parameters do not occur. It is therefore reasonable to develop a model which would combine the advantages of both approaches: the longer lead time of the weighted extrapolation and the detection of sudden changes by monitoring the solar wind parameters. The so called Hybrid Dourbes K (HDK) model is a realization of the abovedescribed concept.

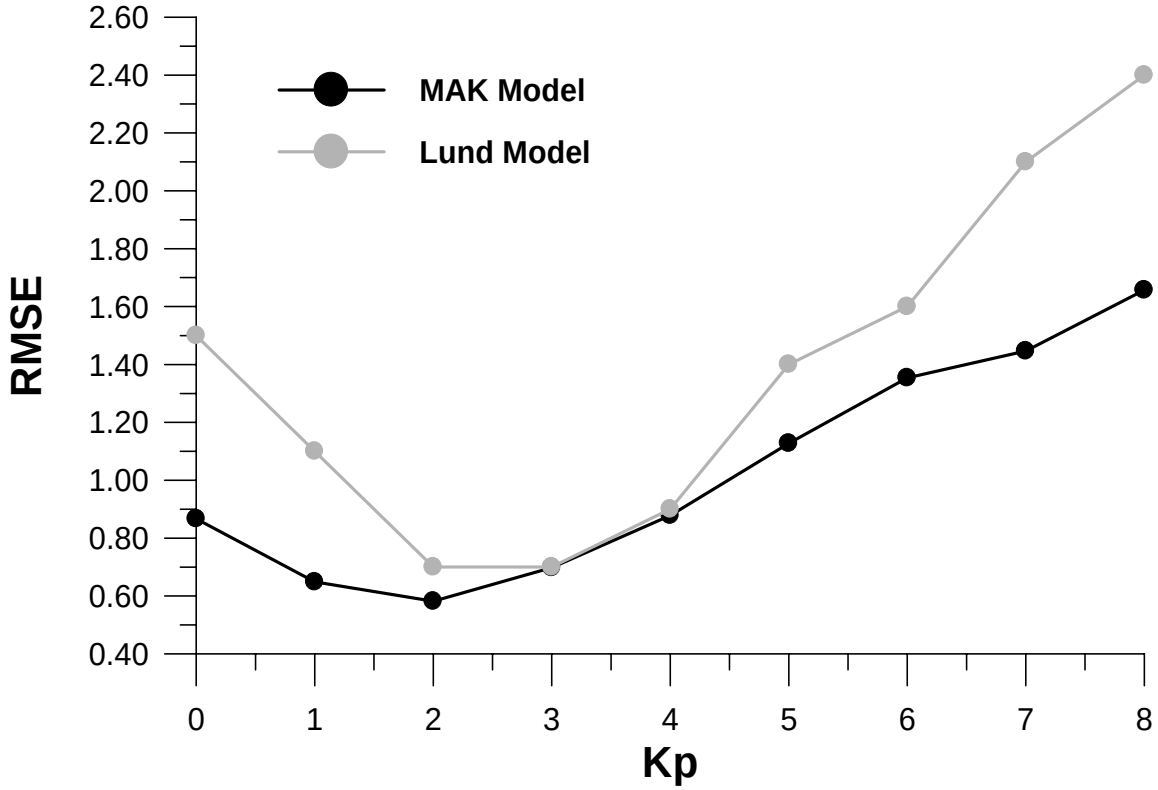

Figure 4. Comparison between MAK and Lund model errors (RMSE)

3. HDK model development

The operation of the Hybrid Dourbes K (HDK) model is demonstrated by the flowchart in Fig.5. The HDK model consists of two branches which are merging in the last stage to produce the final product, e.g. the Dourbes model value of K.

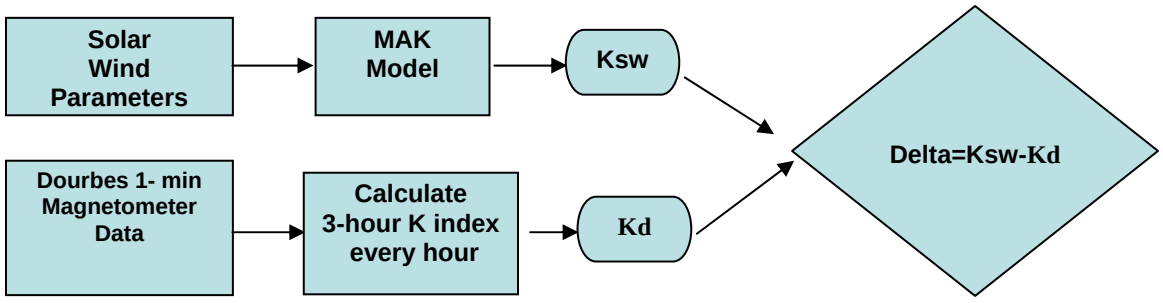

Figure 5: Flow chart of the Hybrid Dourbes K model. The upper branch represents the adoption of MAK-derived Ksw to the derivation of the Dourbes K index. The lower branch represents the production of the K index on a 1-hourly basis.

The upper branch of the model represents the development/operation of the MAK model, by re-calculating model coefficients for Dourbes K data. The lower branch represents the production of the 3-hour index Kd with 1-hour step resolution. At the end of the chart, both branches merge to produce the quantity Delta, being the difference between the solar wind-derived Ksw and the K value derived from magnetograms, i.e. Kd. Thus, the quantity Delta is a key element of the new HDK model.

3.1 Adoption of MAK model to Dourbes K index

Originally, the MAK model coefficients were obtained by fitting the model expressions to Kp values in the period 1998-2004 when ACE data were available. To adapt the model to the Dourbes K data, we have to check first whether the average dependences of K on solar wind parameters remain the same. We assumed that the dependences would be the same if the cross-correlation coefficient between the 3-hour Kp and Dourbes K indices exceeds 0.8. Fig.6 shows the normalized cross-correlation function between the indices versus time lag. The colored lines represent the cross-correlation at different 3-hour local time intervals. The maximum cross-correlation of all time-intervals is achieved at the time lag = 0, which means that there is no delay between the variations of the two indices.

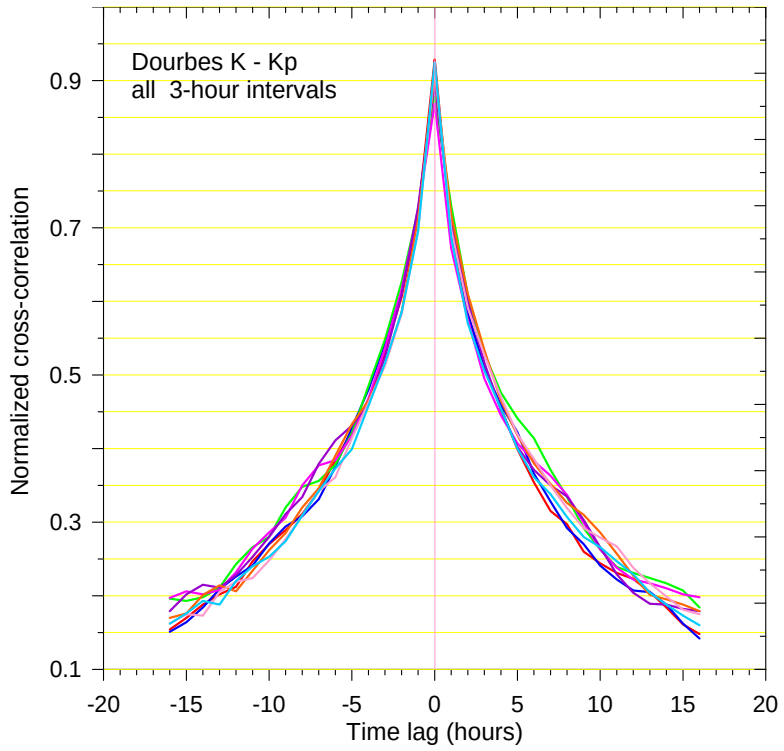

Figure 6: Cross-correlation between the 3-hour Kp and Dourbes K indices for the time lag ±20 hours.

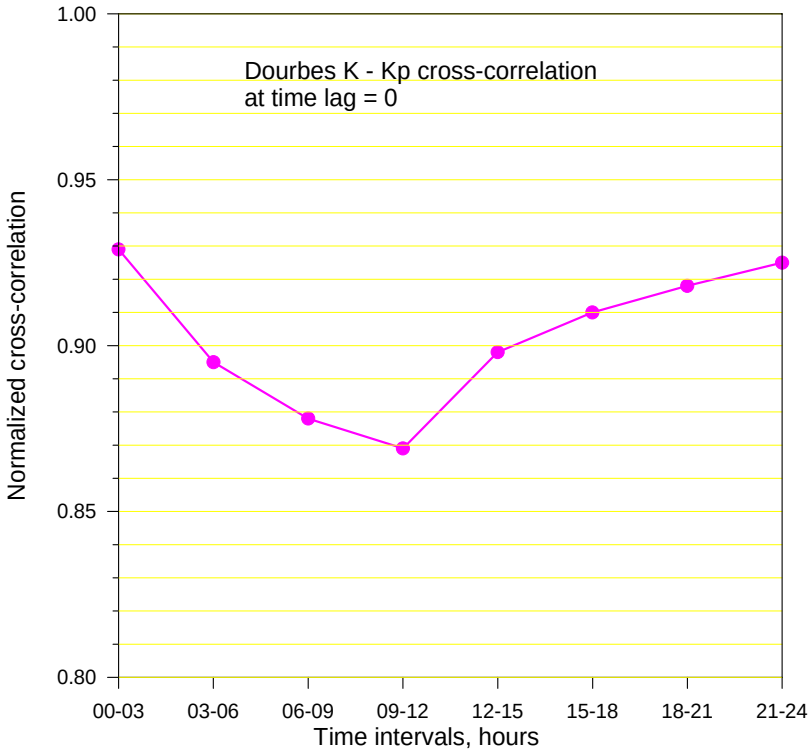

Figure 7: Local time variation of the cross-correlation between 3-hour Kp and Dourbes K indices for the time lag = 0.

Fig.7 shows the cross-correlation between Dourbes K and Kp at the time lag = 0. The maximum of 0.929 is seen in time interval (00-03) hours. The minimum cross-correlation of 0.869, or 6% lower, is in (09-12) hour interval. The local time differences are expected, because of the asymmetry of current system in the auroral oval. The evening and morning electrojets are generally on the night side and then the Dourbes K represents better the magnetic disturbances in the auroral oval (captured by Kp) than it does on the dayside. The high cross-correlation between Dourbes K and Kp with only 6% tolerance proves that the MAK definitions are applicable to Dourbes K index.

The Dourbes K index was linearly interpolated to obtain 1-hour resolution values. The interpolated hourly values are further denoted as Kd. In accordance with the database changes, two new sets of coefficients were obtained. The first set of coefficients ignores their local time dependence. The respective formula for Ksw is:

$$\begin{aligned} Ksw = &- 0.009347\ Bzm^2 + 0.350934\ Bzm + 0.004763\ V - \\ &-0.005151\ P^2 + 0.188879\ P - 0.687897 \end{aligned} \qquad (3)$$

Fig. 7 shows that the cross-correlation between Kp and Dourbes K indices is local time dependent. Previously, we have obtained (results not published yet) that the Kp-based Ksw had also local time dependence. By definition, Kp is a planetary index and should not have a local time dependence. The same is true for Ksw, because solar wind parameters and not local time dependent. The local time dependence for Ksw was obtained by calculating the model coefficients from K values separately for each hour of the day. Precise calculations of Ksw would require using local time dependence of each coefficient a_i. We assumed that the local time dependence of model coefficients is a second order effect and at present HDM model we neglect this correction. Nevertheless, for the sake of completeness, the hourly values of model coefficients were calculated and presented in the following Table 1.

Table 1

h	a1	a2	a3	a4	a5	a0
00	-0.014530	0.447056	0.004705	-0.005838	0.208673	-0.714878
01	-0.011754	0.430674	0.004578	-0.005148	0.185335	-0.678447
02	-0.009146	0.398970	0.004500	-0.005982	0.183642	-0.664761
03	-0.007609	0.362546	0.004371	-0.005287	0.177946	-0.619901
04	-0.008068	0.359614	0.004280	-0.005422	0.183568	-0.653496
05	-0.007885	0.347272	0.004230	-0.005493	0.180562	-0.666528
06	-0.007539	0.331876	0.004224	-0.007483	0.195623	-0.704093
07	-0.008119	0.321497	0.004362	-0.006718	0.189059	-0.788550
08	-0.007363	0.292517	0.004387	-0.007850	0.202172	-0.776112
09	-0.006850	0.267942	0.004280	-0.007654	0.207768	-0.654227
10	-0.007863	0.262154	0.004323	-0.006743	0.212033	-0.621137
11	-0.008664	0.265612	0.004322	-0.006141	0.213254	-0.581020
12	-0.009836	0.278203	0.004403	-0.008798	0.241555	-0.624165
13	-0.010438	0.298423	0.004595	-0.010334	0.255362	-0.700733
14	-0.007332	0.280015	0.004911	-0.009651	0.258763	-0.780298
15	-0.006696	0.297829	0.005209	-0.011783	0.269778	-0.887531
16	-0.007491	0.330674	0.005585	-0.009884	0.252434	-1.020659

17	-0.008580	0.355044	0.005685	-0.006433	0.223470	-1.001479
18	-0.010902	0.398185	0.005654	-0.005592	0.206741	-0.979049
19	-0.013991	0.459716	0.005547	-0.007445	0.234674	-1.011828
20	-0.015818	0.475589	0.005382	-0.008505	0.251381	-0.960969
21	-0.016586	0.471143	0.005216	-0.008048	0.244417	-0.898543
22	-0.018755	0.489312	0.005024	-0.005423	0.215716	-0.812965
23	-0.018681	0.485708	0.004832	-0.005531	0.213477	-0.757474

The model expression (3) will take the form:

$$Ksw[h]= a_1[h]Bzm[h])^2 + a_2[h]Bzm[h]+ a_3[h]V[h]+ +a_4[h](P[h])^2 +a_5[h]P[h] +a_0[h] \quad (4)$$

The time constants T_1 and T_2 were also recalculated. Their average values remain the same as in the Kp-based model: T_1=3 hours and T_2=7 hours which is reasonable in view of the high coherence between Kp and Dourbes K.

3.2 HDK algorithm

The main idea of the combined use of Ksw and Kd is that at any given moment Ksw is corrected by the differences between Ksw and Kd (named Delta) obtained at some past moments. We use the fact that at the current moment *t*, Ksw is a measure of the current conditions in the solar wind (within the last hour, if we deal with its hourly values), while Kd represents the averaged conditions of the Earth magnetic field (at least) in the last 3 hours. So, at the current moment *t*, HDK provides the value of Ksw, corrected with an average of weighted Delta values at some previous hours, including the current, if available.

To perform prediction, the HDK model uses the method of weighted extrapolation of Winner-Hopf. This method was first applied to ionospheric short-term predictions by Muhtarov and Kutiev (1999). In the method, the weighted coefficients, assigned to the past data, are obtained by the autocorrelation function of respective quantity.

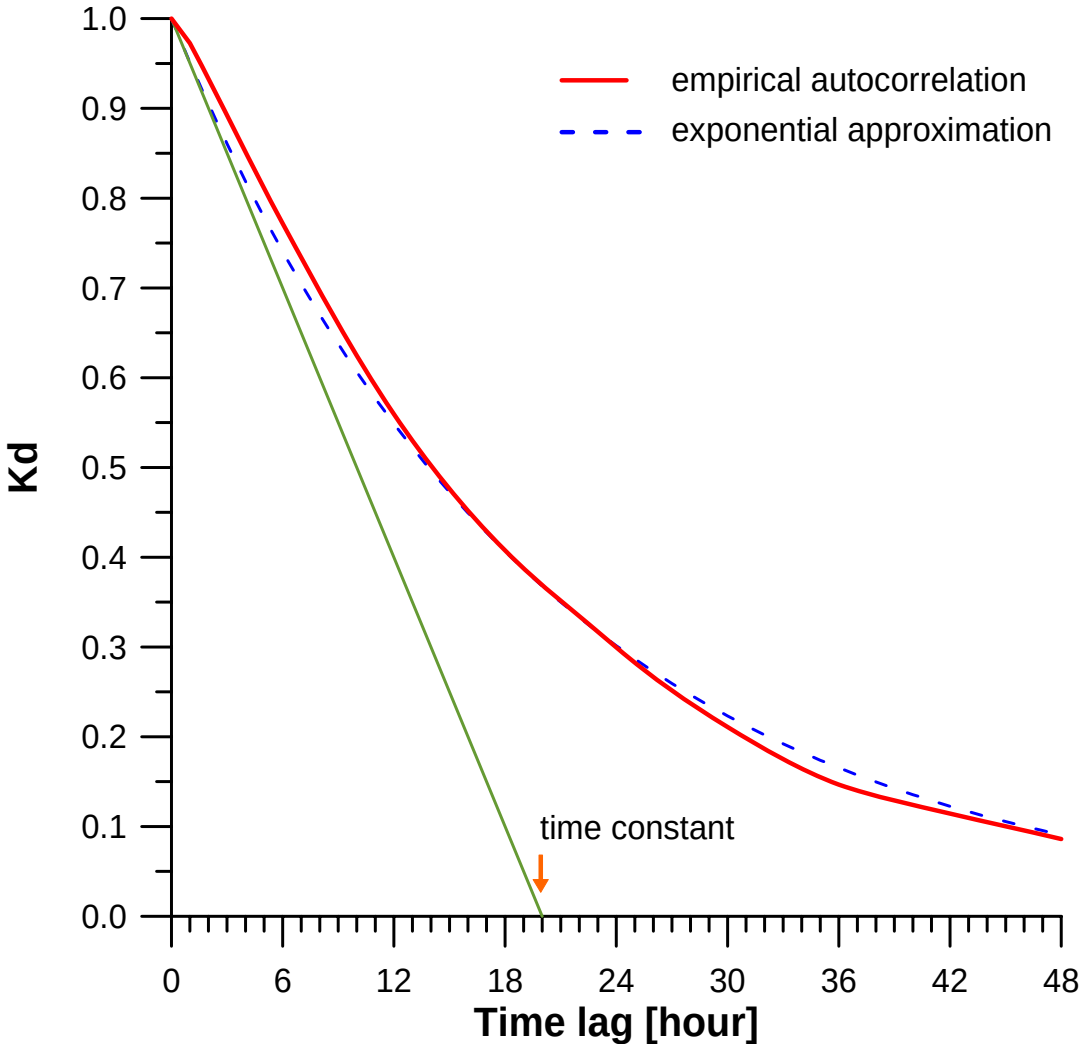

Figure 8. Empirical autocorrelation function (red line) of the hourly Dourbes K (Kd) and its exponential approximation (dashed blue line). The time constant is defined as the time at which the tangent at time lag=0 (green line) crosses the abscissa.

Fig.8 shows the autocorrelation function of Kd. The empirical autocorrelation function is given by the red line, while the blue dashed line represents its exponential approximation. If the autocorrelation function can be approximated by an exponential function, as is the case in most ionospheric processes, the weighted correction can be limited to the nearest (first) past value only.

The time constant of the approximated autocorrelation function is defined as the time at which the tangent at the point where the time lag = 0 crosses the abscissa. At that time the autocorrelation function decreases e times. The time constant of the Kd autocorrelation function is 20 hours.

Fig.9 shows the autocorrelation function of Delta up to 6 hours time lag. The time constant of Delta is 6 hours, much less comparing with that of Kd. If we denote the HDK model value at the current moment *t* with Kdf, the expression is the following:

$$Kdf(t)=Ksw(t) + MDelta + [Kdh(t-1) - Kdsw(t-1) - MDelta]*exp(-1/6) \quad (5)$$

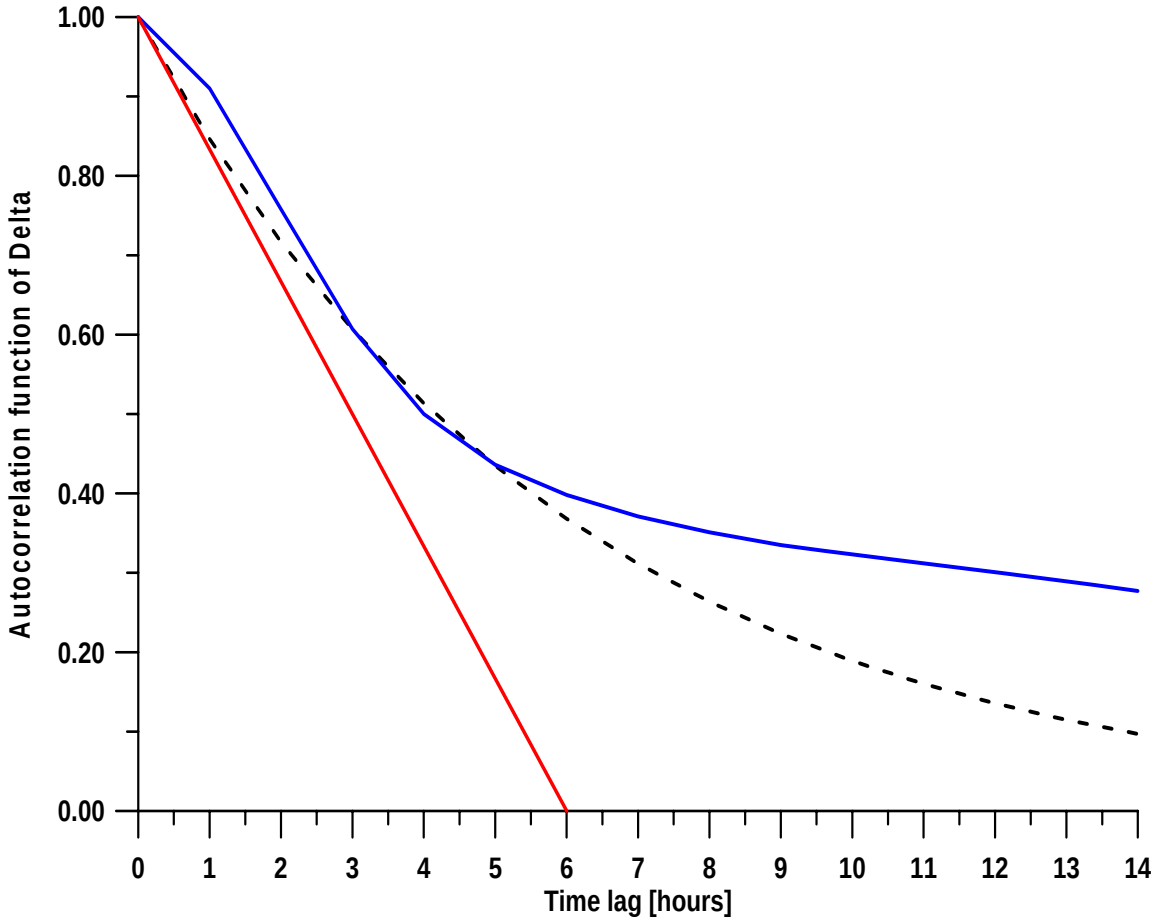

Figure 9: Empirical autocorrelation function of Delta (blue line) and its exponential approximation (dashed line). The time constant of Delta is 6 hours.

MDelta is the averaged Delta in the past 6 hours; (t-1) is the previous hour; the exponent is the weighting factor for the contribution from previous hour (in brackets). HDK algorithm works in the following simple way: At the current moment *t*, the model calculates (predicts) Kdf by using Ksw and the average Delta (MDelta). Kd at the time *t* is not taken into account. Kd actually goes one step behind and contributes to the term in brackets. The latter term is usually small during quiet conditions. If at the point of time *t* the Ksw value suddenly increase, Kdf will increase proportionally, providing that MDelta does not change. At next step, t+1, the term in brackets will be negative (Ksw(t) > Kd(t)) and tends to reduce the increase of Kdf(t+1) due to increasing Ksw values.

If one or two recent Kd values are missing, the model (5) can still calculate Kdf, because instead of Kd from the previous hour, values from earlier hours can be used, with their respective weights. This is important property of the model because it makes possible the use of magnetogram-derived K with larger sampling rate. For example, it is possible to use 3-hour Dourbes K values to obtain MDelta. Then MDelta will be an average of two 3-hour K values only which will increase the error.

Figures 10-12 show examples of HDK output Kdf, along with Ksw and Kd for 3 periods with different geomagnetic activity. Figure 10 represents the period 7–10 November 2004, when two intense geomagnetic storms occurred on 7-8 and 10-11 November. HDK model curve (green line) follows much closer the ground-based Kd, than the solar wind Ksw (red line) does. It is interesting to follow the behavior of Kdf and Ksw during the late hours of 7th November. Kd (blue line) increases sharply following the sudden increase of Bz (not shown here). Ksw also increases but with some delay determined by the time constant T_1. Obviously, in this case T_1 (which is obtained as an average over the whole database) is larger than needed to closely follow Kd. The HDK output Kdf, however, stays closer to Kd due to the correction Delta. The back slope of Ksw cannot be compensated by Delta and the model values of Kdf also deviate from Kd. This is an example of a large discrepancy between HDK model and the data.

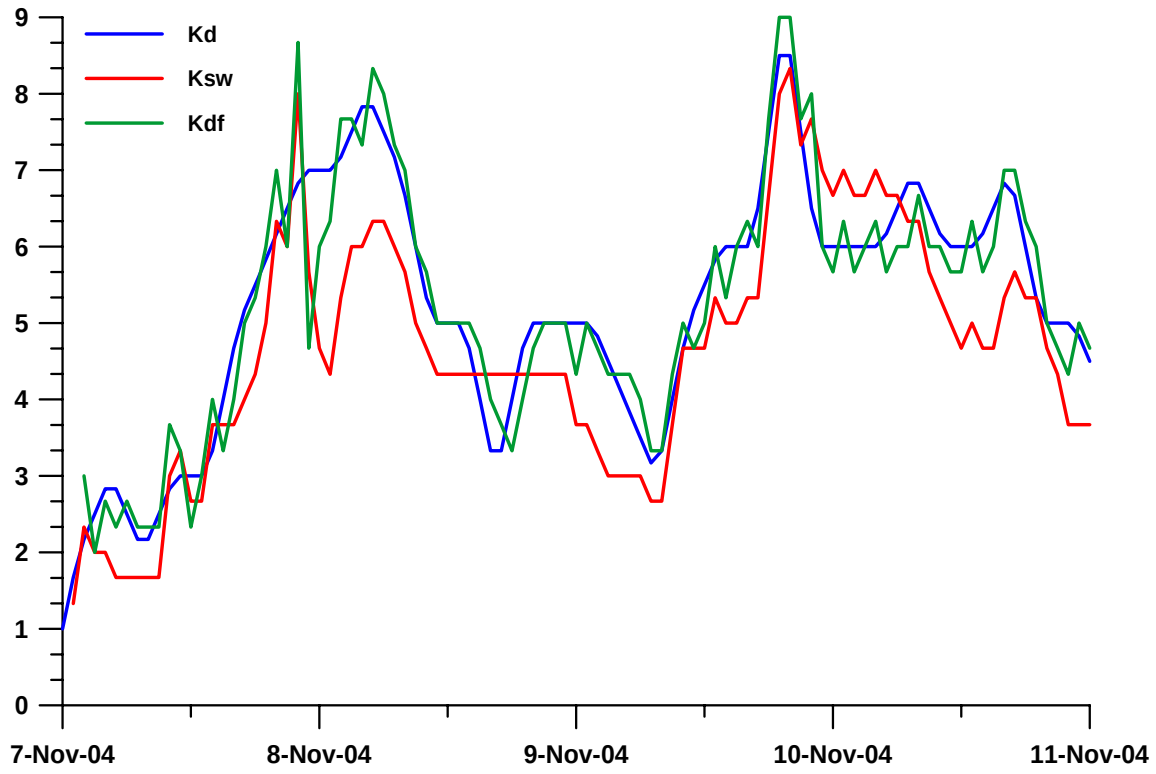

Figure 10: Hourly interpolated Dourbes K (Kdh), solar wind derived Kdsw , and the predicted Kdf during the period 7-10 November 2004.

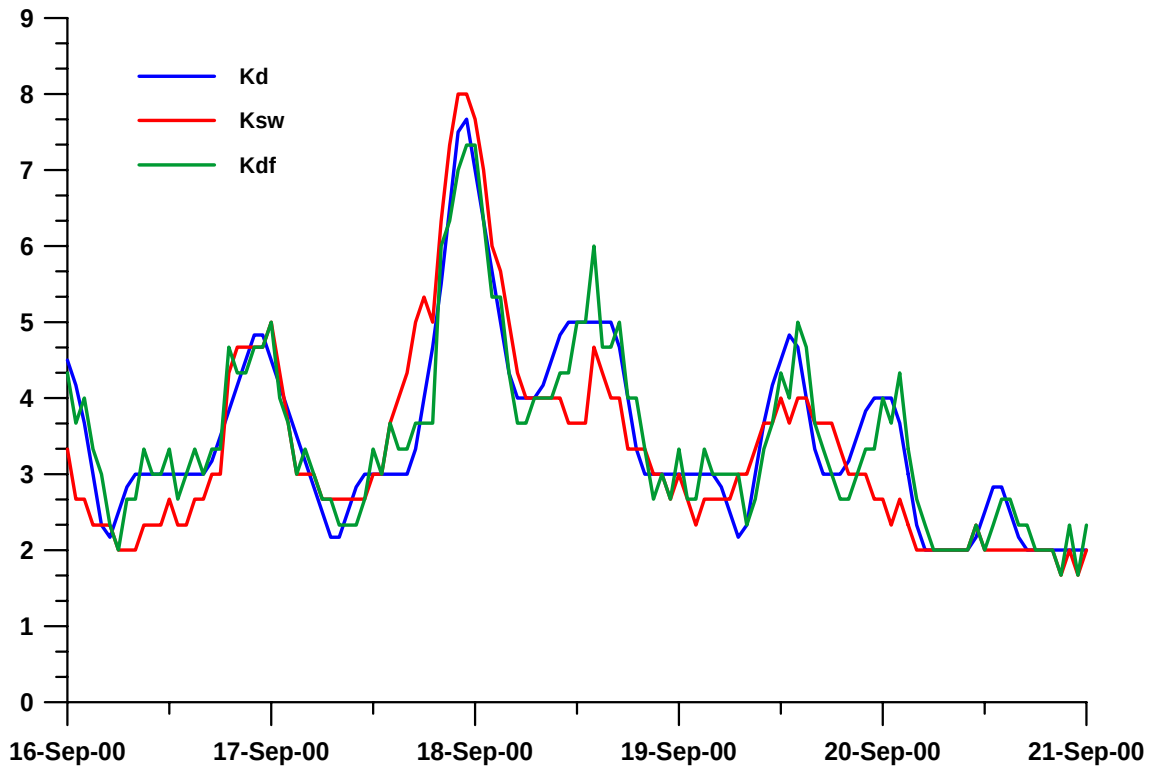

Figure 11: Hourly interpolated Dourbes K (Kdh), solar wind derived Kdsw , and the predicted Kdf during the period 16-20 September 2000.

Figure 11 shows a period with several modest storms (Kd not exceeding 5) and one major storm around 18 September 2000. The agreement between Kdf and Kd is excellent even in the periods when Ksw deviate form Kd.

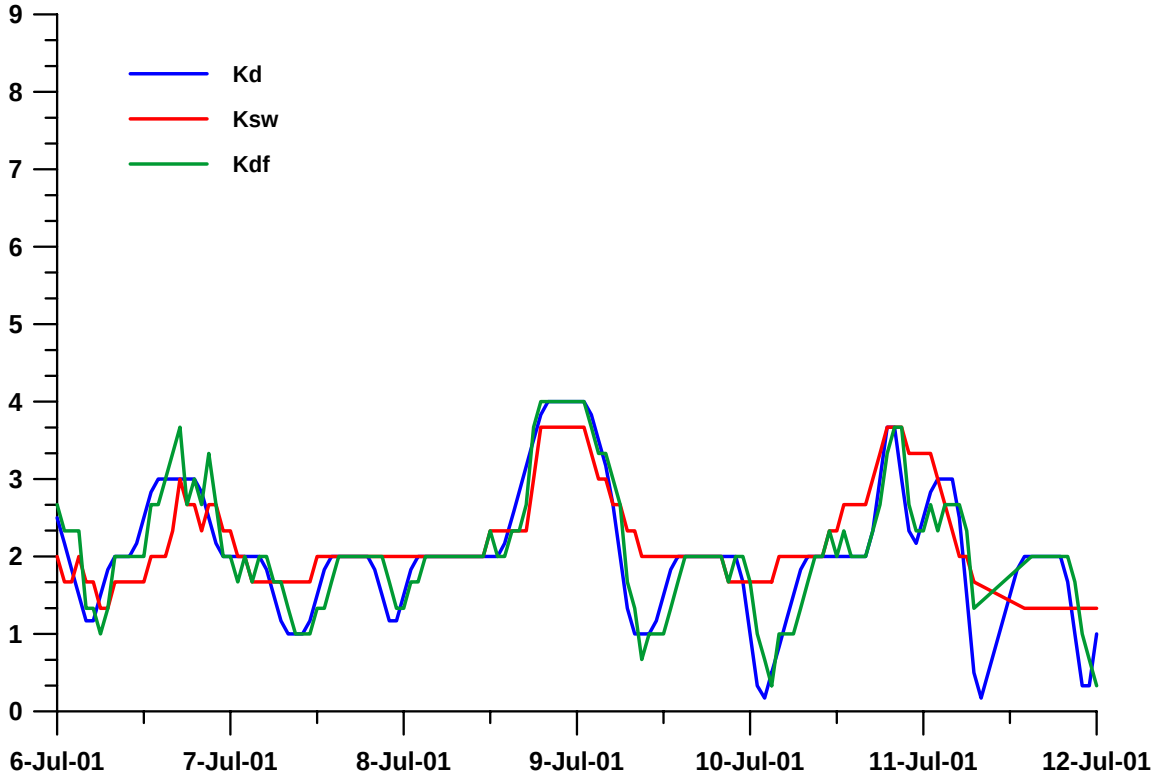

Figure 12: Hourly interpolated Dourbes K (Kdh), solar wind derived Kdsw , and the predicted Kdf during the period 6-11 July 2001.

Figure 12 presents a case at low geomagnetic activity (Kdh below 4). The model HKD is capable to reproduce even the smallest Kd values, when usually the models failed.

3.3 Error assessment and verification of HDK model

The hybrid model was found to perform well. Obviously, the concept of combining the advantages of MAK model and magnetogram-derived K index turns out to be very fruitful. The model error of using equation (5) is 0.38 KU. The high correlation between Kdf and Kd is confirmed in Fig.13 which shows the cross-correlation between the two quantities in the time lags –48, 48 hours. The peak of the function at time lag = 0 is 0.97. In Fig.14 comparison is made between the MAK and HDK model errors as depending on the K magnitude. For the whole range of K, the reduction of model error is 0.2 – 0.5 KU. The error reduction is especially large at higher K values, when predictions are most important for the NGSS applications. The use of the quantity Delta equalizes the error in the whole K range (0-9), which makes HDK model suitable for prediction.

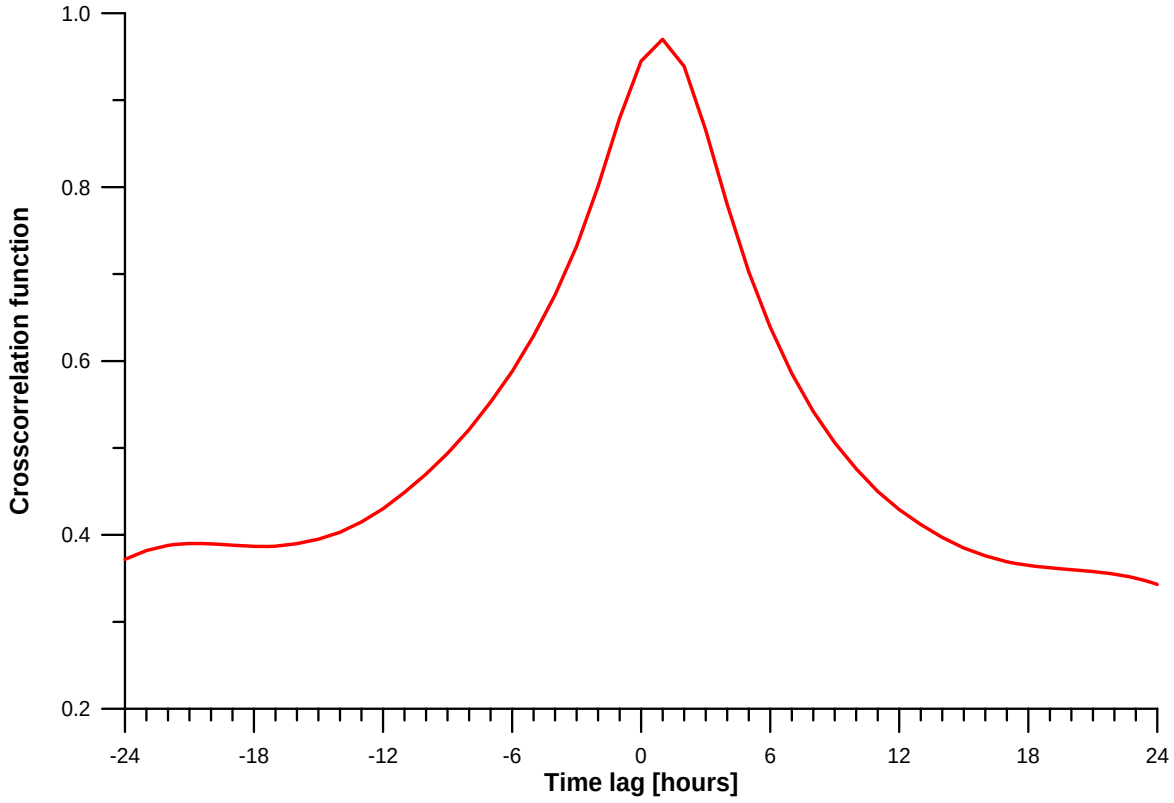

Figure 13: Cross-correlation between input Kd and predicted Kdf

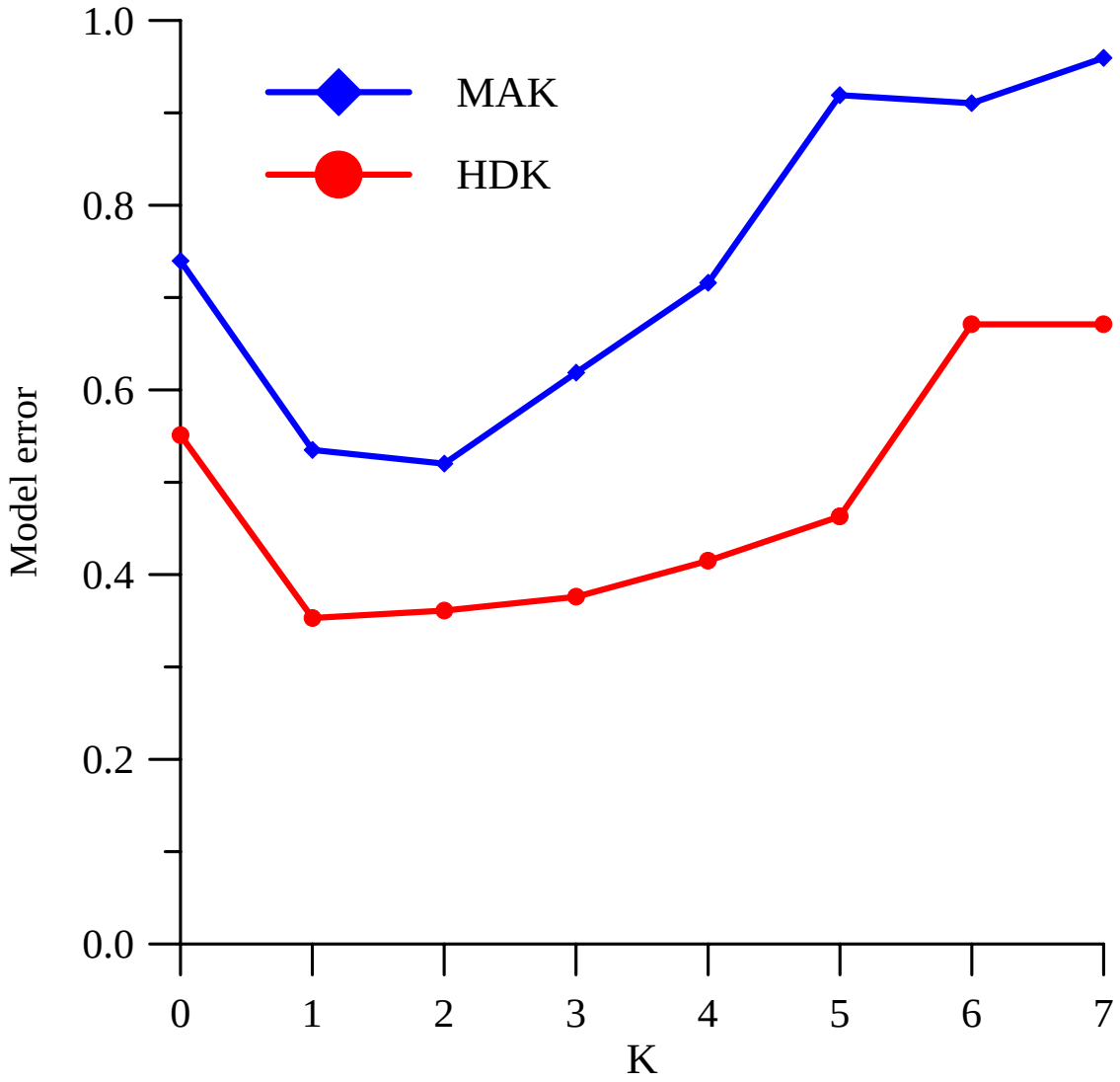

Figure 14: Model errors (standard deviation between the model values and data) of MAK (blue line) and HDK (red line).

As mentioned above, the use of the 3-hour K values in MDelta yields larger model error (Fig.15). The blue curve represents HDK error by using the 3-hour K values only and the red curve is the error by using hourly Kd values. The difference here is of order of 0.4 - 0.6 KU in the whole K range. We have to remind again that interpolated hourly Kd values have lower dispersion than the 3-hour K values, because the interpolated values triple the number of data points without contributing with their own scatter.

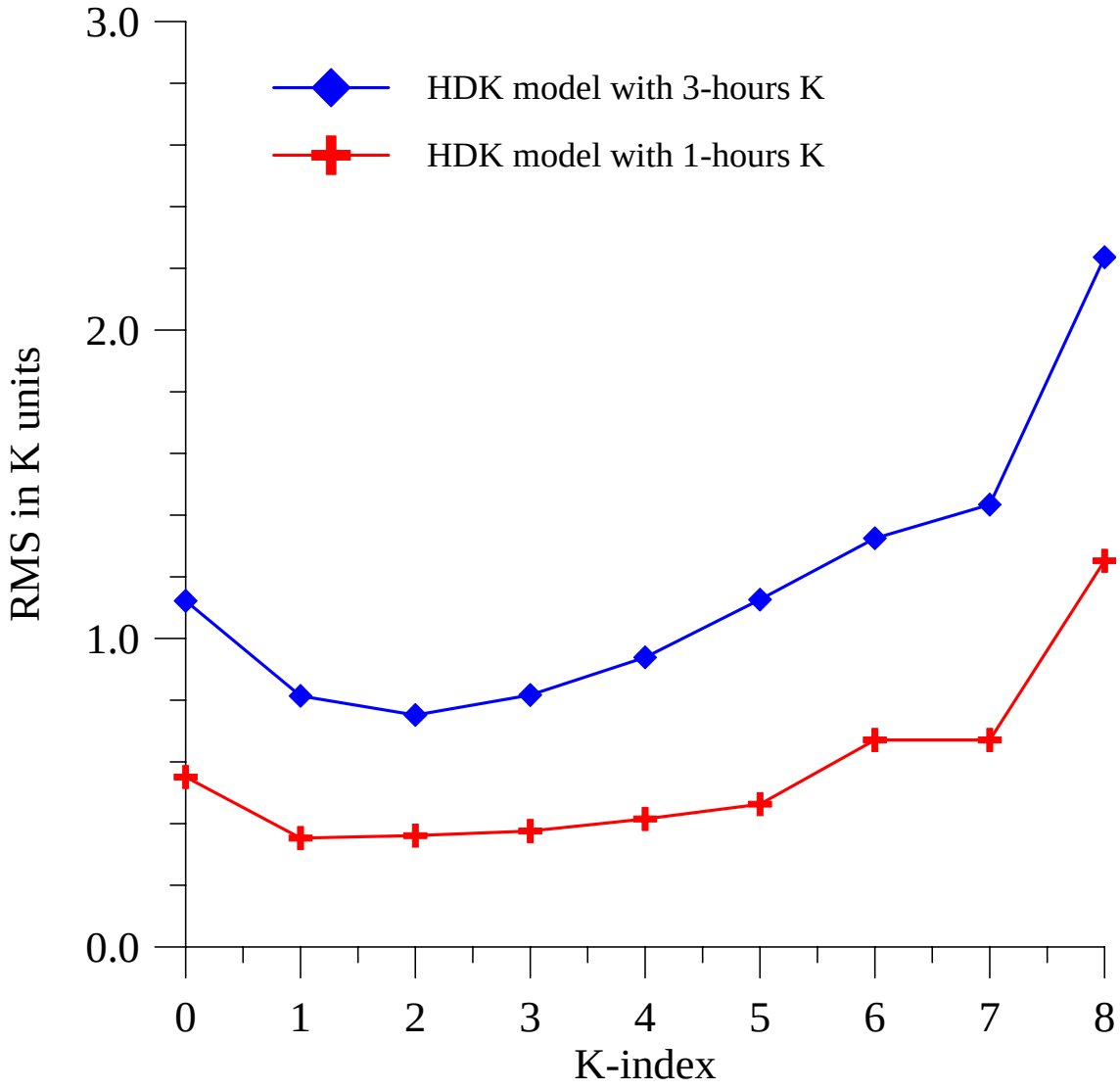

Figure 15: Model errors of HDK using 3-hour Dourbes K index (blue line) and hourly sampled Kd values.

4. Development of the HDK forecasting technique

Kdf values exhibit a good reproducibility. Figure 16 represents its autocorrelation function (blue line). This function is approximated by an exponent (dashed line) with a time constant T=13 hours. To perform prediction of Kdf, we use the same method of Winner-Hopf, as we did for Delta. If we have the last obtained Kdf value at moment *t*, the predicted value Kdfp at the future moment ($t+\tau$) is defined as:

$$Kdfp(t+\tau) = Kdfmean + [Kdf(t)-Kdfmean]*\exp(-\tau/13) \quad (6)$$

Here Kdfmean, in analogy with MDelta, is the average of Kdf values for some hours in the past. The theory requires obtaining Kdfmean as average of 13 past Kdf values, but practically this average is very close the average Kd value in the whole database. That is why we take Kdfmean = 2.42, the same for all predictions. Expression (6) clearly shows the main property of the weighted extrapolation method: with increasing Kdfp tends to the average Kdfmean.

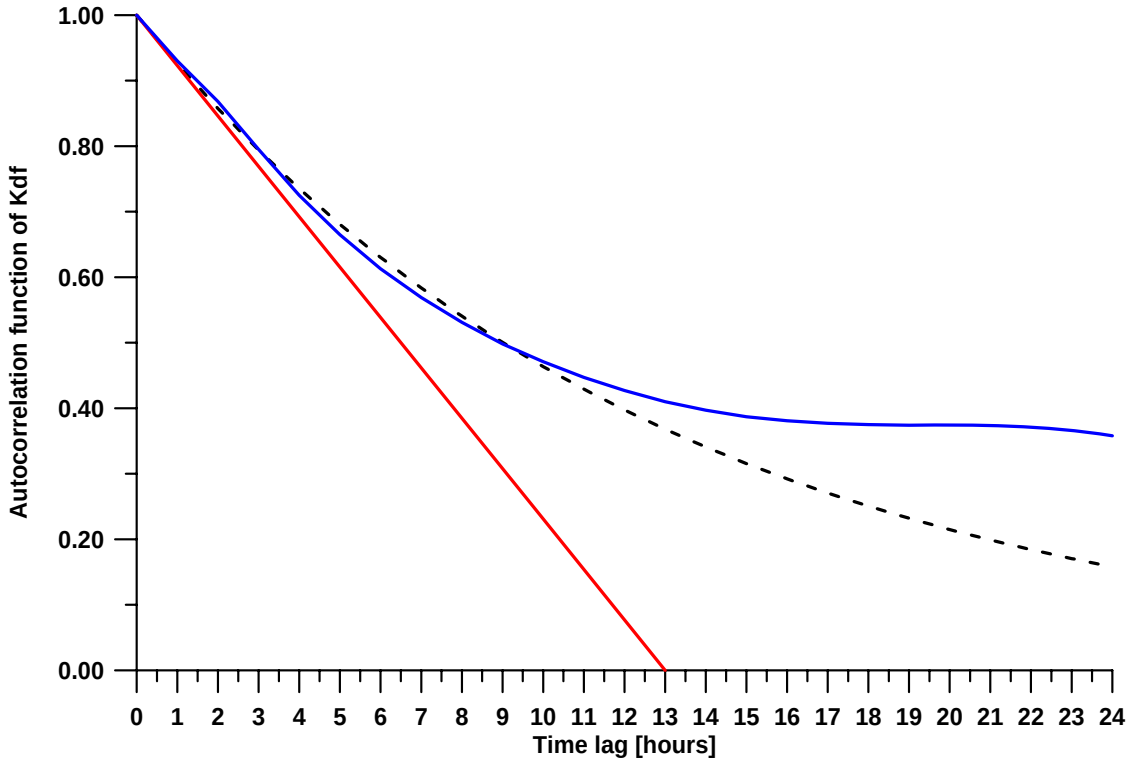

Figure 16: Autocorrelation function of predicted Kdf

Fig.17 shows the prediction error as function of the time of prediction (lead time). Prediction error is defined as standard deviation of predicted Kdfp from the actual Kdf values. Similar to Fig.15, the blue line represents the prediction error obtained by using 3-hour K values and red line representing prediction error obtained by using the hourly (interpolated) Kd values. The algorithm (6) assures that predictions error up to 6 hours ahead does not exceed 1.0 KU. This is, of course, statistically averaged error; for individual cases the error could exceed the theoretical value.

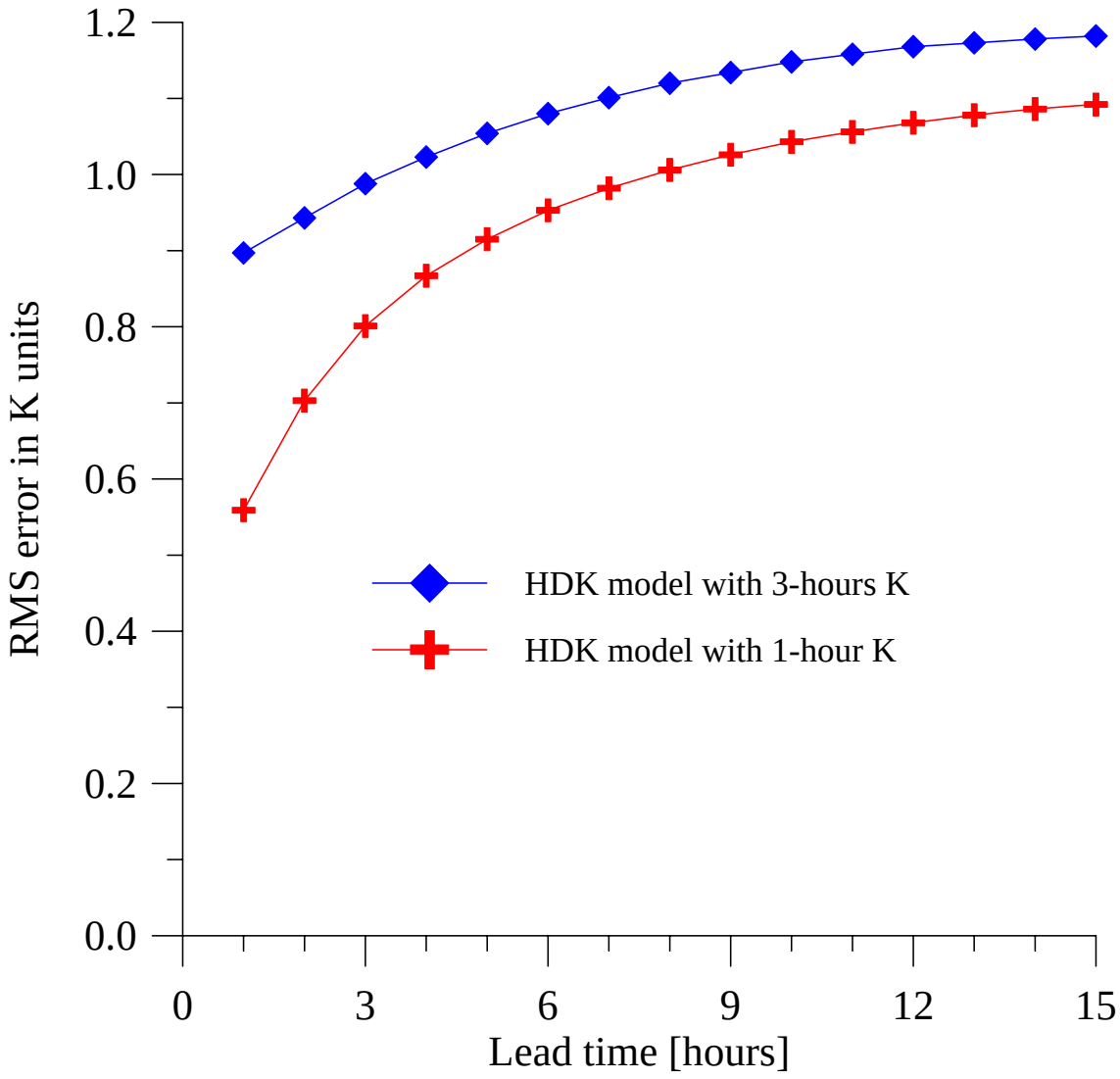

Figure 17: Prediction error (standard deviation between prediction and HDK-defined Kdf for 1-hour (red line) and 3-hour (blue line) sampling rate.

5. Comparison with NN forecasting models

To our knowledge, HDK is the first from the group of empirical (linear regression) models, combining solar wind and ground-based magnetic data for forecasting K index. HDK provides analytical expressions for nowcasting and forecasting K. There exist, however, models based on NN technique, which involve Kp along with solar wind parameters as a input for their calculations. Wing et al. (2005) developed a prediction model, named APL and compared their results with those of three other NN models: Costello (Costello, 1997), NARMAX (Balikhin et al., 2001), and Lund (Boberg et al., 2000). In this comparison, Wing et al. (2005) have provided the correlation coefficient ***r*** only but not RMS errors. Therefore we had to calculate the HDK correlation coefficients and present the comparison in Table 2.

Table 2: Correlation coefficient ***r*** estimated by different models

model	type	***r***	input parameters
Castello	NN	0.75	V, Bz, |**B**|
NARMAX	NN	0.77	V, Bz, |**B**|, Kp(t-3)
Lund	NN	0.77	V, n, Bz,
APL	NN	0.92	V, Bz, |**B**|, nowcast Kp
HDK	empirical	0.95	V, P, Bz, nowcast K

The third column in the table shows the correlation coefficients and the fourth column shows the input parameters of the models; |**B**| is the absolute value of IMF. HDK correlation coefficient is calculated between Kd and Kdf values. It is clear that APL and HDK models show higher ***r*** due to the fact that they combine ground-based K with solar wind parameters. Wing et al. (2005) did not discuss why NARMAX, which also used past Kp values in its calculations, show much smaller correlation coefficient.

Wing et al. (2005) developed two models for predicting Kp 1 and 4 hours ahead. Notice that the NN methods cannot provide prediction formula, function of the lead time as that of (6). Therefore, each prediction time needs separate model. For a lead time of 1 and 4 hours, Wing et al. (2005) provides ***r*** = 0.92 and 0.79 respectively.

Fig.18 presents the correlation coefficient between measured Kd and predicted Kdfp values. The lead time = 0 shows the correlation coefficient between Kd and Kdf. For prediction 1 and 4 hours lead time HDK correlations are lower: 0.88 and 0.68 against 0.92 and 0.79 of Wing et al. (2005). Unfortunately, we do not have enough information to discuss their results: what is database used, how they obtain the nowcast value of Kp, etc.

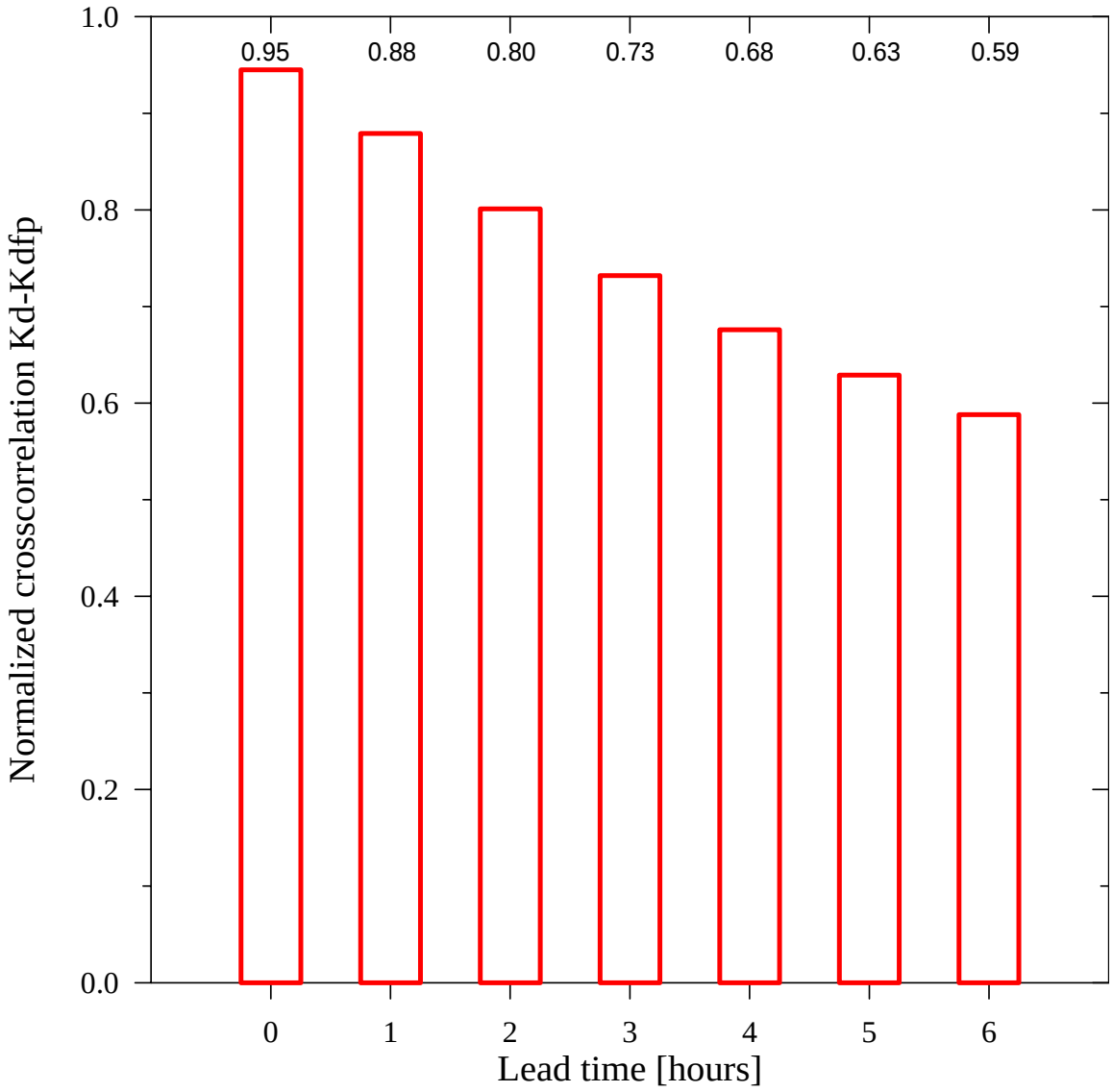

Figure 18: Vertical bars represent correlation coefficient between Kd and predicted Kdfp for a lead time of 1 to 6 hours. The lead time = 0 shows the correlation coefficient between Kd and Kdf. Correlation coefficient magnitude is shown at the upper axis.

6. Prediction and forecast

The meaning of the terms “prediction” and “forecast” used in the report should be clearly distinguished. Prediction is a feature of the model, its capacity to guess a future state. Forecast is a service. Usually, the prediction model is a part of the forecasting software, which has additionally data collection part, I/O modules, data adjustment (correction) part, etc. In some forecasting software, as those developed in the framework of COST 251 and 247 actions (Kutiev et al., 1999; Muhtarov et al., 2001) data adjustment is a separate part from the models. The data adjustment technique is important for forecasting, because it keeps model prediction (statistically average) close to the current data. In these cited papers, data adjustment is performed by correcting the model prediction by the currently obtained difference between prediction and data. In the HDK model this adjustment is part of the algorithm. Using the analogy with the above cited works, we can regard MAK model as prediction model and the other branch as data adjustment. Therefore, the HDK model is both prediction and forecasting software, with I/O part included, of course.

How far ahead the forecasting can be made? Theoretically there is no restriction. Every user can decide how accurate the forecast should be. Fig.17 shows the expected error up to 6 hours lead time. Theoretically, when the lead time approaches the time constant (in our case T=13 hours), the predicted value approaches the average (Kdfmean) and the prediction error becomes close to the standard deviation of data around the average. This means that for predictions, exceeding 10-13 hours, it is better to use directly the average value of Kd (for our database it is 2.42), instead of making complex calculations.

Another question should be also considered. It is about the apparent controversy between the real prediction of Ksw and statistically based prediction of Kdf. Indeed, we cannot predict any changes in K index, if we do not see the respective changes in solar wind parameters. In the MAK section above, we mentioned that the maximum lead time to predict the K changes caused by solar wind is 3 hours. Expression (5) shows the way the prediction works. On the other side, the HDK model value Kdf exhibits a good reproducibility with a time constant of 13 hours. Obviously, both predictions should be used together: we can make hourly predictions up to 13 hours ahead and when sudden change occurs in Ksw, Kdf will react immediately through (5) and changes in turn Kdfp in (6). So, when a change in Ksw occurs, Kdf and the whole prediction series from the 1^{st} to the 13^{th} hours ahead change simultaneously. Every last Kdfp (at the current time *t*) recalculate through (6) all prediction values with τ from 1 to 13.

7. Conclusion

A new model for predicting the geomagnetic K index has been developed based on the combined use of solar wind parameters and ground-based magnetic data. The present approach implements the previously developed solar wind based MAK model by calibrating its values with magnetogram-derived K index. The new model, applied to the K index issued at the Dourbes Center of Geophysics of the Belgian Royal Meteorological Institute, is named Hybrid Dourbes K (HDK) model. The HDK model combines the advantages of predicting the sudden changes of geomagnetic activity induced by solar wind with the longer term predictability of the K index.

The MAK model coefficients were recalculated by fitting the model expression (2) to IMF Bz, solar wind velocity and dynamic pressure and Dourbes K index data. The database used for modeling consists of the hourly values of solar wind parameters and the hourly interpolated 3-hour Dourbes K index (Kd), collected in the period 1998-2004. The HDK model output, the quantity Kdf, is obtained by MAK model output Ksw, corrected with the average difference between several past values of Kd and Ksw. The model error of the new quantity Kdf is found to be 0.38 KU, or nearly twice less than that of the MAK model. Kdf has a good predictability. Prediction made by weighted extrapolation 6 hours ahead carries an error of 1.0 KU, while for the first 1 hour the error is 0.58 KU only.

The results shown in the report clearly demonstrate that the HDK model is capable of forecasing Dourbes K index in an hourly base with better accuracy that the other existing forecasting models.

8. References

Andonov, B., Muhtarov, P., Kutiev, I., 2004. Analogue model relating Kp index to solar wind parameters. J. Atm. Solar-Terr. Phys., 66, 11, 927-932.

Baker, D., Klimas, A., McPherron, R., Buchner, J., 1990. The evolution from weak to strong geomagnetic: an interpretation in terms of deterministic chaos. Geophysical Research Letters 17 (1), 41-44.

Balikhin, M.A., Boaghe, O.M., Billings, S.A., Alleyne, H., 2001. Terrestrial magnetosphere as a non-linear resonator. Geophys. Res. Lett., 28, 1123-1126.

Bartels, J., Heck, N.H., Johnson, H.F., 1939. The three-hour-range index measuring geomagnetic activity. Terr. Magn. Atmos. Electr., 44, 411-454.

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. J. Geophys. Res. 106, 30047-30066.

Boberg, F., Wintoft, P., Lundstedt, H., 2000. Real Kp predictions from solar wind data using neural networks. Physics Chemistry Earth 25 (4), 275-280.

Costello, K.A., 1997. Moving the Rice MSFM into a Real-Time Forecast Mode Using Solar Wind Driven Forecast Models, Ph.D. dissertation, Rice University, Houston, TX, June 1997.

Della-Rose, D.J., Sojka, J.J., Zhu, L. 1999. Resolving geomagnetic disturbances using “K-like” geomagnetic indices with variable time intervals. J. Atmospheric and Solar-Terrestrial Physics 61, 1179-1194.

Hones, E. Jr., 1979. Transient phenomena in the magnetotail and their relation to substorms. Space Sciences Reviews 23, 393-412.

Kamide, Y., Baumjohann, W., Daglis, I., Gonzalez, W., Grande, M., Joselyn, J., McPherron, R., Phillips, J., Reeves, E., Rostoker, G., Sharma, A., Singer, H., Tsurutani, B., Vasyliunas, V., 1998. Current understanding of magnetic storms: Storm-substorm relationships. J. Geophys. Res. 103 (8), 17705-17728.

Klimas, A., Baker, D., Roberts, D., Fairfield, D., Buchner, J., 1992. A nonlinear dynamic analogue model of geomagnetic activity. Journal Geophysical Research 97 (A8), 12253-12266.

Klimas, A., Baker, D., Roberts, D., 1991. Linear prediction filters for linear and nonlinear modeled geomagnetic activity. Geophysical Research Letters 18 (8), 1635-1638.

Kutiev I., Muhtarov, P., Cander, Lj., Levy, M., 1999. Short-term prediction of ionospheric parameters based on autocorrelation analysis, Annali di Geofisica, 42(1), 121-124.

Mayaud, P.N., 1980. Derivation, Meaning and Use of Geomagnetic Indices, Geophys. Monogr. Ser., vol. 22, 155-161, AGU, Washington DC.

Muhtarov, P., Kutiev, I., Cander, Lj., Zolesi, B., de Franceschi, G., Levy, M., Dick, M., 2001. European ionospheric forecast and mapping, Phys. Chem. Earth, 25, 5, 347-351.

Muhtarov P., Kutiev, I., 1999. Autocorrelation method for temporal interpolation and short-term prediction of ionospheric data, Radio Science, 34, 2, 459-463.

Muhtarov, P., Andonov, B., 2000. Improved relationship between the IMF component B_z and Kp index, Bulgarian Geophysical Journal 26 (1-4), 165-172.

Nebdi, H., Warnant, R., Lejeune, S., 2004. Study of the correlation between small-scale TEC variability and geomagnetic activity. Technical Report of Work Package 222, Solar Influences Data Centre Space Weather Pilot Project, ESA contract 16913/03/NL/LvH.

Takahashi, K., Toth, B.A., Olson, J.V., 2001. An automated procedure for near-real-time Kp estimates, J. Geophys. Res., 106, 21017-21032.

Thomsen, M., 2004. Why Kp is a good measure of magnetospheric convection. Space Weather, 2, S11004, doi:10.1029/2004SW000089.

Vassilliadis, D., Sharma, A., Papadopoulos, K., 1993. An Empirical Model Relating the Auroral Geomagnetic Activity to the Interplanetary Magnetic Field. Geophysical Research Letters 20 (16), 1731-1734.

Warnant, R., Kutiev, I., Marinov, P., Bavier, M., Lejeune, S., 2007. Ionospheric and geomagnetic conditions during periods of degraded GPS position accuracy: 1. Monitoring variability in TEC which degrades the accuracy of Real Time Kinematics GPS applications, Adv. Space Res., 39, 875-880.

Warnant, R., Kutiev, I., Marinov, P., Bavier, M., Lejeune, S., 2007. Ionospheric and geomagnetic conditions during periods of degraded GPS position accuracy: 2. RTK events during disturbed and quiet geomagnetic conditions, Adv. Space Res., 39, 881-888.

Wing, S., Johnson, J.R., Jen, J., Meng, C.-I., et al., 2005. Kp forecast models, J. Geophys. Res., 110, A04203, doi: 10.1029/2004JA010500.