Deliverable 4.1

Provision of software tool for forecasting indices

Project acronym: *AFFECTS*
Project title: *Advanced Forecast For Ensuring Communications Through Space*
Grant Agreement number: *263506*
Coordinator: *Volker Bothmer*

Project co-funded by the European Commission,
Seventh Framework Programme

Funding Scheme: *FP7-SPACE-2010-1*

Due date of deliverable:	August 31, 2012
Actual submission date:	August 17, 2012
Start date of the project:	March 1, 2011
Project duration:	3 years

Work package:	4 "Forecasting Tools and Modelling"
Task(s):	4.2
Lead beneficiary for this deliverable:	SRI NASU-NSAU
Editor:	Aleksei Parnowski
Authors:	A. Parnowski, A. Polonska, O. Semeniv, V. Yatsenko, O. Cheremnykh, N. Salnikov, I. Kremenetsky, V. Kuntsevich
Quality reviewer:	Igor Kremenetsky, Volker Bothmer

Project co-funded by the European Commission within the Seventh Framework Programme (2007 - 2013)		
Dissemination level		
PU	**Public**	**PU**
PP	**Restricted to other programme participants (including the Commission Services)**	
RE	**Restricted to a group specified by the consortium (including the Commission Services)**	
CO	**Confidential, only for members of the consortium (including the Commission Services)**	

Document Control

This document has no controlled or maintained paper copies. The master document will be accessible through the AFFECTS website upon approval. Changes to this document will be made available to the scientific community through updates of this document on the AFFECTS website.

Issue record

Version	Date	Author(s)	Reason for modification	Status
1.0	2012-08-17	Aleksei Parnowski	N/A	Submitted to coordinator
1.1	2012-09-30	Volker Bothmer	Quality review	
1.2	2012-10-15	Aleksei Parnowski	revision of comments by PI	Submitted to EC
1.3	2012-02-27	Aleksei Parnowski	revision of comments by the Technical Reviewer	submitted to EC

NOTICE

The contents of this document are the copyright of the AFFECTS consortium and shall not be copied in whole, in part of otherwise reproduced (whether by photographic, reprographic or any other method) and the contents thereof shall not be divulged to any other person or organisation without prior written permission. Such consent is hereby automatically given to all members who have entered into the AFFECTS Consortium Agreement, dated 26/01/2011 and to the European Commission to use and disseminate.

This report has been prepared under the scope of the AFFECTS project in collaboration with the following participants of Work Package 4 "Forecasting tools and modelling" under the Lead of SRI NASU-NSAU:

Space Research Institute of the National Academy of Sciences of Ukraine and National Space Agency of Ukraine, Ukraine (SRI NASU-NSAU)

Phone: +380 93 3264229, Fax: +380 44 5264124, http://www.ikd.kiev.ua/
Aleksei Parnowski, e-mail: parnowski@ikd.kiev.ua

Georg-August-Universität Göttingen, Stiftung Öffentlichen Rechts, Germany (UGOE)

Phone: +49 551 39 5044, Fax: -5043, http://www.astro.physik.uni-goettingen.de/
Volker Bothmer, e-mail: bothmer@astro.physik.uni-goettingen.de

Acknowledgements

We acknowledge GFZ Potsdam for the NRT K_P data, the Kyoto WDC for Geomagnetism for the NRT D_{ST} data, and the SPDF at NASA GSFC for the OMNI data.

The OMNI data were obtained from the GSFC/SPDF/NSSDC FTP server at ftp://nssdcftp.gsfc.nasa.gov/pub/spacecraft_data/omni

We thank the ACE MAG and SWEPAM instruments teams and the ACE Science Center for providing the ACE data.

We acknowledge the CELIAS/MTOF experiment on the Solar Heliospheric Observatory (SOHO) spacecraft. SOHO is a joint European Space Agency, United States National Aeronautics and Space Administration mission.

Contents

List of acronyms

ACE	Advanced Composition Explorer
ASCII	American Standard Code for Information Interchange
CELIAS	Charge, Element, and Isotope Analysis System
CME	Coronal mass ejection
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre)
DoW	Description of Work
FSI	Forecast System Ionosphere
GFZ	GeoForschungZentrum (German Research Centre for Geosciences)
GSFC	Goddard Spaceflight Center
IAGA	International Association of Geomagnetism and Aeronomy
IMF	Interplanetary magnetic field
LASP	University of Colorado Laboratory of Atmospheric and Space Physics
LC	Linear correlation coefficient
MAG	Magnetometer
MSE	Mean square error
MTOF	Mass Time-Of-Flight spectrometer
NASA	National Aeronautics and Space Administration
NICT	National Institute of Information and Communications Technology
NRT	Near-real time
NSSDC	National Space Science Data Center
PC	Personal computer
PE	Prediction efficiency
PM	Proton Monitor
QC	Quality check
RAM	Random access memory
RWC	Regional Warning Centre
SOHO	Solar and Heliospheric Observatory
SPDF	Space Physics Data Facility
SRI NASU-NSAU	Space Research Institute of the National Academy of Sciences of Ukraine and the National Space Agency of Ukraine
SRI RAS	Space Research Institute of the Russian Academy of Sciences
SS	Forecast skill score
SW	Solar wind

SWACI	Space Weather Application Centre Ionosphere
SWE	Solar Wind Experiment
SWEPAM	Solar Wind Electron, Proton, and Alpha Monitor
SWx	Space weather
TEC	Total electron content
USAF	United States Air Force
UT	Universal time
WDC	World Data Centre
WINDMI	Solar Wind Driven Magnetosphere-Ionosphere System
XML	Extensible markup language

1. Introduction

This report describes a "Software tool for forecasting indices", from now on termed the "Geomagnetic forecast software", as designed in order to meet the requirements of Deliverable D4.1 in WP4 "Forecasting tools and modelling" of the AFFECTS project.

The Geomagnetic forecast software consists of three parts allowing users to forecast the geomagnetic indices D_{ST} and K_P in near-real time. The first part is a modelling processor, which constructs the forecasting models of various SWx parameters, including the geomagnetic indices, using historical space- and ground-borne datasets for training. The second part is a forecast processor, which evaluates the models constructed by the modelling processor against near-real-time data, thus producing forecasts. The third part is a preprocessor, which is a maintenance module. It takes care of missing or corrupt input data, provides some basic quality control and generates metadata.

We tried 3 different methods of model construction: the dynamic-information approach with genetic algorithms of structure identification; the approach involving minimax and guaranteed estimation of model parameters with structure enumeration; and the linear regression approach with variance analysis. The latter was chosen as the most appropriate of them for NRT forecasting of the geomagnetic indices.

2. Description of the geomagnetic indices

Most SWx events affect the Earth magnetic field measured by magnetic observatories around the world (geomagnetic field). The changes of the geomagnetic field vary at different locations. For this reason, the individual data are specifically processed in order to eliminate local effects, such as temperature variations of the magnetometer and ground conductivity, before, from a global data set, the geomagnetic indices are determined. There are various indices with different definitions, each having its advantages and drawbacks (e.g. a_P, aa, D_{ST}). Some indices are used for general purpose, some are highly specialised. The definition of geomagnetic indices is supervised by IAGA. The indices relevant to this deliverable are described below.

Planetary K_P index is constructed from 3-hourly peak variations of the most disturbed horizontal magnetic field components, after removing the quiet-day variation pattern [11]. K_P is measured at 13 mid- and high-latitude observatories. 7 observatories are located in Northern Europe, 4 in North America, 1 in Australia and 1 in New Zealand. The range is then converted into a local *K* index (first introduced 1938 for the magnetic observatory Niemegk near Potsdam) taking the values 0 to 9 according to a quasi-logarithmic scale, which is station specific; this is done in an attempt to normalize the frequency of occurrence of the different sizes of disturbances. Since *K* is still remaining a local index, describing disturbances in the vicinity of each observatory, in the next step, according to the geographic and geomagnetic coordinates of the observatories, the annual cycle of daily variations are eliminated through conversion tables using statistical methods. After applying the conversion tables, a standardized index *Ks* for each of the 13 selected observatories is determined. In contrast to the *K* values, the *Ks* index is expressed in a scale of thirds (28 values):

```
0o, 0+, 1-, 1o, 1+, 2-, 2o, 2+, ... , 8o, 8+, 9-, 9o
```

The main purpose of the standardized index *Ks* is to provide a basis for the global geomagnetic index K_P which is the average of a number of “K_P stations”, originally 11. The *Ks* data for the 2 stations Brorfelde and Lovö, as well as for Eyrewell and Canberra, are combined so that their average enters into the final calculation, the divisor thus remaining 11. This index was introduced in 1949 by Bartels at the University of Göttingen. The linear counterpart of the quasi-logarithmic K_P index is the linear a_P index and its daily sum, the A_P index. K_P and a_P are currently maintained by the GFZ German Research Centre for Geosciences [2].

Advantages: depending on a station’s latitude, the index comprises contributions from the 2 major magnetospheric / ionospheric current systems: the intensity and positions of the auroral electrojet and the equatorial magnetospheric ring current. It is probably the most widely used index in the world, extending back in history to 1932. K_P proxies are available in NRT.

Drawbacks: low cadence (3 hours), saturation of values for extreme events, superposition of spatial and intensity variations of electrojet currents.

Disturbance storm-time index (D_{ST}) is the average hourly variation of the horizontal (H) component of the geomagnetic field measured at 4 – 5 since 1997 – near-equatorial magnetic observatories. The quiet solar dynamo (Sq) is subtracted from the variations before averaging. This index is associated primarily with the magnetospheric ring current. The early definition dates back to 1958 by Kertz from the University of Göttingen and to the definitive one provided by Sugiura in 1969. It is currently maintained by the Kyoto WDC for Geomagnetism [1].

Advantages: higher cadence than K_P, but shorter time-series, proxies are available in NRT, less dependent on spatial variations compared to electrojet currents.

Drawbacks: provides no geomagnetic information for high-latitude areas, ring current asymmetries not taken into account, and substantial delays in calculation of science-quality data.

Analysis of the magnetospheric dynamics using Kolmogorov-Sinai entropy and the Lyapunov exponents performed at SRI NASU-NSAU shows that the D_{ST} index can be forecasted up to 4 hours ahead with static models and up to 5 hours with adaptive models. This is important because a good correlation of D_{ST} with ΔTEC variations has been found by DLR Neustrelitz (see AFFECTS D4.4).

3. A brief description of the regression modelling method

Currently the optimal combination of virtues and vices is provided by forecast methods involving time series analysis and data mining. They provide the lead time up to a few hours with high accuracy. The multidimensional time series analysis can be performed using different mathematical methods. The most widely used versions are artificial neural networks, evolutionary algorithms, and correlation analysis. Most of these methods have a common feature: they lead to a regression relation at some point, so it seems natural to skip all the preliminary steps and instantly use the regression analysis. This approach can provide accurate short-term and, to a certain extent, medium-term forecasts and can yield new information about the underlying physical processes. Here only some principal ideas of this method will be described; for specific details the reader is referred to the article [3] for details.

3.1 General ideas

The problem of construction of the forecasting model is considered from the point of view of the systems theory. In its framework, this problem consists of two problems: identification of the model structure, i.e. determining the drivers affecting the forecasted value, and the identification of the model parameters, i.e. determining how exactly the forecasted value depends on these drivers. The second problem is well-studied and has a number of readily available solutions. The first problem, however, is much more difficult and complicated. First of all, even for stationary linear systems, this is an ill-posed problem, because it is both underdetermined (by the number of measured values) and overdetermined (by the number of data points) at the same time. In our case the system (the magnetosphere) is dynamical, which means that its properties change with time, and strongly nonlinear. No formal solutions exist for this case and research activities in this area were very limited until recently. To make this problem manageable, we constrain ourselves to Kolmogorov-Gabor polynomial class of models and treat the uncertainties as stochastic deviations.

Consider a system with an unknown number K_{tot} of inputs u_k and 1 output y (the predictand), which is also one of inputs u_k. At each moment of time t we know only $K < K_{\text{tot}}$ inputs $u_k(t), k = \overline{1,K}$ ($\overline{1,K}$ means all integer numbers from 1 to K inclusively) and the output $y(t)$. Then at an arbitrary step T we can write the predicted value of the system's output in the form

$$y(T+\Theta) = y^*(T+\Theta) + \Delta y(T+\Theta), \quad (1)$$

where Θ is the lead time (the number of hours the forecasted value is ahead of the last measured value), $y^*(T+\Theta)$ is the forecast, and $\Delta y(T+\Theta)$ is the uncertainty, which we assume to be stochastic. We are also forced to assume that all values are distributed normally to be able to use the methods of mathematical statistics, though this is, of course, not always true. We also assume that the statistical properties of the dynamical system do not change on the time scale Θ. The predictand $y^*(T+\Theta)$ is expressed through a partial regression relation:

$$y^*(T+\Theta) = \sum_{i=1}^{m} C_i x_i(T), \quad (2)$$

where $x_i, i = \overline{1,m}$ are the regressors, which are arbitrary functions of input quantities $u_k(t)$, which are already measured at the time T when the forecast is made, $C_i, i = \overline{1,m}$ are the regression coefficients, and m is the number of variable regressors. We choose the regressors x_i in the form of products of powers of the input quantities $u_k(t)$

$$x_i(t) = \prod_{k=1}^{K} u_k^{p_k}(t-l), \; l = \overline{0,L}, \qquad (3)$$

where p_k are powers, which can be equal to zero or any natural number, l is the lag, and L is the maximal lag. We determine the coefficients C_i by the generalised least squares method over a large training sample with equal statistical weights of all points.

The statistical significances of the regressors are determined according to Fisher's F-test. This test allows to separate significant and insignificant regressors. The insignificant parameters are then rejected and the routine is repeated until the regression contains only significant regressors. The execution time of each F-test is proportional to m^4. For this reason, we should not add all the regressors at once, but rather add them gradually, increasing the model complexity in several steps. We use the following routine for the choice of the regressors: first we construct an autoregression model with $x_i(t) = y(t-l), l = \overline{0,L}$, then we add all the other inputs with lags (linear model) and finally we construct nonlinear combinations of the most significant regressors (nonlinear model). To avoid overfitting, we then reject insignificant regressors over a different sample, which we call a tuning sample.

The performance of these models is tested on a third sample, which is called validation sample. The obtained performance scores are compared to those of a persistence model $y_0^*(t+\Theta) = y(t)$. We rate our models in MSE, LC, PE, and SS performance scores. The definitions of these scores are given in [4, 5].

3.2 Application to SWx forecasting

To construct the models we used the OMNI2 database [6], maintained by NASA GSFC, SPDF and NSSDC. We selected 3 samples from this database: years 1976 to 2000 for model training, 2001 to 2008 for model tuning (these two have approximately equal number of data points), and 2011 for performance evaluation. We did not use the data from 2009 and 2010 due to anomalously quiet SW conditions at that time. We also used the archive of the D_{ST} index from Kyoto WDC for Geomagnetism (WDC-C2). We use the datasets with hourly cadence.

Naturally, we used only those parameters, which are available in NRT. They are IMF (total intensity, 2 angular and 3 Cartesian components) and SW plasma parameters (density, proton temperature, and velocity). These quantities enter the models with lags up to 24 hours. We also used the previous values of the predictand with lags up to 27 days (1 Carrington Rotation of the Sun). This allows our models to take into account recurrent space weather events. To simulate diurnal and seasonal variations we also added 4 synthetic values, which are simply sine and cosine functions with periods of 12 hours and 6 months. Together with 3 geomagnetic indices this makes 16 input quantities, but only 14 can enter the model simultaneously.

This method can be used to forecast any values, not only geomagnetic indices. In particular, it can be used to fill the gaps in NRT data. In gap-filling mode we used only the IMF and SW plasma parameters. We assumed that the data gaps most frequently occur simultaneously in all SW plasma parameters, so we used IMF values measured at the same time and older, and SW plasma parameters — 1 hour ago and older.

The performance scores of the developed models are listed in Table 1. These values can still be improved.

Table 1: Performance scores of the developed models over the validation sample (2011)

Predictand	Θ, hrs	MSE, data units		LC, %		PE, %		SS, %
		model	persistence	model	persistence	model	persistence	
D_{ST}, nT	1	3.0	4.0	95.4	96.6	95.9	92.9	41.8
	2	5.1	6.5	91.7	90.8	88.3	81.1	38.2
	3	6.7	8.5	87.6	85.3	80.2	69.8	34.5
	4	7.6	9.2	84.4	80.5	74.1	60.1	35.1
$K_P \cdot 10$	3	7.5	8.4	77.2	76.4	61.9	51.9	20.6
a_P, nT	3	7.6	7.9	72.1	74.5	52.4	48.3	7.9
n, cm^{-3}	0*	1.8	2.1	92.4	91.0	86.3	81.8	24.5
V, km/s	0*	11.0	12.0	97.5	98.9	98.5	98.2	16.6

*Gap-filling mode

3.3 Results of NRT performance tests

The numbers in Table 1 are based on science-quality data. It is, however, much more interesting to see how the method copes with NRT data. Unfortunately, the number of intense SWx events, which happened during the development of the geomagnetic forecast software, is insufficient to conduct proper statistical studies of its performance in NRT. For this reason, we have no choice but to reduce the NRT tests to a case study. As an example we will consider the event of March 7-10, 2012.

The event started on March 7, 2012 with a side blow from a CME generated by an X1.1 flare on March 5 after a long period of southward-directed IMF. This first G2 storm scored the K_P index of 6 and lasted over 12 hours. Around midnight on March 7 an X5.4 flare has generated an Earth-directed CME, almost immediately causing an R3 blackout and an S3 radiation storm. It was soon followed by an X1.3 flare, but whether this one was followed by a CME is not clear. This is a very important issue, since all known superstorms were caused by several sequential CMEs. According to SOHO/CELIAS/MTOF/PM measurements at 1000Z the solar wind velocity increased to 700 km/s with a peak value of 800 km/s. On March 8 at 1105Z the CME hit the Earth causing a 58 nT sudden impulse and instantly increasing K_P to 5. The B_Z component was northward for the most of the event's duration. For this reason, this storm was much weaker than it potentially could be. An additional challenge was brought by the corruption of ACE/SWEPAM data, apparently due to radiation effects.

The results of NRT forecasting made by a preliminary version of the geomagnetic forecast software are shown in Figure 1. The X-axis is the universal time in hours; the Y-axis is the D_{ST} index in nanoteslas. The black solid line with blank circles is the quicklook D_{ST} index from Kyoto WDC for Geomagnetism. Each circle represents a data point. The blue solid line with filled circles is the forecast issued by the preliminary geomagnetic forecast software with 3 hours lead time. Each circle represents a moment of time, for which a forecast was issued.

We compared the performance of the preliminary geomagnetic forecast software with the available online geomagnetic forecast services described in the Section 4.3.

For the D_{ST} forecast the results were the following. Temerin and Li model (3 hours lead time) overestimated the magnitude of the first storm by 50%, and Wintoft D_{ST} model (1 hour lead time) totally missed the sudden impulse, but otherwise they provided a reasonable forecast. The NICT model (1 hour lead time) overestimated the first storm's magnitude by 25% and missed the sudden impulse. Podladchikova model overestimated the magnitude of the first storm by 25%. It should be noted that this last model outputs not the running value of the D_{ST} index itself, but rather an estimation of its peak value. For this reason, the lead time cannot be clearly defined for this model, but typically the storm is forecasted 2-3 hours before the commencement. This approach does not give the information on the onset time.

The preliminary geomagnetic forecast software predicted the storm onset 1-2 hours later than in reality and missed the sudden impulse, and sometimes lagged 1-2 hours behind the measured value, thus providing 1-2 hours lead time instead of 3 hours. Otherwise, the forecast was reasonable.

For the K_P forecasts the models behaved in the following way. Wing model became inoperational since March 7 2012 08:30 UTC and remained in such state through the whole event, and Wintoft K_P model provided unrealistically low K_P forecasts (less then 1 K_P unit).

The preliminary geomagnetic forecast software underestimated the K_P by about 3 times, but remained operational. This underestimation was caused by the erroneous solar wind velocity data provided by the ACE/SWEPAM instrument due to an accompanying radiation storm.

The performance of the Rice University model was not checked.

A detailed account for this case study is given in [7].

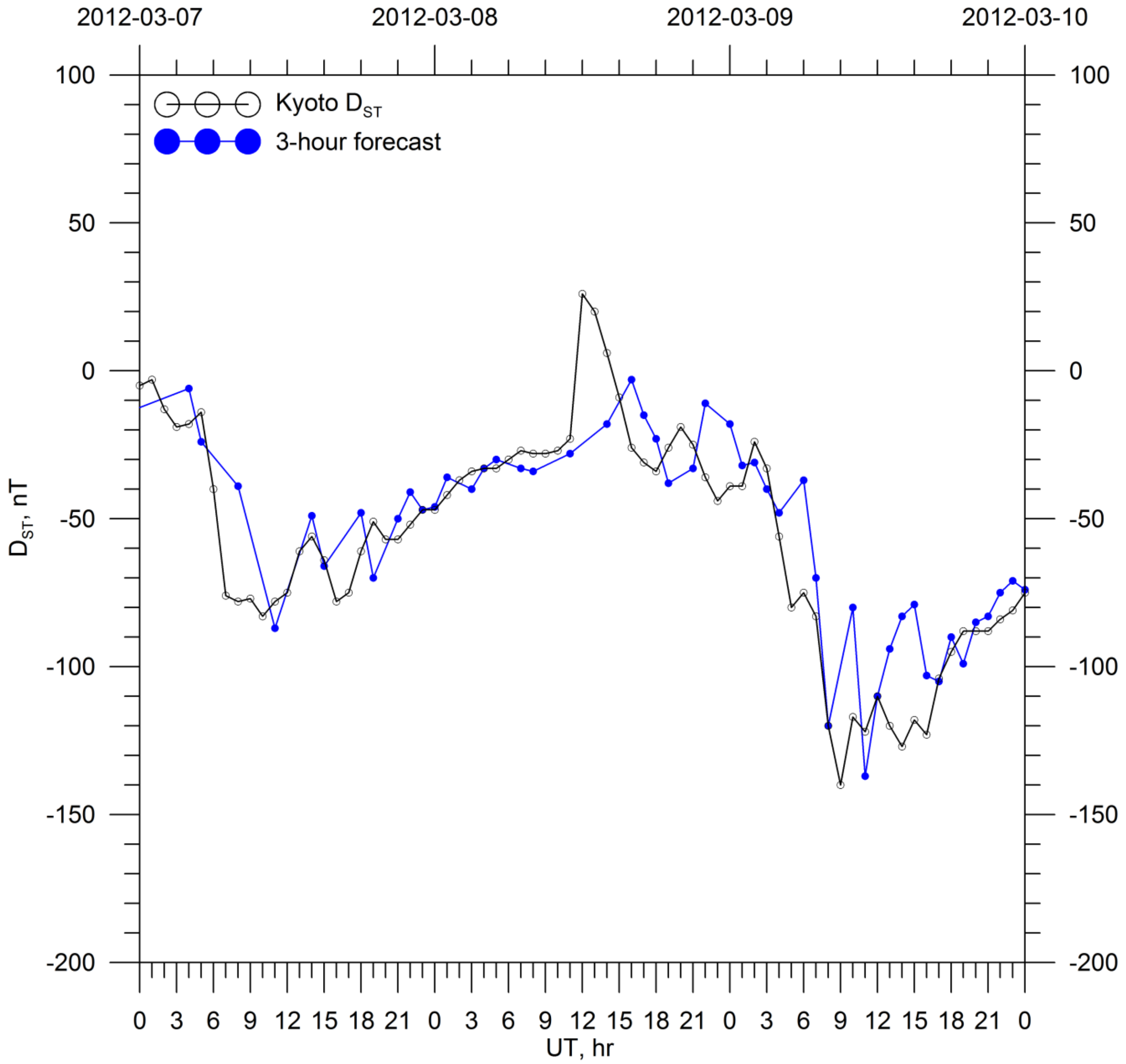

Figure 1: A NRT forecast of March 7-10, 2012 event (D_{ST} index, 3 hours lead time).

4. Provision of software tool for forecasting indices

4.1 Aims and role in the FSI

It is planned to use the geomagnetic forecast software in AFFECTS for 2 distinctive tasks. The first one is to forecast the geomagnetic indices D_{ST}, K_P, and a_P, with D_{ST} being the primary goal. The lead times are 1 to 4 hours for D_{ST}, and 3 hours for K_P and a_P. The second task is to fill gaps in NRT input values, specifically in solar wind velocity and density. This is a highly experimental feature and no warranties are given regarding its performance.

The geomagnetic forecast software will be implemented as an integral part (module 1) of the FSI as part of the Task 5.2 of WP5 as foreseen in the DoW [8]. It will interact with other modules as described in the deliverable report D5.1 [9]. To summarize it briefly, the command interface will be implemented via standard SWACI Job Order Files; the input data will be supplied through FSI modules 4 and 5 and the SWACI PickupPoint; level 1 QC will be performed internally by module 1 and reflected in the metadata; the forecasts together with the relevant metadata will be output to the FSI, which will handle the visualisation, archiving, and level 2 and 3 QC. These forecasts will be also utilized by FSI module 2.

4.2 Software implementation

The geomagnetic forecast software consists of three parts. The first part is the modelling processor, which is used to construct the models using the regression modelling method, described above. It is used offline and will not be integrated into FSI. The second part is the forecasting processor, which is used to evaluate the models using NRT data and to generate metadata. It will be the core component of the FSI Geomagnetic Forecast Module. The third part is a preprocessor, which parses the Job Order File, procures and compiles the input data. It will be an interface part of the FSI Geomagnetic Forecast Module.

The models constructed for AFFECTS are based on the OMNI2 database. We used the samples described in Section 3.2. These models were produced by the modelling processor as ASCII files, which are machine-readable and, to a certain extent, human-readable. There are 2 files per each model: the first one contains information about the model structure and coefficients and the other one about the errors (covariance matrix). The forecasting processor uses these models together with the NRT data to forecast the geomagnetic indices D_{ST} with 1, 2, 3, and 4 hours lead time, K_P and a_P with 3 hours lead time. The required NRT data consists of hourly ACE/MAG and ACE/SWEPAM data, and hourly D_{ST} index or 3-hourly K_P index data in WDC format depending on the predictand. All data are required to cover the current and the previous months. The ACE data originate from NOAA-SWPC, D_{ST} data from Kyoto WDC for Geomagnetism, and K_P data from GFZ Potsdam, so we forecast the official rather than estimated K_P index. These data are delivered through SWACI.

Both the modelling and the forecasting processors are written in FORTRAN90 programming language following a modular approach, which simplifies future modifications and ensures interoperability. The most important progress since the beginning of the project is that both of these systems are fully automatic and operate in hands-off mode without the need for human intervention. This allows constructing much more complex and hence accurate models with the same effort. Both processors were made with as little assumptions concerning the nature of the data as possible, and get this information from a configuration file. This approach ensures that these processors can be used in most typical scenarios without the need to recompile their codes. This is done as a part of the effort to ensure the operability of the system after the end of the project's lifetime, as foreseen in the DoW [8]. In operational environment the forecasting processor will be implemented as a subroutine, callable from within the preprocessor for seamless integration.

Currently, both the modelling and the forecasting processors are single-threaded x86 console win32 applications. This is quite adequate for the forecasting processor, which is fast enough to be used in operational systems – its typical runtime is a few seconds on an average PC regardless of the model complexity. However, this significantly limits the speed of the modelling processor. A typical time required to construct a fairly simple D_{ST} model with 3 hours lead time and cubic nonlinearity is about 40 minutes on an average PC. For complex models with high degree of nonlinearity this time can increase to several days. However, the algorithm allows good parallelization, and can be made multi-threaded at a later stage of development. Another way of improving its speed is switching to an x86-64 compiler, which will allow storing the design matrices in the RAM, which exceed the 2GB limit for 32-bit processes under typical conditions.

The preprocessor is written in C++ programming language following an object-oriented approach. It looks for a XML Job Order File, whose name is provided as a command-line argument. If the command line is absent or the file with provided filename does not exist, it looks for a Job Order File with a standard filename in the same directory. It reads all the "Processing_Parameter" records in the Job Order File, assuming that the last record contains the name of the composite input file. Afterwards it parses the rest of the Job Order File looking for the names of input and output directories. If at least one of the indicated input files exists, the preprocessor will compose the input files into the output file.

Ideally, 8 input files, as described above, are expected. The format of the files should match those on the respective NRT data services. If none of the input files exist or the Job Order File is missing, the preprocessor switches to the autonomous mode, in which it downloads the above input files from the relevant NRT data services and cleans up its directory from old files. At the end the preprocessor outputs a report file, which contains the names of executed tasks and status flags (0 = OK, 1 = error). In operational environment, the preprocessor will call the forecasting module instead of writing a composite input file and handle the output of forecast results and metadata.

4.3 Similar geomagnetic forecast systems

There are several other geomagnetic forecast services available on the web:

1) USAF Weather Agency Wing K_P model (1, 4 hours), http://www.swpc.noaa.gov/wingkp/index.html
2) Rice Space Institute K_P, D_{ST} and AE models (1, 3, 6 hours), http://mms.rice.edu/realtime/forecast.html
3) RWC Sweden K_P model (3 hours), http://rwc.lund.irf.se/rwc/kp/

4) RWC Sweden D_{ST} model (1 hour), http://rwc.lund.irf.se/rwc/dst/
5) NICT D_{ST} model (1 hour), http://www2.nict.go.jp/aeri/swe/swx/ace/nnw/
6) LASP Temerin & Li D_{ST} and AL models (3 hours), http://lasp.colorado.edu/space_weather/dsttemerin/dsttemerin.html
7) SRI RAS Podladchikova D_{ST} model (3 hours), http://spaceweather.ru/node/20

There are also a few older forecast models, including University of Texas at Austin WINDMI model (http://orion.ph.utexas.edu/~windmi/realtime/) and the Naval Research Laboratory model (http://wwwppd.nrl.navy.mil/whatsnew/prediction/), which ceased to operate at the time this report was written.

A systematic comparison between these and the AFFECTS geomagnetic forecast software has not been performed. However, we performed a comparison of the performance of these models with the ones described in this report during the event described in Section 3.3.

5. Summary and conclusion

The following activities were carried out with regard to Deliverable 4.1:

1. The dynamic-information approach with genetic algorithms of structure identification was applied and its performance evaluated.
2. The approach involving minimax and guaranteed estimation of model parameters with structure enumeration was applied and its performance evaluated [10].
3. The regression modelling method was applied and its performance evaluated [3]. It was chosen for implementation in the geomagnetic forecast software.
4. Forecast models for the D_{ST} index with 1, 2, 3, and 4 hours lead time, for the K_P index with 3 hours lead time, and for the a_P index with 3 hours lead time were constructed.
5. The layout of the geomagnetic forecast software was designed taking into account the requirements posed by FSI and SWACI.
6. Models for filling the gaps in solar wind density and velocity NRT data were constructed.
7. The geomagnetic forecast software was developed and tested on archived and NRT data [7].

The AFFECTS geomagnetic forecast software is now ready for offline operations. The integration into FSI will begin in October 2012 in the framework of Task 5.2. Upon integration it will be available to users from inside as well as outside the AFFECTS consortium via the SWACI web portal.

6. Recommendations for future development

Future development is envisaged in two directions: improvement of forecasting algorithms and improvement of software implementation. The currently foreseen improvement of the forecasting algorithm is switching from hourly averages to full-cadence data, which would reduce the bias. The next step in software implementation is parallelizing the code and switching to 64-bit compilers. It is also planned to use SOHO/CELIAS/MTOF/PM and WIND/SWE data in parallel to ACE/SWEPAM data to detect corrupt measurements automatically.

References

1. On D_{ST} Index. http://wdc.kugi.kyoto-u.ac.jp/dstdir/dst2/onDstindex.html
2. Geomagnetic K_P Index. http://www-app3.gfz-potsdam.de/kp_index/description.html
3. Parnowski A.S., 2011. Regression modelling of geomagnetic activity, *J. Phys. Studies*, **15**(2), 2002.
4. Forecast Verification. http://www.cawcr.gov.au/projects/verification/
5. Forecast Verification Glossary. http://www.swpc.noaa.gov/forecast_verification/Glossary.html
6. OMNI2 database. ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/omni
7. Parnowski A.S., Polonska A.Yu., Semeniv O.V., 2012. The women day storm, arXiv.org e-print [arXiv:1203.1951].
8. AFFECTS, Annex I "Description of Work", Project full title: "Advanced Forecast For Ensuring Communications Through Space", Grant Agreement no: 263506, Date: 2011-01-25.
9. System Architecture Document, AFFECTS Deliverable D5.1.
10. Kremenetskiy I.A., Salnikov N.N., 2011. Minimax Approach to Magnetic Storms Forecasting (Dst-index Forecasting), *J. Automation and Information Sciences*, **43**(3), P.67-82. DOI: 10.1615/JAutomatInfScien.v43.i3.70
11. Siebert, M., *Maßzahlen der erdmagnetischen Aktivität*, in *Handbuch der Physik*, vol. **49**/3, 206-275, Springer, Berlin Heidelberg, 1971