

Modelling and Prediction of Global Magnetic Disturbance in Near-Earth Space: a Case Study for Kp Index Using NARX Models

Jose Roberto Ayala Solares¹, Hua-Liang Wei¹, R. J. Boynton¹, Simon N.

Walker¹, and Stephen A. Billings¹

¹Department of Automatic Control and Systems Engineering, Faculty of Engineering, The University of Sheffield, United Kingdom.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2016SW001463

Abstract. Severe geomagnetic disturbances can be hazardous for modern technological systems. The reliable forecast of parameters related to the state of the magnetosphere can facilitate the mitigation of adverse effects of space weather. This study is devoted to the modeling and forecasting of the evolution of the Kp index related to global geomagnetic disturbances. Throughout this work the Nonlinear AutoRegressive with eXogenous inputs (NARX) methodology is applied. Two approaches are presented: i) a recursive sliding window approach, and ii) a direct approach. These two approaches are studied separately and are then compared to evaluate their performances. It is shown that the direct approach outperforms the recursive approach, but both tend to produce predictions slightly biased from the true values for low and high disturbances.

1. Introduction

The operation of many modern technological systems is vulnerable to space weather disturbances. Severe geomagnetic disturbances, such as magnetic storms, can have severe adverse effects on power grids, navigation systems and affect satellite drag. Forecasts of space weather hazards can assist reliable operation of these technological systems. However, a physical model of the solar-terrestrial system that can be used to forecast the evolution of the magnetosphere has not been developed yet, because of the complexity of the dynamical processes involved.

The Kp index is one of the most widely used indices for quantifying geomagnetic activity. It stands for *planetarische Kennziffer*, which means planetary index in German. *Thomsen* [2004] concluded that the Kp index is a good measure of the strength of magnetospheric convection because of its dependence on the latitude of the auroral current region. This index is computed by taking the weighted average of K indices at 13 ground magnetic field observatories. The values of Kp range from 0 (very quiet) to 9 (very disturbed) in 28 discrete steps, resulting in values of 0, 0+, 1-, 1, 1+, 2-, 2, 2+, ..., 9 [*Wing et al.*, 2005].

The Kp index is known to be correlated with solar wind observations [*Newell et al.*, 2007; *Elliott et al.*, 2013]. This has enabled the development of models that attempt to forecast Kp . The most popular models are

based on artificial neural networks, which are considered black-box models [Detman and Joselyn, 1999; Boberg *et al.*, 2000]. For instance, in Wing *et al.* [2005], three neural networks were trained with solar wind data and are now used to nowcast Kp index, producing hourly and 4-hourly forecasts of the Kp , updated every 15 minutes. In Bala and Reiff [2012], an improved neural network was trained using the Boyle index in order to generate 1-, 3- and 6-hour ahead predictions. The Liu Kp model consists of a neural network trained with autoregressive values of Kp and solar wind data, and is able to predict Kp values up to 3.5 hours in advance [Liu *et al.*, 2013]. A comparative study between neural networks and support vector machines was done in Ji *et al.* [2013]. These authors found that the best model is a neural network trained with the same inputs as the Liu Kp model. A probabilistic approach was taken in Wang *et al.* [2015] where the Kp range is divided in 4 groups and 1268 models were compared in terms of accuracy, reliability, discrimination capability, and forecast skill.

In general, there are two approaches for the modelling of magnetic disturbances. The first one consists of the derivation of a mathematical model that contains comprehensive physical insight into all events and processes that take part in space weather dynamics and disturbances [Wei *et al.*, 2004a]. Such a model can then be used to analyse and forecast future events. Nevertheless, it is obvious that such a model is intractable to ob-

tain given the difficulty to fully describe the intrinsic mechanics. The second is a data-based modelling or system identification approach. System identification is an interesting research area with important applications in science and engineering. It consists in finding a mathematical model from discrete-time observational data in order to characterise the behaviour of a system [Billings, 2013; Wei et al., 2004a].

Since the 1980s, several approaches have been developed in the nonlinear realm of system identification given the fact that most real world problems are nonlinear in nature and conventional linear modelling techniques are not sufficient to characterize nonlinear processes of interest [Pope and Rayner, 1994; Billings, 2013]. One popular approach is the Nonlinear AutoRegressive with eXogenous inputs (NARX) methodology, which has been successfully used to identify nonlinear systems [Billings, 2013; Boaghe et al., 2001; Balikhin et al., 2011]. The NARX approach can detect an appropriate model structure and select the most important model terms from a dictionary consisting of a great number of candidate model terms.

In recent years, several variants have been proposed that improve the performance of the original NARX algorithm. Such variations include the use of more complex and flexible predefined functions such as wavelets [Alexandridis and Zapranis, 2013; Billings and Wei, 2005a, b], radial basis functions [Billings et al., 2007; Wei et al., 2007, 2004a], and ridge basis

functions [Wei et al., 2015], together with an improved search mechanism such as the common model structure selection [Wei and Billings, 2008a; Li et al., 2013, 2015], iterative search [Guo et al., 2015a], incorporation of weak derivatives information [Guo et al., 2015b], and other dependency metrics [Koller and Sahami, 1995; Billings and Wei, 2007; Wei and Billings, 2008b; Wang et al., 2013; Speed, 2011; Reshef et al., 2011; Székely et al., 2007; Székely and Rizzo, 2013; Piroddi and Spinelli, 2003; Ayala Solares and Wei, 2015].

The NARX approach produces transparent and interpretable models in which the contribution of each model term to the output signal can be evaluated. This methodology has been previously used to model space weather phenomena. For example, it was used to model the evolution of energetic electrons fluxes at geostationary orbit [Balikhin et al., 2011], to obtain the most influential coupling functions that affect the evolution of the magnetosphere [Boynnton et al., 2011], to predict the *Dst* index using multiresolution wavelet models [Wei et al., 2004a], to build a multiscale radial basis function network to forecast the geomagnetic activity of the *Dst* index [Wei et al., 2007], and to unravel the time varying relationship between the solar wind and the SYM-H index [Beharrell and Honary, 2016], among others. Furthermore, NARX models can be used to compute the generalized frequency response functions (GFRFs) in order to perform frequency domain

analysis [Billings, 2013]. This technique has been used previously to study the spectral properties of the *Dst* index dynamics [Balikhin *et al.*, 2001] and to identify types of nonlinearities involved in the energy storage process in the magnetosphere [Boaghe *et al.*, 2001].

In this paper we investigate the use of NARX models to forecast the *Kp* index. In particular we are interested in forecasts at four different horizons: 3, 6, 12 and 24 hours ahead. To do so, we explore two approaches. The first one consists of a recursive sliding window scheme in which we employ a window period of 6 months to train a model and use it to forecast future values based on previous predictions. The second approach involves the identification of a specific model for each horizon of interest using a fixed dataset of 6 months.

This paper is organised as follows. Section 2 contains a brief summary of the nonlinear system identification methodology, together with a discussion of the Orthogonal Forward Regression algorithm. In section 3 the dataset used for the analysis is described. The NARX recursive approach is developed in section 4, while section 5 is dedicated to the direct approach. Section 6 compares both approaches. The work is concluded in section 7.

2. Nonlinear System Identification

The aim of the system identification modelling approach is to find a model from observational data that can capture as close as possible the relation-

ship between a system input and output [Söderström and Stoica, 1989; Billings, 2013]. Linear system identification has been a popular, widely used approach. Nevertheless, the high complexity of most real-life systems compromise the linearity assumption [Pope and Rayner, 1994]. To overcome this issue, several studies have been performed in the nonlinear realm [Billings, 2013]. In particular, the Nonlinear AutoRegressive with eXogenous inputs (NARX) methodology has become a powerful tool for nonlinear system identification problems [Billings, 2013; Wei et al., 2004a, b; Rashid et al., 2012].

Model structure detection is a challenging task in dynamic system identification. This topic has been studied extensively and a vast amount of information can be found in the literature. Model structure detection has been tackled using different methods, such as clustering [Aguirre and Jácôme, 1998; Feil et al., 2004], the Least Absolute Shrinkage and Selection Operator (LASSO) [Kukreja et al., 2006; Qin et al., 2012], elastic nets [Zou and Hastie, 2005; Hong and Chen, 2012], genetic programming [Sette and Boullart, 2001; Madár et al., 2005], the Orthogonal Forward Regression (OFR) using the Error Reduction Ratio (ERR) approach [Wei et al., 2004b], and the bagging methodology [Ayala Solares and Wei, 2015]. The second step is parameter estimation, which is typically performed using the traditional least squares method, gradient descent and the Metropolis-Hastings

algorithm [Baldacchino et al., 2012; Teixeira and Aguirre, 2011]. The final step is model validation, for which several authors have developed different approaches. In Billings and Voon [1986], a set of statistical correlation tests have been developed that can be used for validation of a nonlinear input-output model.

2.1. Appropriate Model Term Selection

Consider the NARX model:

$$y(k) = f\left(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u)\right) + e(k) \quad (1)$$

where $f(\cdot)$ is a function to be determined from data, $u(k)$ and $y(k)$ are the system input and output signal respectively, $e(k)$ is system noise (with $k = 1, 2, \dots, N$), and the maximum lags for the input and output signals are n_u and n_y [Wei and Billings, 2008b]. Most approaches assume that the function $f(\cdot)$ can be approximated by a linear combination of a predefined set of functions $\phi_i(\varphi(k))$, therefore equation (1) can be expressed in a linear-in-the-parameters form

$$y(k) = \sum_{i=1}^M \theta_i \phi_i(\varphi(k)) + e(k) \quad (2)$$

where θ_i are the coefficients to be estimated, $\phi_i(\varphi(k))$ are the predefined functions that depend on the regressor vector $\varphi(k) =$

$\left[y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u) \right]^T$ of past outputs and inputs, and M is the number of functions in the set.

The most popular algorithm for NARX modelling is the Orthogonal Forward Regression (OFR) algorithm [Guo *et al.*, 2015a; Billings, 2013]. OFR is a stepwise algorithm [Billings *et al.*, 1989], which follows a recursive-partitioning procedures [Dietterich, 2002] to identify a parsimonious NARX model [Wei and Billings, 2008b; Aguirre and Letellier, 2009]. One of the most commonly used NARX models is the polynomial NARX representation, where equation (2) can be written as

$$y(k) = \theta_0 + \sum_{i_1=1}^n \theta_{i_1} x_{i_1}(k) + \sum_{i_1=1}^n \sum_{i_2=i_1}^n \theta_{i_1 i_2} x_{i_1}(k) x_{i_2}(k) + \dots + \sum_{i_1=1}^n \dots \sum_{i_\ell=i_{\ell-1}}^n \theta_{i_1 i_2 \dots i_\ell} x_{i_1}(k) x_{i_2}(k) \dots x_{i_\ell}(k) + e(k) \quad (3)$$

where

$$x_m(k) = \begin{cases} y(k-m) & 1 \leq m \leq n_y \\ u(k-m+n_y) & n_y+1 \leq m \leq n = n_y+n_u \end{cases} \quad (4)$$

and ℓ is the nonlinear degree of the model. A NARX model of order ℓ means that the order of each term in the model is not higher than ℓ . The total number of potential terms in a polynomial NARX model is given by

$$M = \binom{n+\ell}{\ell} = \frac{(n+\ell)!}{n! \cdot \ell!} \quad (5)$$

The OFR algorithm performs a stepwise regression procedure to identify the most significant model terms. To achieve this, it uses the Error Reduction Ratio (ERR) index to measure the significance of each candidate model term [Billings, 2013]. This index can be evaluated by calculating the normalised energy coefficient $C(\mathbf{x}, \mathbf{y})$ between two associated vectors \mathbf{x} and \mathbf{y} [Billings and Wei, 2007]

$$C(\mathbf{x}, \mathbf{y}) = \frac{(\mathbf{x}^T \mathbf{y})^2}{(\mathbf{x}^T \mathbf{x})(\mathbf{y}^T \mathbf{y})} \quad (6)$$

In recent years, several variants of the algorithm have been proposed that modify the predefined functions, the dependency metric or the search mechanism in order to enhance its performance. In particular, the ERR index only detects linear dependencies; so new metrics have been proposed to capture nonlinear dependencies [Billings and Wei, 2007; Wei and Billings, 2008b], i.e. entropy, mutual information [Koller and Sahami, 1995; Billings and Wei, 2007; Wei and Billings, 2008b; Wang et al., 2013], simulation error [Piroddi and Spinelli, 2003] and distance correlation [Ayala Solares and Wei, 2015].

Most of these variants are able to obtain good one-step ahead (OSA) predictions,

$$\hat{y}(k) = f\left(y(k-1), y(k-2), \dots, y(k-n_y), u(k-1), u(k-2), \dots, u(k-n_u)\right) \quad (7)$$

However, because the NARX model (1) depends on past outputs, a more reliable way to check the validity of the model is through the model predicted output (MPO), which uses past predicted outputs to estimate future ones, and to provide details about the stability and predictability range of the model,

$$\hat{y}(k) = f\left(\hat{y}(k-1), \hat{y}(k-2), \dots, \hat{y}(k-n_y), u(k-1), u(k-2), \dots, u(k-n_u)\right) \quad (8)$$

In the literature, some authors have adapted the original OFR algorithm to optimize directly the MPO in order to obtain a better long-term prediction. However these modified versions tend to be computationally expensive during the feature selection step, and a much better alternative is to use the iterative or ultra-orthogonalisation approach [Guo *et al.*, 2015a, b].

Furthermore, in many real applications, multiple step-ahead predictions are of interest. For an autonomous system (e.g. a time series process without external input), the system output value at the current time instant k , i.e. $y(k)$, may be predicted using previous observations at time instants $k-s$, $k-s-1$, etc., and the predicted value $\hat{y}(k)$ is called the s -step

ahead prediction. For an input-output system, the s -step ahead prediction $\hat{y}(k)$ is often estimated using previous output measurements $y(k-s)$, $y(k-s-1)$, ... , and previous input values $u(k-1)$, $u(k-2)$, ... , etc. So, for an input-output system model, the s -step ahead prediction is defined with respect to the system output; it is actually still one-step ahead prediction with respect to the system input.

3. Dataset description

Every three hours throughout the day, 13 ground-based magnetic field observatories located at geomagnetic latitudes between 48° and 63° around the world, record the largest magnetic change that their instruments measure. This change is denoted as the K index, which is given on a quasi-logarithmic scale from 0 ($< 5nT$) to 9 ($> 500nT$) [Boberg *et al.*, 2000]. The average of these observations is known as the Kp index. This determines how disturbed the Earth's magnetosphere is on a scale that goes from 0 (very quiet) to 9 (very disturbed) in 28 discrete steps, resulting in values of 0, 0+, 1-, 1, 1+, 2-, 2, 2+, ..., 9 [Boberg *et al.*, 2000; Wing *et al.*, 2005]. In this paper, these values are rescaled to be represented by the numbers 0, 0.3, 0.7, 1, ..., 9.

In general, large Kp values can indicate a more active terrestrial magnetosphere due to a solar storm, or a sudden rearrangement of the Earth's magnetosphere due to the solar wind .

The datasets used in this paper consist of the variables shown in Table 1. These were measured during the year 2000. The inputs are taken from the low resolution OMNI dataset, which consist of hourly average near-Earth solar wind magnetic field and plasma data from several spacecraft in geocentric or L1 (Lagrange point) orbits. The data period used for this study employed four spacecraft: IMP 8, WIND, Geotail and ACE. The output is the Kp index which, as mentioned before, is measured every three hours. In order to match the time resolutions between the input and output signals, the observed Kp values are interpolated to 1-hour resolution by simply repeating the last measured value during the next two hours.

Given that the variable of interest is the Kp index, its distribution for year 2000 is shown in Fig. 1. This highlights that high values of Kp are rare, which makes their prediction a challenging task.

4. Sliding Window Models and Recursive Predictions

This approach uses a window of a fixed length to build a single model using the data within the window frame as the training set. This model is used to make 3-, 6-, 12- and 24-hour ahead predictions based on the model simulated values (i.e. the model predicted outputs - MPOs) as shown in equation (8). Once this is done, the window is moved forward by one time step, a new model is built and subsequently used to forecast the next 3-,

6-, 12- and 24-hour ahead Kp values. This way, the training and validation sets are mutually exclusive.

Every time the window frame moves forward, a new NARX model is trained. The training process uses the adaptive orthogonal search algorithm described in *Billings and Wei* [2008]. A nonlinear model term and variable selection procedure proposed in *Wei et al.* [2004b] was applied, and numerical experimental results suggested that $n_y = 4$ and $n_u = 2$ was an appropriate choice. Accordingly, the NARX model structure is given by

$$\begin{aligned} \widehat{Kp}(k) = f & \left(Kp(k-1), \dots, Kp(k-4), \right. \\ & V(k-1), V(k-2), Bs(k-1), Bs(k-2), \\ & VBs(k-1), VBs(k-2), p(k-1), p(k-2), \\ & \left. \sqrt{p(k-1)}, \sqrt{p(k-2)} \right) \end{aligned} \quad (9)$$

where $f(\cdot)$ is chosen to be a polynomial of nonlinear degree $\ell = 2$, $Kp(k)$ is the measured Kp index at time k , and $\widehat{Kp}(k)$ is the predicted Kp index at time k . In our analysis, the window length is of 6 months, therefore the initial training and validation sets correspond to the first and second half of year 2000, respectively. As the window frame moves forward, the validation set size decreases. The reason to choose a window length of 6 months is because for the NARX methodology typically just a few hundred data samples are required to estimate a model, which can be important

in many applications where it is unrealistic to perform long experiments [Billings, 2013].

The results for this approach are shown in Fig. 2. Here it can be seen that there is a bias for low and high magnetic disturbances. Furthermore, for high values of Kp ($Kp \geq 8$) the error bars become odd and difficult to interpret. This is due to the fact that there are very few occurrences of high-value Kp indexes, so few predictions are made in such cases and hence they tend to be underpredicted. Such characteristics have been previously reported in *Detman and Joselyn* [1999]; *Boberg et al.* [2000], where it is argued that a model will perform well for the most common training values, while predictions for others will be poor.

To quantify our results, the root mean squared error (RMSE), correlation coefficient (ρ) and prediction efficiency (PE) are computed. The latter is defined as

$$PE = 1 - \frac{\sigma_{error}^2}{\sigma_{measured}^2} \quad (10)$$

where $\sigma_{measured}^2$ is the variance of the measured Kp values, and σ_{error}^2 is the variance of the error between the measured Kp values and the predicted ones. These metrics are shown in Table 2.

The error time series for each of the four horizons of interest are shown in Fig. 3. It can be seen that the error is notoriously high at the middle of

July. Fig. 4 shows a glimpse of this period where it can be seen that high activity of the terrestrial magnetosphere was recorded between July 13th-17th, 2000. Such an activity was not properly forecasted by this approach. In addition, Table 3 shows a statistical summary of the error time series in Fig. 3. In general, it can be concluded that this approach tends to overpredict the Kp index given that both the median and the mean are negative. Furthermore, as the number of hours to predict ahead increases, the forecasts are less accurate because the interquartile range (1st quartile - 3rd quartile) increases.

5. Direct approach

The second modelling technique investigated in this paper involves use of what is termed the direct approach. Instead of training a model many times and using it recursively to calculate forecasts, the direct approach obtains a separate model for a horizon h of interest. In such a case, equation (1) becomes

$$y(k) = f\left(y(k-h), y(k-h-1), \dots, y(k-h-n_y), u(k-1), u(k-2), \dots, u(k-n_u)\right) + e(k)$$

The main advantage of the direct approach is that it only requires the computation of h -step-ahead predictions. This means that the output at the present time k , $y(k)$, is predicted using the past values $y(k-h)$,

$y(k-h-1), \dots, y(k-h-n_y), u(k-1), u(k-2), \dots, u(k-n_u)$, where it is assumed that these are known [Wei *et al.*, 2007].

In similarity to the sliding window approach, $n_y = 4$ and $n_u = 2$ were chosen, and the training process uses the adaptive orthogonal search algorithm described in Billings and Wei [2008]. Accordingly, the NARX model structure is given by

$$\begin{aligned} \widehat{Kp}(k) = f & \left(Kp(k-h), \dots, Kp(k-h-3), \right. \\ & V(k-1), V(k-2), Bs(k-1), Bs(k-2), \\ & VBs(k-1), VBs(k-2), p(k-1), p(k-2), \\ & \left. \sqrt{p(k-1)}, \sqrt{p(k-2)} \right) \end{aligned} \quad (11)$$

where $f(\cdot)$ is chosen to be a polynomial of nonlinear degree $\ell = 2$, $Kp(k)$ is the measured Kp index at time k , and $\widehat{Kp}(k)$ is the predicted Kp index at time k . In this analysis, the first six months of year 2000 are used for training while the second half of the year is used for validation.

The models identified by the NARX methodology for each horizon are listed below:

- 3-hours ahead

$$\begin{aligned}
\widehat{Kp}(k) = & 0.325543Kp(k-3) - 0.000043\mathbf{V}(\mathbf{k}-1)\sqrt{p(\mathbf{k}-1)} + 0.673034\mathbf{Bs}(\mathbf{k}-1) \\
& - 0.164093\mathbf{Bs}(\mathbf{k}-1)\sqrt{p(\mathbf{k}-1)} - 0.000003V(k-1)^2 + 0.000217V(k-1) \cdot Bs(k-2) \\
& - 0.006701Bs(k-1) \cdot Bs(k-2) - 0.005810Bs(k-1) \cdot p(k-2) - 2.179360 \\
& + 0.753122\sqrt{p(k-1)} + 0.006105V(k-1) - 0.387292VBs(k-1) \\
& + 0.136271VBs(k-1)\sqrt{p(k-1)} \tag{12}
\end{aligned}$$

- 6-hours ahead

$$\begin{aligned}
\widehat{Kp}(k) = & -0.000191\mathbf{V}(\mathbf{k}-1)\sqrt{p(\mathbf{k}-1)} + 0.852464\mathbf{Bs}(\mathbf{k}-1) + 0.158716Kp(k-6) \\
& - 0.172607\mathbf{Bs}(\mathbf{k}-1)\sqrt{p(\mathbf{k}-1)} + 0.000340V(k-1) \cdot Bs(k-2) - 0.000003V(k-1)^2 \\
& - 0.058229Bs(k-1) \cdot Bs(k-2) - 0.007989Bs(k-1) \cdot p(k-2) \\
& + 0.009495\sqrt{p(k-1)}\sqrt{p(k-2)} + 0.000962p(k-1) \cdot p(k-2) - 2.749889 \\
& + 0.007744V(k-1) + 0.958020\sqrt{p(k-1)} - 0.514336VBs(k-1) \\
& + 0.113874VBs(k-1)\sqrt{p(k-1)} + 0.011219VBs(k-1)^2 + 0.009277VBs(k-2)^2 \\
& + 0.032255Bs(k-2) \cdot VBs(k-1) \tag{13}
\end{aligned}$$

- 12-hours ahead

$$\begin{aligned}
\widehat{Kp}(k) = & 0.001618\mathbf{V}(\mathbf{k}-1)\sqrt{\mathbf{p}(\mathbf{k}-1)} + 0.748665\mathbf{Bs}(\mathbf{k}-1) - 0.268901\mathbf{Bs}(\mathbf{k}-1)\sqrt{\mathbf{p}(\mathbf{k}-1)} \\
& - 0.000229Kp(k-12) \cdot V(k-1) + 0.203764Bs(k-2) - 0.017656Kp(k-12) \cdot p(k-1) \\
& - 0.007676Bs(k-1) \cdot Bs(k-2) - 1.606480 - 0.000324V(k-1) \cdot p(k-1) \\
& - 0.003098p(k-1)\sqrt{p(k-1)} + 0.000312V(k-1) \cdot Bs(k-1) + 0.265301Kp(k-12) \\
& + 0.003683V(k-1) + 0.286045p(k-1) - 0.012219Kp(k-12) \cdot VBs(k-2) \\
& - 0.531734VBs(k-1) + 0.195865VBs(k-1)\sqrt{p(k-1)} \quad (14)
\end{aligned}$$

- 24-hours ahead

$$\begin{aligned}
\widehat{Kp}(k) = & 0.000066\mathbf{V}(\mathbf{k}-1)\sqrt{\mathbf{p}(\mathbf{k}-1)} + 0.838922\mathbf{Bs}(\mathbf{k}-1) - 0.213375\mathbf{Bs}(\mathbf{k}-1)\sqrt{\mathbf{p}(\mathbf{k}-1)} \\
& + 0.011558Kp(k-24) \cdot Bs(k-2) - 0.000004V(k-1)^2 + 0.269300Bs(k-2) \\
& - 0.066312Bs(k-1) \cdot Bs(k-2) - 3.080364 + 1.023429\sqrt{p(k-1)} \\
& + 0.008776V(k-1) + 0.014446Bs(k-1)^2 - 0.573961VBs(k-1) \\
& + 0.120880VBs(k-1)\sqrt{p(k-1)} + 0.007968Kp(k-25)^2 + 0.012127Bs(k-2)^2 \\
& + 0.034862VBs(k-1) \cdot VBs(k-2) - 0.121102VBs(k-2) \\
& + 0.000240V(k-2) \cdot VBs(k-1) \quad (15)
\end{aligned}$$

The results of this approach are shown in Fig. 5. They display a similar pattern to the sliding window approach, i.e. there is a bias for low and high magnetic disturbances, and the error bars for high values of Kp ($Kp \geq 8$)

become less meaningful. Once again, these characteristics are due to the uncommon number of cases of high values of the Kp index compared with the most common Kp values related with quiet activity periods of the magnetosphere.

To quantify our results, the root mean squared error (RMSE), correlation coefficient (ρ) and prediction efficiency (PE) are computed. These metrics are shown in Table 4.

The error for each of the four horizons of interest is respectively shown in Fig. 6. Once again, there is a notoriously high error at the middle of July, corresponding to a period of high geomagnetic activity, as mentioned above.

A glimpse of this period is shown in Fig. 7. In addition, Table 5 shows a statistical summary of the error time series in Fig. 6. In general, it can be concluded that on average, this approach tends to slightly underpredict the Kp index given that the means are positive. Furthermore, as the number of hours to predict ahead increases, the forecasts are less accurate because the interquartile range (1st quartile - 3rd quartile) increases, as expected.

6. Model Comparison

A quick view to Tables 3 and 5 shows that the direct approach provides better forecasts than the sliding window approach because the means and medians are closer to zero, and the interquartile ranges are smaller. To better visualise this difference, a randomly selected 30-day interval on the

second half of year 2000 is taken. The features dynamics are shown in Fig. 8.

The model forecasts using both approaches during this 30-day interval are shown in Figs. 9-10.

To quantify our results, the root mean squared error (RMSE), correlation coefficient (ρ) and prediction efficiency (PE) are computed. These metrics are shown in Table 6.

These results show that better forecast accuracy is obtained by the direct approach. This is an expected result given that the sliding window approach uses model predicted outputs from a single model, and long-term forecasts tend to deviate from true values as time goes on. On the other hand, the direct approach uses a separate model for each horizon and relies on single calculations for h -step-ahead predictions. However, both approaches show that predictions for low and high disturbances are slightly biased from the true values. This observation is coincident with previous findings reported in *Detman and Joselyn* [1999] and *Boberg et al.* [2000], where a model will perform well for the most common training values, while predictions for others will be poor. Another explanation is that this comes as a trade-off for using a regression model to predict a categorical output variable.

Comparing the results obtained with those presented in *Wing et al.* [2005], the values of the two model performance metrics (i.e. prediction perfor-

mance and correlation coefficient) calculated from our results are slightly lower. This may be explained from several factors: i) all the data for all input and output variables used for model estimation in this study are raw data sampled hourly where no pre-processing (e.g. smoothing, interpretation, etc.) was performed; ii) the model input variables used in this work are not exactly the same as those used in previous studies; iii) some coefficients required by the models, for example the maximum lags of the input and output variables, may need to be optimised further. Note that one of the objectives of this work is to generate compact transparent models to show how Kp index depends on solar wind parameters and geomagnetic field indices, and then use such models to do further analysis including forecast. As shown in models 12-15, an important contribution obtained from the direct approach is that there are three significant model terms that are shared by all the models. These are shown in bold in the equations above. The values of the three terms, together with the Kp index, are normalised, and the associated scatter plots are shown in Fig. 11 (note that the normalisation of the values is just to facilitate the visualisation and comparison of the scatter plots). The importance of the first selected common model term $V(k-1)\sqrt{p(k-1)}$ may be roughly explained by its relevance with Kp when measuring the correlation coefficient ($\rho = 0.6149$). Model terms

ranked later would normally not be so important as the top ones, and their correlation with the Kp signal becomes very weak.

The importance of the model terms selected in the equations above is not always measured by the values or amplitude of these model terms. A model term with a high (or low) value does not necessarily mean a high (or low) value in Kp index, as its change is an outcome of combined and weighted interactions of many lagged input variables. Experience shows that top model terms can reflect the major varying trend of the output signals, while model terms ranked later can be useful in revealing local and relatively minor changes. While the role of solar wind speed and dynamic pressure as drivers of the Kp index has been confirmed by previous studies, this work provides some further information with an explicit format of these input variables, showing what kind of interactions of these drivers make a contribution to the change of the Kp index. This is important for further understanding and analysis of the dependent relationship of the Kp index on solar wind speed and dynamic pressure, etc.

7. Conclusion

In this paper, we have applied the NARX modelling methodology to the forecasting of the Kp index. We have obtained a number of models using two different implementation approaches: namely, recursive prediction approach based on sliding windows and a direct approach which can di-

rectly generate h -hour ahead predictions ($h = 3, 6, 12$ and 24 in our case studies). In general, good forecasts were obtained for both short and long-term prediction using the estimated NARX models, but the direct approach outperforms the recursive approach. Nevertheless, both approaches tend to show that predictions for low and high disturbances are slightly biased from the true values. As previously reported, such a bias is a result of the uneven distribution in the output signal, and in this paper, the use of a regression model to predict a categorical output variable may also play a role on this matter. An interesting property obtained from the direct approach is a set of significant model terms that are shared by all the models, regardless of the time horizon of interest. While the role of the solar wind speed and dynamic pressure as drivers of the Kp index has been confirmed by previous studies, the present work produced some further information showing the relative contributions made by these drivers to the changes in the Kp index. This is useful for further understanding the relationship of the Kp index to solar wind. It was noticed that the values of prediction performance and correlation coefficient relating to our models are slightly lower than those reported by *Wing et al.* [2005] and possible reasons were briefly discussed. To improve the overall performance of the proposed models, the following investigations will be considered: i) in dynamic regression modeling, the choice of maximum lags for both input and output variables is

important, therefore it is highly desirable to introduce an adaptive maximum lag selection scheme to accommodate the non-stationary features of both the input and output sequences; ii) the raw Kp data are categorical, the recently developed logistic-NARX model may be more appropriate to deal with the Kp prediction problem where the output signal is categorical while the input variables are time-continuous.

8. Acknowledgement

The authors acknowledge the financial support to J. R. Ayala Solares from the University of Sheffield and the Mexican National Council of Science and Technology (CONACYT). The authors gratefully acknowledge that part of this work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/I011056/1 and Platform Grant EP/H00453X/1, and EU Horizon 2020 Research and Innovation Programme Action Framework under grant agreement No 637302.

The authors would like to thank Professor Michael Balikhin for his constructive comments and suggestions for improving the paper. The solar wind and Kp index data were provided by the OMNIWeb Service (<http://omniweb.gsfc.nasa.gov>), which is gratefully acknowledged.

Plots and numerical results of the paper are available from the author on request; more results will be available from a publicly accessible website later.

References

Aguirre, L. A., and C. J ac ome (1998), Cluster analysis of NARMAX models for signal-dependent systems, in *Control Theory and Applications, IEE Proceedings-*, vol. 145, pp. 409–414, IET.

Aguirre, L. A., and C. Letellier (2009), Modeling Nonlinear Dynamics and Chaos: A Review, *Mathematical Problems in Engineering*, 2009(35).

Alexandridis, A. K., and A. D. Zaprani (2013), Wavelet neural networks: A practical guide, *Neural Networks*, 42(0), 1–27, doi: <http://dx.doi.org/10.1016/j.neunet.2013.01.008>.

Ayala Solares, J., and H.-L. Wei (2015), Nonlinear model structure detection and parameter estimation using a novel bagging method based on distance correlation metric, *Nonlinear Dynamics*, pp. 1–15.

Bala, R., and P. Reiff (2012), Improvements in short-term forecasting of geomagnetic activity, *Space Weather*, 10(6), doi:10.1029/2012SW000779, s06001.

Baldacchino, T., S. R. Anderson, and V. Kadirkamanathan (2012), Structure detection and parameter estimation for NARX models in a unified EM framework, *Automatica*, 48(5), 857–865.

Balikhin, M. A., O. M. Boaghe, S. A. Billings, and H. S. C. K. Alleyne (2001), Terrestrial magnetosphere as a nonlinear resonator, *Geophysical Research Letters*, 28(6), 1123–1126, doi:10.1029/2000GL000112.

Balikhin, M. A., R. J. Boynton, S. N. Walker, J. E. Borovsky, S. A. Billings, and H. L. Wei (2011), Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit, *Geophysical Research Letters*, 38(18), doi:10.1029/2011GL048980, 118105.

Beharrell, M. J., and F. Honary (2016), Decoding Solar Wind–Magnetosphere Coupling, *Space Weather*, doi:10.1002/2016SW001467, 2016SW001467.

Billings, S. A. (2013), *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*, Wiley.

Billings, S. A., and W. S. F. Voon (1986), Correlation based model validity tests for nonlinear models, *International Journal of Control*, 44(1), 235–244.

Billings, S. A., and H.-L. Wei (2005a), The wavelet-NARMAX representation: a hybrid model structure combining polynomial models with multiresolution wavelet decompositions, *International Journal of Systems Science*, 36(3), 137–152.

Billings, S. A., and H.-L. Wei (2005b), A New Class of Wavelet Networks for Nonlinear System Identification, *Neural Networks, IEEE Transactions on*, 16(4), 862–874.

Billings, S. A., and H.-L. Wei (2007), Sparse Model Identification Using a Forward Orthogonal Regression Algorithm Aided by Mutual Information,

IEEE Transactions on Neural Networks, 18(1), 306–310.

Billings, S. A., and H.-L. Wei (2008), An adaptive orthogonal search algorithm for model subset selection and non-linear system identification, *International Journal of Control*, 81(5), 714–724.

Billings, S. A., S. Chen, and R. J. Backhouse (1989), The identification of linear and non-linear models of a turbocharged automotive diesel engine, *Mechanical Systems and Signal Processing*, 3(2), 123–142.

Billings, S. A., H.-L. Wei, and M. A. Balikhin (2007), Generalized multiscale radial basis function networks, *Neural Networks*, 20(10), 1081–1094.

Boaghe, O. M., M. A. Balikhin, S. A. Billings, and H. Alleyne (2001), Identification of nonlinear processes in the magnetospheric dynamics and forecasting of Dst index, *Journal of Geophysical Research: Space Physics*, 106(A12), 30,047–30,066, doi:10.1029/2000JA900162.

Boberg, F., P. Wintoft, and H. Lundstedt (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(4), 275 – 280, doi:http://dx.doi.org/10.1016/S1464-1917(00)00016-7.

Boynton, R. J., M. A. Balikhin, S. A. Billings, H. L. Wei, and N. Ganushkina (2011), Using the NARMAX OLS-ERR algorithm to obtain the most influential coupling functions that affect the evolution of the magnetosphere, *Journal of Geophysical Research: Space Physics*, 116(A5), doi:

10.1029/2010JA015505, a05218.

Detman, T., and J. Joselyn (1999), Real-time Kp predictions from ACE real time solar wind, *AIP Conference Proceedings*, 471(1), 729–732.

Dietterich, T. G. (2002), Machine Learning for Sequential Data: A Review, in *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 15–30, Springer.

Elliott, H. A., J.-M. Jahn, and D. J. McComas (2013), The Kp index and solar wind speed relationship: Insights for improving space weather forecasts, *Space Weather*, 11(6), 339–349, doi:10.1002/swe.20053.

Feil, B., J. Abonyi, and F. Szeifert (2004), Model order selection of nonlinear input-output models—a clustering based approach, *Journal of Process Control*, 14(6), 593–602.

Guo, Y., L. Guo, S. Billings, and H.-L. Wei (2015a), An iterative orthogonal forward regression algorithm, *International Journal of Systems Science*, 46(5), 776–789, doi:10.1080/00207721.2014.981237.

Guo, Y., L. Z. Guo, S. A. Billings, and H.-L. Wei (2015b), Ultra-orthogonal forward regression algorithms for the identification of non-linear dynamic systems, *Neurocomputing*.

Hong, X., and S. Chen (2012), An elastic net orthogonal forward regression algorithm, in *16th IFAC Symposium on System Identification*, pp. 1814–1819.

Ji, E.-Y., Y.-J. Moon, J. Park, J.-Y. Lee, and D.-H. Lee (2013), Comparison of neural network and support vector machine methods for Kp forecasting, *Journal of Geophysical Research: Space Physics*, 118(8), 5109–5117, doi:10.1002/jgra.50500.

Koller, D., and M. Sahami (1995), Toward optimal feature selection, in *In 13th International Conference on Machine Learning*.

Kukreja, S. L., J. Lofberg, and M. J. Brenner (2006), A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification, in *System Identification*, vol. 14, pp. 814–819.

Li, P., H.-L. Wei, S. A. Billings, M. A. Balikhin, and R. Boynton (2013), Nonlinear model identification from multiple data sets using an orthogonal forward search algorithm, *Journal of Computational and Nonlinear Dynamics*, 8(4), 10.

Li, Y., H.-L. Wei, S. A. Billings, and P. Sarrigiannis (2015), Identification of nonlinear time-varying systems using an online sliding-window and common model structure selection (CMSS) approach with applications to EEG, *International Journal of Systems Science*, pp. 1–11.

Liu, Y., B. X. Luo, and S. Q. Liu (2013), Kp Forecast Models Based on Neural Networks, *Manned Spaceflight*, 27(2).

Madár, J., J. Abonyi, and F. Szeifert (2005), Genetic programming for the identification of nonlinear input-output models, *Industrial & Engineering*

Chemistry Research, 44(9), 3178–3186.

Newell, P. T., T. Sotirelis, K. Liou, C.-I. Meng, and F. J. Rich (2007),

A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables, *Journal of Geophysical Research: Space Physics*, 112(A1), n/a–n/a, doi:10.1029/2006JA012015, a01206.

Piroddi, L., and W. Spinelli (2003), An identification algorithm for polynomial NARX models based on simulation error minimization, *International Journal of Control*, 76(17), 1767–1781.

Pope, K. J., and P. J. W. Rayner (1994), Non-linear system identification

using Bayesian inference, in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, vol. IV, pp. 457 – 460.

Pope, K. J., and P. J. W. Rayner (1994), Non-linear system identification

using Bayesian inference, in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, vol. IV, pp.

457 – 460.

457 – 460.

Qin, P., R. Nishii, and Z.-J. Yang (2012), Selection of NARX models estimated using weighted least squares method via GIC-based method and

L1-norm regularization methods, *Nonlinear Dynamics*, 70(3), 1831–1846,

doi:10.1007/s11071-012-0576-y.

doi:10.1007/s11071-012-0576-y.

Rashid, M. T., M. Frasca, A. A. Ali, R. S. Ali, L. Fortuna, and M. G. Xibilia

(2012), Nonlinear model identification for Artemia population motion,

Nonlinear Dynamics, 69(4), 2237–2243, doi:10.1007/s11071-012-0422-2.

Reshef, D. N., Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean,

P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti (2011),

Detecting novel associations in large data sets, *Science*, 334(6062), 1518–1524, doi:10.1126/science.1205438.

Sette, S., and L. Boullart (2001), Genetic programming: principles and applications, *Engineering Applications of Artificial Intelligence*, 14(6), 727–736.

Söderström, T., and P. Stoica (1989), *System Identification*, Prentice Hall.

Speed, T. (2011), A correlation for the 21st century, *Science*, 334(6062), 1502–1503, doi:10.1126/science.1215894.

Székely, G. J., and M. L. Rizzo (2013), Energy statistics: A class of statistics based on distances, *Journal of Statistical Planning and Inference*, 143(8), 1249–1272.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007), Measuring and testing dependence by correlation of distances, *The Annals of Statistics*, 35(6), 2769–2794.

Teixeira, B. O., and L. A. Aguirre (2011), Using uncertain prior knowledge to improve identified nonlinear dynamic models, *Journal of Process Control*, 21(1), 82–91.

Thomsen, M. F. (2004), Why K_p is such a good measure of magnetospheric convection, *Space Weather*, 2(11), n/a–n/a, doi:10.1029/2004SW000089, s11004.

Wang, J., Q. Zhong, S. Liu, J. Miao, F. Liu, Z. Li, and W. Tang (2015), Statistical analysis and verification of 3-hourly geomagnetic activity probability predictions, *Space Weather*, *13*(12), 831–852, doi: 10.1002/2015SW001251, 2015SW001251.

Wang, S., H.-L. Wei, D. Coca, and S. A. Billings (2013), Model term selection for spatio-temporal system identification using mutual information, *International Journal of Systems Science*, *44*(2), 223–231.

Wei, H.-L., and S. A. Billings (2008a), Constructing an overall dynamical model for a system with changing design parameter properties, *International Journal of Modelling, Identification and Control*, *5*(2), 93–104.

Wei, H.-L., and S. A. Billings (2008b), Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information, *International Journal of Modelling, Identification and Control*, *3*(4), 341–356.

Wei, H.-L., S. A. Billings, and M. A. Balikhin (2004a), Prediction of the Dst index using multiresolution wavelet models, *Journal of Geophysical Research: Space Physics*, *109*(A7).

Wei, H.-L., S. A. Billings, and J. Liu (2004b), Term and variable selection for non-linear system identification, *International Journal of Control*, *77*(1), 86–110.

Wei, H.-L., D.-Q. Zhu, S. A. Billings, and M. A. Balikhin (2007), Forecasting the geomagnetic activity of the Dst index using multiscale radial basis function networks, *Advances in Space Research*, *40*(12), 1863–1870.

Wei, H.-L., M. A. Balikhin, and S. N. Walker (2015), A new ridge basis function neural network for data-driven modeling and prediction, in *Computer Science & Education (ICCSE), 2015 10th International Conference on*, pp. 125–130, IEEE.

Wing, S., J. R. Johnson, J. Jen, C.-I. Meng, D. G. Sibeck, K. Bechtold, J. Freeman, K. Costello, M. Balikhin, and K. Takahashi (2005), Kp forecast models, *Journal of Geophysical Research: Space Physics*, *110*(A4), a04203.

Zou, H., and T. Hastie (2005), Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.

Table 1. Dataset variables.

Variable	Symbol	Description
Input	V	Solar wind speed [km/s]
	B_s	Southward interplanetary magnetic field [nT]
	VB_s	Southward interplanetary magnetic field [$VB_s = V \cdot B_s/1000$]
	p	Solar wind pressure [nPa]
	\sqrt{p}	Square root of solar wind pressure
Output	K_p	K_p index (variable of interest)

Table 2. Evaluation metrics for each of the four horizons of interest obtained with the sliding window approach.

Horizon	RMSE	ρ	PE
3	0.7935	0.8590	0.7359
6	0.9014	0.8159	0.6598
12	0.9513	0.7991	0.6225
24	0.9624	0.7972	0.6149

Table 3. Statistical summary for the error time series shown in Fig. 3.

Statistic	Forecast			
	3	6	12	24
Minimum	-3.1980	-2.7180	-2.5570	-2.4860
1st Quartile	-0.5394	-0.6409	-0.7106	-0.7270
Median	-0.0843	-0.1032	-0.1391	-0.1497
Mean	-0.0303	-0.0491	-0.0711	-0.0830
3rd Quartile	0.4084	0.4440	0.4454	0.4447
Maximum	4.7170	5.7520	6.3050	6.3900

Table 4. Evaluation metrics for each of the four horizons of interest obtained with the direct approach.

Horizon	RMSE	ρ	PE
3	0.7593	0.8711	0.7585
6	0.8328	0.8424	0.7096
12	0.8623	0.8305	0.6895
24	0.8719	0.8265	0.6824

Table 5. Statistical summary for the error time series shown in Fig. 6.

Statistic	Forecast			
	3	6	12	24
Minimum	-2.3890	-2.6040	-2.8780	-3.5550
1st Quartile	-0.4436	-0.4842	-0.4910	-0.5073
Median	-0.0138	-0.0096	0.0068	-0.0107
Mean	0.0373	0.0433	0.0575	0.0446
3rd Quartile	0.4625	0.4950	0.5210	0.5005
Maximum	4.6880	5.6140	5.9440	5.7260

Table 6. Evaluation metrics for each of the four horizons of interest using the sliding window and direct approaches during a 30-day interval between September and October of year 2000.

Horizon	Approach	RMSE	ρ	PE
3	Window	0.8308	0.8874	0.7828
	Direct	0.7582	0.9156	0.8287
6	Window	0.9298	0.8628	0.7283
	Direct	0.8053	0.9071	0.8105
12	Window	0.9546	0.8728	0.7138
	Direct	0.8537	0.9054	0.7919
24	Window	0.9569	0.8804	0.7125
	Direct	0.8588	0.8875	0.7831

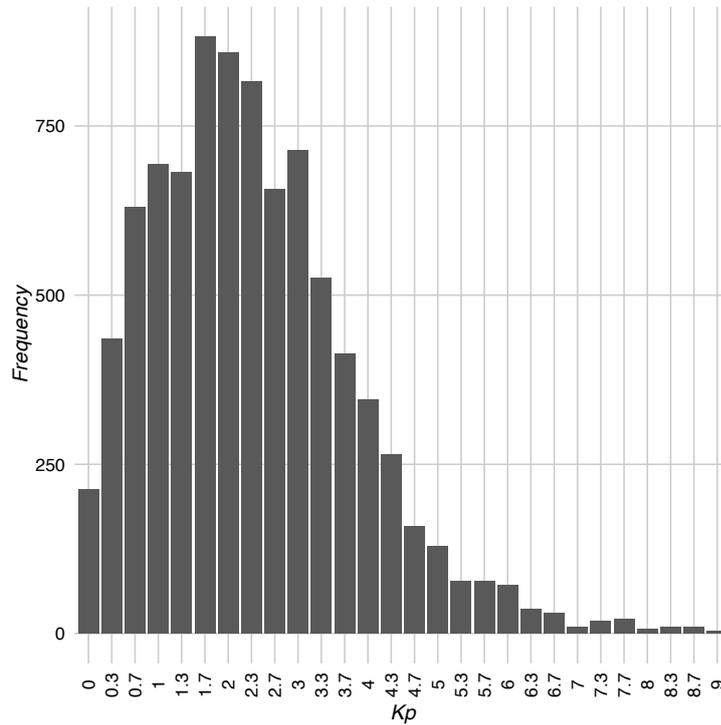


Figure 1. Histogram of the Kp index for year 2000. High Kp values of 5 to 9 are relatively rare.

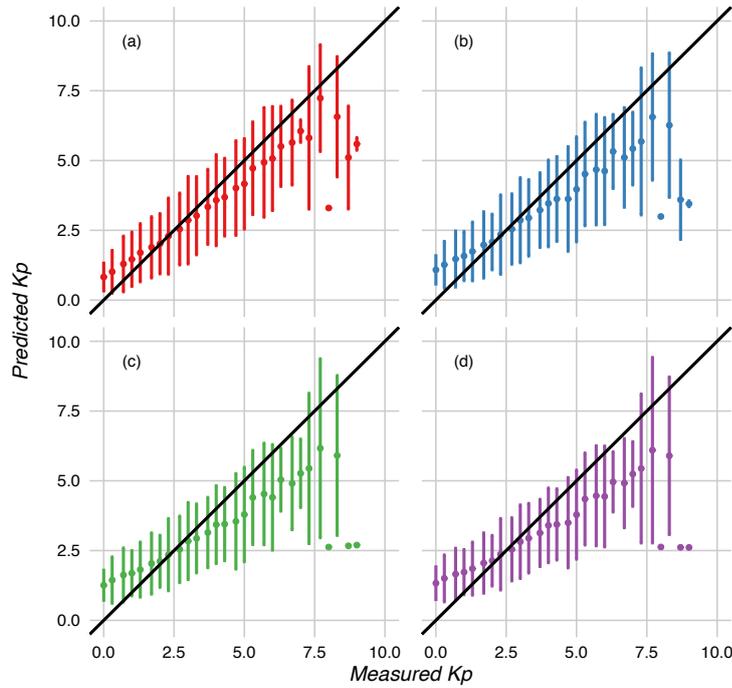


Figure 2. Comparison between the measured Kp index and predictions made for (a) 3, (b) 6, (c) 12, and (d) 24 hours ahead using the sliding window approach. The black line represents the ideal case when the prediction is equal to the measured Kp index. The points and bars correspond to the means and one standard deviations of the predictions made for each of the 28 Kp values.

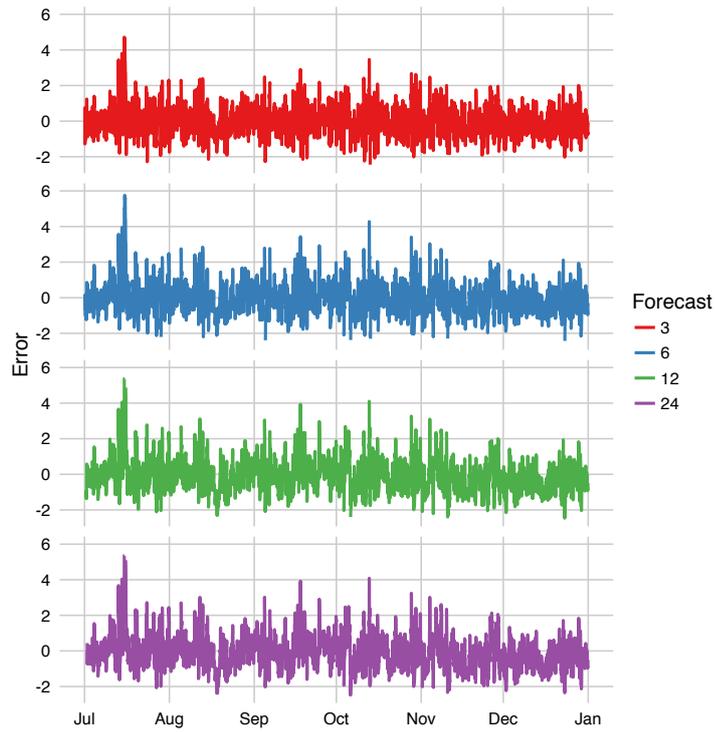


Figure 3. Error time series for the four horizons of interest obtained with the sliding window approach.

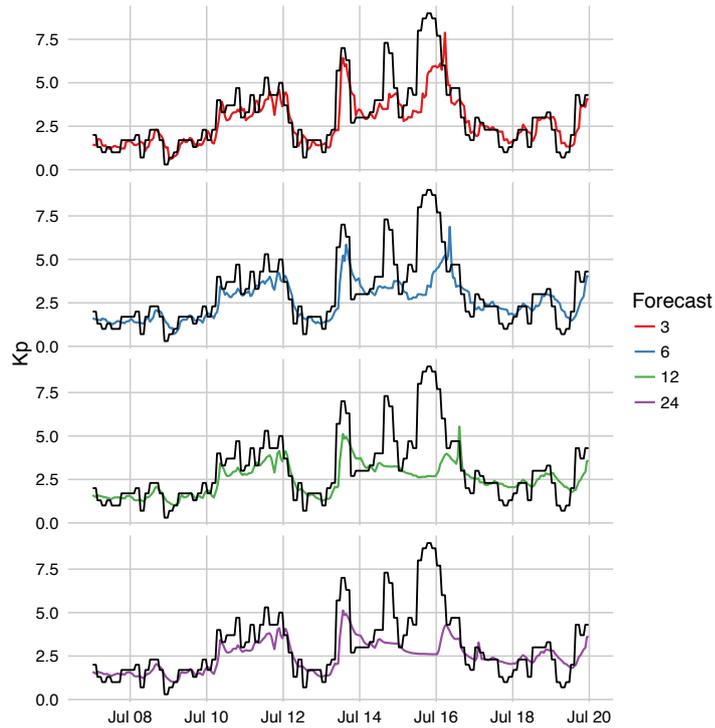


Figure 4. Predictions of the Kp index for the four horizons of interest during the middle of July 2000 using the sliding window approach. The black line corresponds to the measured Kp values.

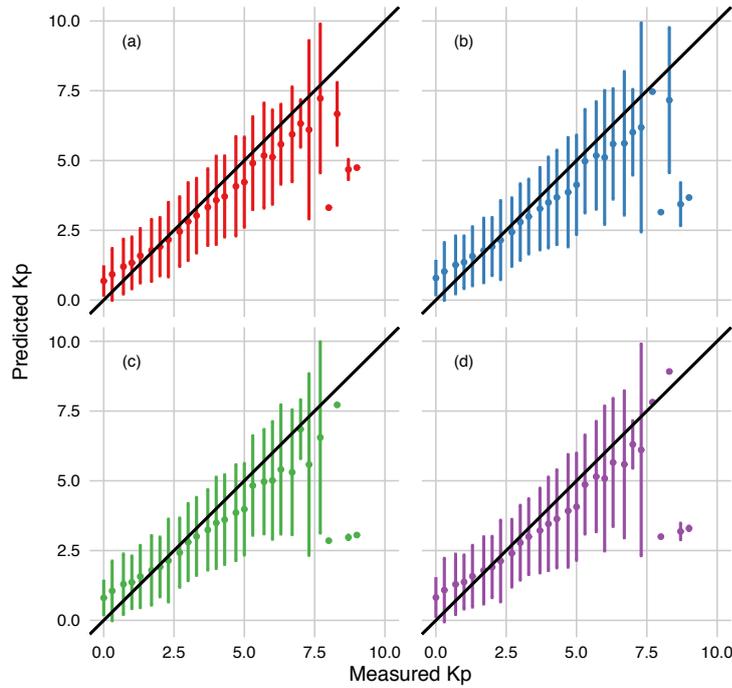


Figure 5. Comparison between the measured Kp index and predictions made for (a) 3, (b) 6, (c) 12, and (d) 24 hours ahead using the direct approach. The black line represents the ideal case when the prediction is equal to the measured Kp index. The points and bars correspond to the means and one standard deviations of the predictions made for each of the 28 Kp values.

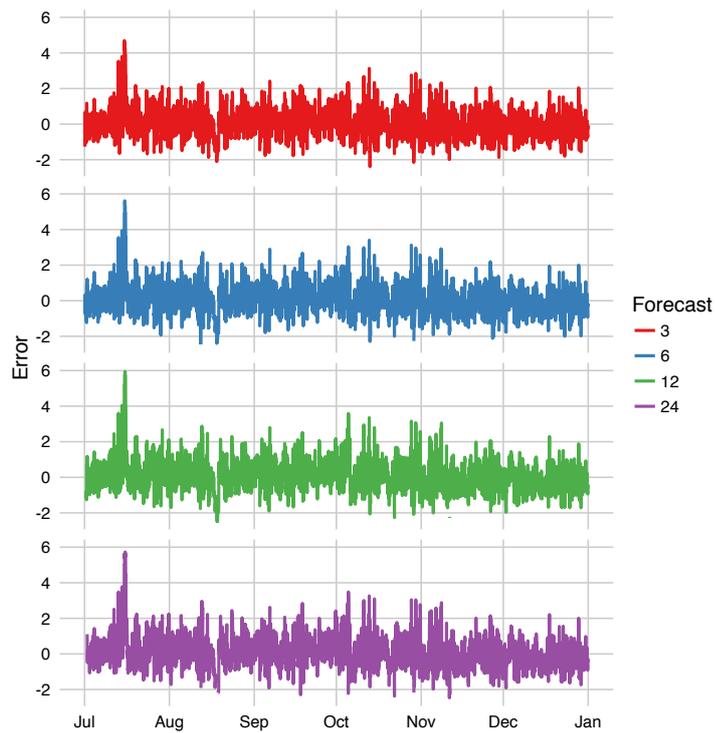


Figure 6. Error time series for the four horizons of interest obtained with the direct approach.

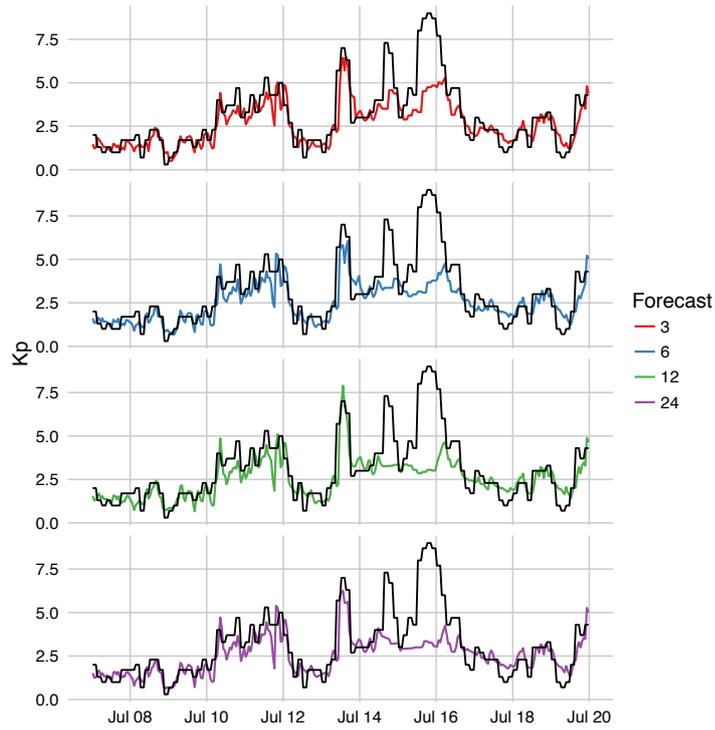


Figure 7. Predictions of the Kp index for the four horizons of interest during the middle of July 2000 using the direct approach. The black line corresponds to the measured Kp values.

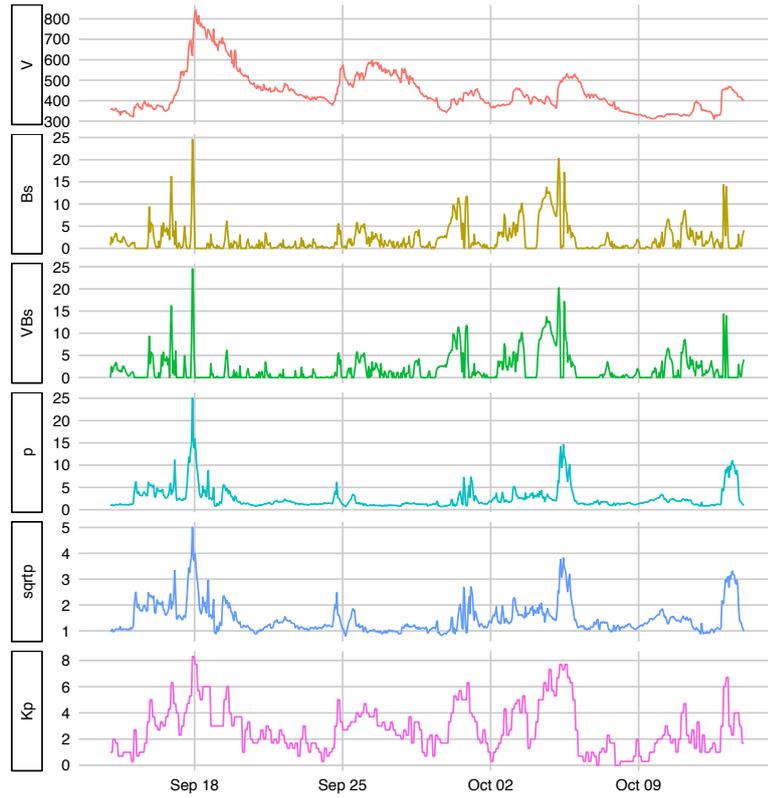


Figure 8. Feature dynamics for a randomly selected 30-day period on the second half of year 2000. The variable sqrtp corresponds to $\sqrt{p(t)}$.

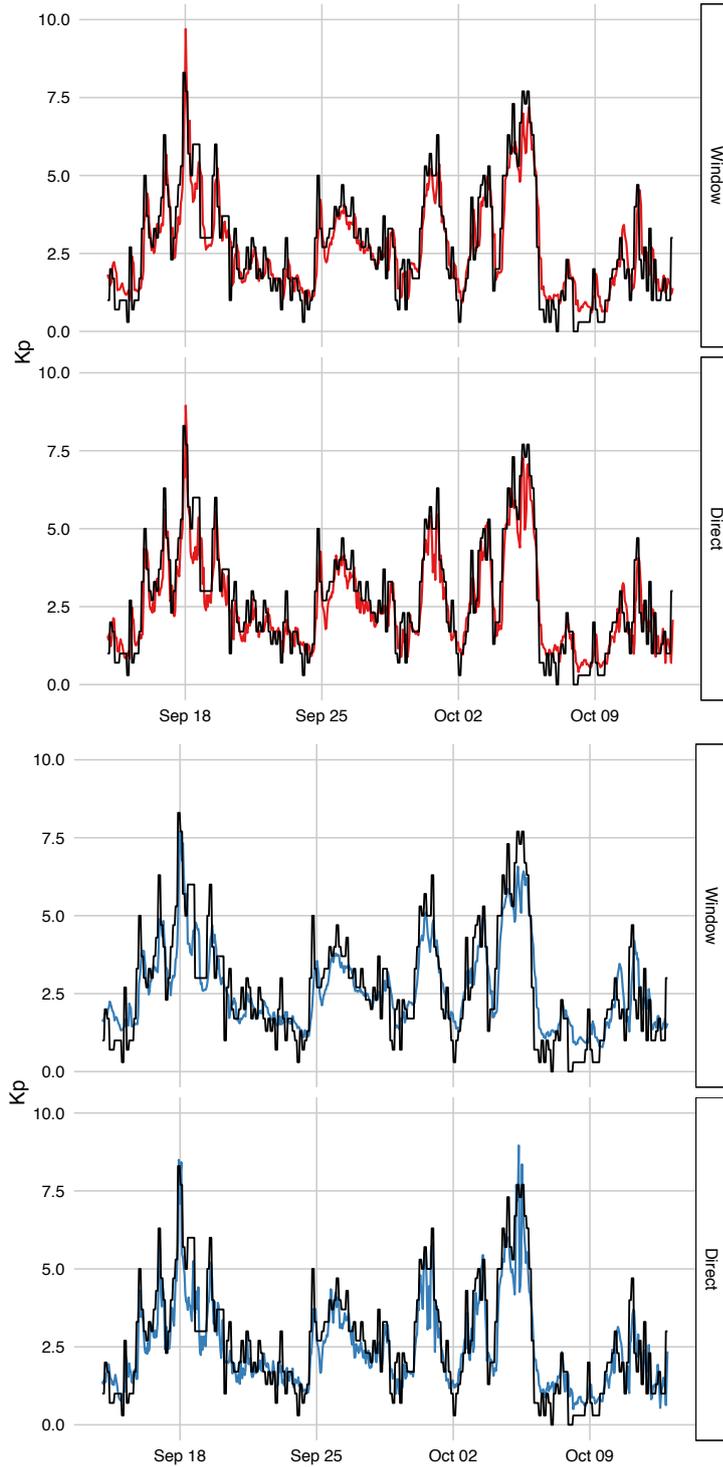


Figure 9. Comparison between the sliding window and direct approaches for 3-hour ahead predictions of the Kp index during a 30-day interval between September and October of year 2000. The black line corresponds to the measured Kp values.

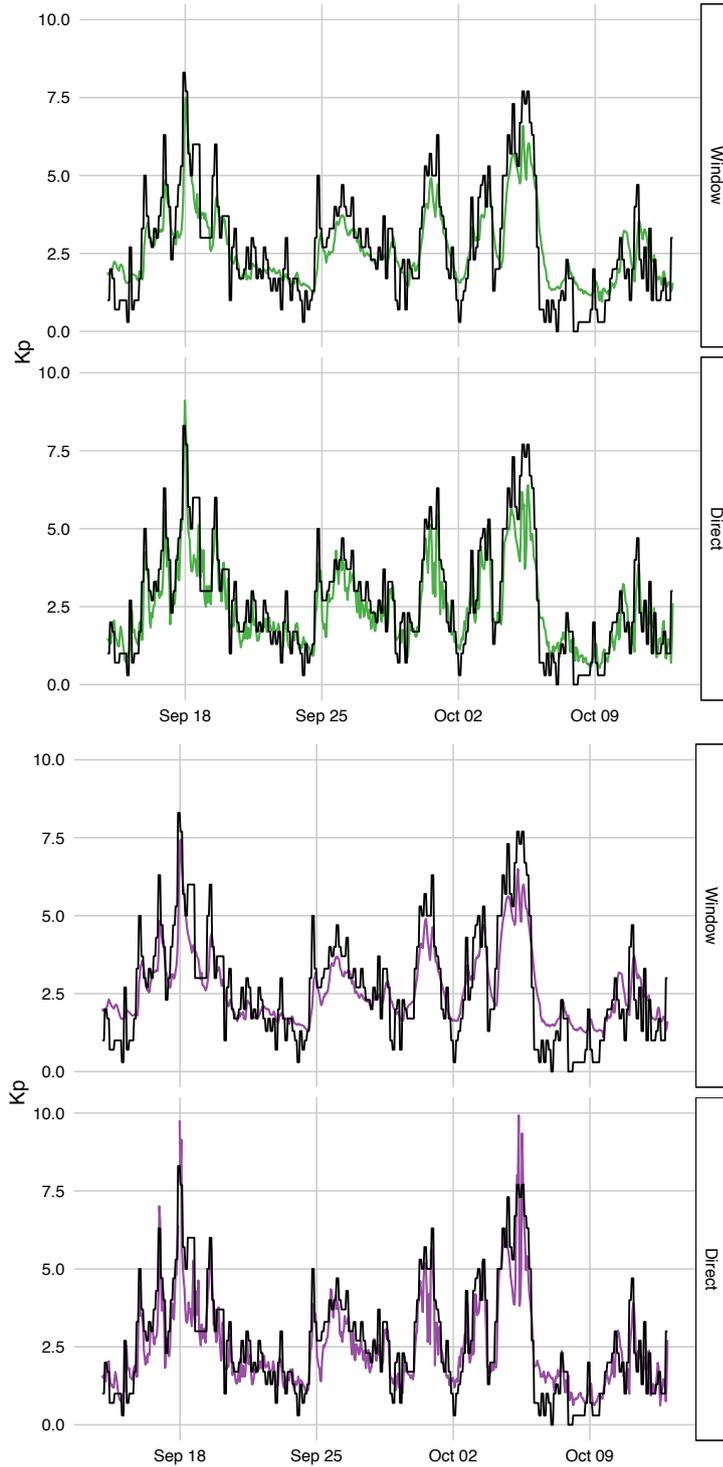


Figure 10. Comparison between the sliding window and direct approaches for 24-hour ahead predictions of the Kp index during a 30-day interval between September and October of year 2000. The black line corresponds to the measured Kp values.

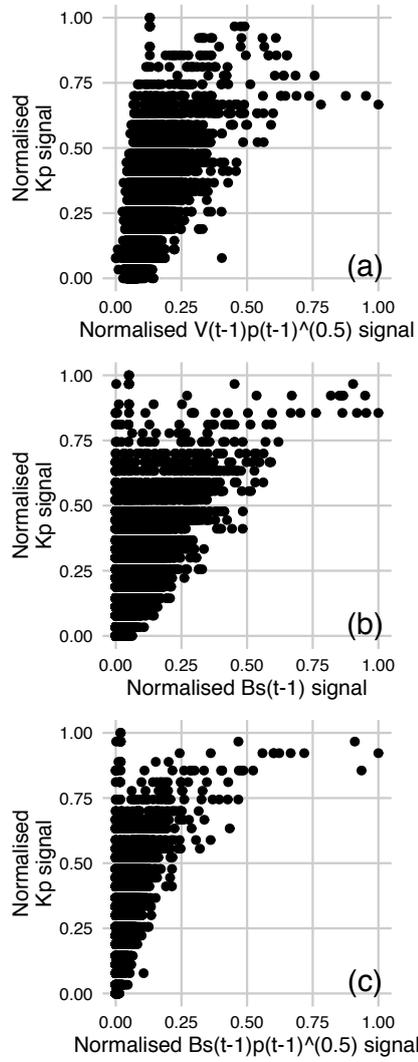


Figure 11. Top three significant model terms shared by all models in the direct approach. The correlation coefficients are (a) 0.6149, (b) 0.5571 and (c) 0.5437.