Real-time Prediction of Magnetospheric Activity Using the Boyle Index

Ramkumar Bala
Department of Physics and Astronomy, Rice University, 6100 Main Street, TX 77005, USA

P. H. Reiff
Department of Physics and Astronomy, Rice University, 6100 Main Street, TX 77005, USA

J. E. Landivar
Department of Physics, The University of Texas at Arlington, Arlington, TX 76019, USA

Abstract.
We present a new algorithm with an improvement in the accuracy and lead-time in short term space weather predictions by coupling the Boyle Index, \( \Phi = 10^{-4}\nu^2 + 11.7B \sin^3(\theta/2) \) kV, to artificial neural networks. The algorithm takes inputs from ACE and a handful of ground-based magnetometers to predict the next upcoming Kp in real time. The model yields a correlation coefficient of over 86% when predicting Kp with a lead-time of 1 hour and over 85% for a 2-hour ahead prediction, significantly larger than the Kp persistence of 0.80. The Boyle Index, available in near-real-time from http://space.rice.edu/ISTP/wind.html, has been in use for over 5 years now to predict geomagnetic activity. The logarithm of both 3-hour and 1-hour averages of the Boyle Index correlates well with the following Kp: Kp = 8.93 \log_{10} < BI > - 12.55. Using the Boyle Index alone, the algorithm yields a correlation coefficient of 85% when predicting Kp with a lead-time of 1 hour and over 84% for a 3-hour ahead prediction, nearly as good as when using Kp in the history but without any possibility of “persistence contamination”. Although the Boyle Index generally overestimates the polar cap potential for severe events, it does predict that severe activity will occur. Also, 1-hour <Boyle Index> value less than 100 kV is a good indicator that the magnetosphere will be quiet. However, some storm events with Kp>6 occur when the Boyle Index is relatively low; the new algorithm is successful in predicting those events by capturing the influence of preconditioning.

1. Introduction

The Boyle Index (BI) is an empirically derived formula that approximates the steady state polar cap potential (PCP) and has an empirically derived functional form given by \( \Phi = 10^{-4}\nu^2 + 11.7B \sin^3(\theta/2) \) kV, where \( \nu \) is the solar wind velocity in km/sec, B is the magnitude of the interplanetary magnetic field (IMF) in nanoteslas, and \( \theta = \arccos(B_Z/B_{GSM}) \) [Boyle et al., 1997]. Besides its simple scalar functional form, the IMF and solar wind dependent viscous term makes it a good indicator as to the state of the magnetosphere system [see similar functions in Newell et al., 2007]; the PCP also corresponds to the ionospheric plasma flow that is a low-altitude magnetospheric phenomenon. In general, during the periods when solar wind and IMF are steady for several hours, an hourly-averaged BI is a good predictor of the polar cap potential drop, provided the BI is less than 160kV. However, it overestimates the PCP when the BI exceeds 160kV. Theoretical studies have suggested that polar cap potential drop reaches saturation during periods of strong and southward IMF [Hill et al., 1976]. Recently, polar cap saturation has been observed and reported by various authors using different techniques [e.g., Hairston et al., 2003; MacDougall et al., 2006]. Furthermore, Hairston et al. [2003], using the DMSP spacecraft observations of the October and November 2003 superstorms, showed that saturation of the polar cap potential generally follows the Hill-Siscoe model [Siscoe et al., 2002a], with saturation potential in the range 160-250 kV.

Kp is a geomagnetic activity index based on a quasi-logarithmic scale and is calculated using weighted average of 13 ground-based magnetometers situated between 48° and 60° magnetic latitude. The values of Kp vary from 0 (signifying a quiet period) to 9 (severe activity) in 28 quantized levels and therefore take values from 0, 0+, 1-, 1, 1+, 2-, ... 9. In the past, several studies have shown a good correlation between Kp and IMF and with other parameters of the solar wind and it is now fairly well known that the magnetosphere responds to variation in the solar wind parameters [e.g., Papitashvili et al., 2000]. Recently, Johnson and Wing [2005] have discovered a significant nonlinearity in the Kp time series and have attributed this to solar cycle dependence of internal magnetospheric dynamics.

While Kp has a crucial importance in space weather for being a convenient measure of the geomagnetic activity and as a parameter used in the magnetospheric and ionospheric models, it has its own deficiencies. The official values of Kp are not made available until the end of every month and therefore, due to processing delays, are less suited for some real time applications such as space weather forecasts; however, some near-real-time proxies are now available. Since it is a 3-hour average, it is not ideal for timing events. Recently, several groups have tried to address these issues involving Kp and bridge the gap so it could be made to serve in real time. Several algorithms have been developed to nowcast Kp [e.g., Takahashi et al., 2001] and make predictive estimates of Kp ahead of time [e.g., Wing et al., 2005; Costello 1997; Boberg et al., 2000]. In fact, one such algorithm based on an Artificial Neural Network (ANN), developed by Costello [1997] and currently being used by the US
Air Force, take solar wind data as inputs and generates Kp every 20 minutes with an early warning of up to 30 minutes in advance. Given the accuracy of these Kp estimates and the generally short forecast times, there is an opportunity for significant improvement.

With the advent of the Advanced Composition Explorer (ACE) spacecraft and its ability to provide IMF conditions of upstream Earth and solar wind data in real time, it is possible and relatively straightforward to estimate Kp up to 3 hours in advance. ACE provides a broad spectrum of conditions responsible for geomagnetic activity as the storm drivers during the ACE era have been the strongest or at least as strong as those observed by spacecraft during the pre-ACE era [Vogt et al., 2006]. A long time line of observations from ACE, with up to a minute resolution, and the history of Kp dating back to 1932, provides an excellent opportunity to develop a prediction algorithm for the next upcoming 3 hour Kp. In the past, various statistical correlation techniques have been used to infer the magnetosphere’s response time to the changing IMF and solar wind conditions. In this study, the statistical correlations between the natural logarithm of BI and Kp are explored. Since the BI was derived using steady-state conditions, we exploit the neural network to explore the effects of time variability, including preconditioning, which may be non-linear.

The goal of this paper is to design a computer algorithm with good prediction capabilities for moderate to severe storms and to forecast Kp up to approximately 2-3 hours ahead in near-real time. Artificial neural networks (ANNs) have been a useful tool for temporal predictions in problems involving non-linearity. Numerous applications have been developed recently using such networks in the context of space weather and geomagnetic storms [e.g., Costello 1997; Wing et al., 2005; Wu et al., 1997]. In this study, we primarily focus on training and validating an ANN using the time histories of BI and Kp to develop a new Kp forecast model. Here we have chosen the standard multilayered perceptron-backpropagation network for the purposes of training, validation, and testing.

The BI plot is available in real time from http://space.rice.edu/IISTP/wind.html. Subscribers to the “spacairt” mailing list receive email notices whenever the 10-minute BI average exceeds 200 kV (red alerts). In over five years of real-time operations since October 2003, no major storm (Kp>6) has been missed. However, there were some false alarms, mainly owing to a 10-minute averaging time rather than a one or three hour average. In this work, we train a neural network and extend our operation to forecast Kp in four different modes: (1) a model that takes solar wind and magnetospheric data from ACE and Kp from the National Weather Service and predicts Kp 1 hour ahead; (2) a model that takes the same input as above but predicts Kp 2 hours ahead; (3) a model that takes only solar wind and magnetospheric data from ACE to derive the BI and hence predict Kp 1 hour ahead; and (4) a model that predicts Kp 3 hours ahead using the BI. In order to achieve that, we use both an hourly and the conventional 3-hour averaged Kp and BI. Of course, the 1-hour Kp is merely the oversampled 3-hour Kp, ascribed to each hour. In the training (and of course in the real-time predictions), we never use a known Kp index value to predict itself; if the previous hours Kp index is not known, the prior measured Kp index is just duplicated. The operational modes of three of our models described here are similar to the APL models presented in Wing et al. [2005]. We have done a head-to-head comparison of our forecast model with the Costello geomagnetic activity index model. Finally, this paper will demonstrate an improved real-time prediction capability for time resolutions of up to 2 hours in advance.

2. Data and Technique

The BI’s used in this study were computed using archived data from ACE (1998-2007) while the official values of 3 hour averaged Kp were obtained from GeoForschungsZentrum (GFZ), Potsdam, Germany. We have also used solar wind and IMF data from WIND (1995-1997) and IMP-8 (1995-1997) respectively to derive the BI, making the total available time line of observations span 13 years and covering an entire solar cycle. Although the official Kp record is fairly uninterrupted, gaps do exist in the IMF data from ACE and WIND. For real-time forecasts, the National Weather Service (http://sec.noaa.gov/) provides regular updates of estimated Kp (3 hour average) while ACE level 2 data products of SWEPAM (solar wind plasma) and MAG (magnetic field) provides an uninterrupted data set to derive the BI.

3. Statistical Correlations

A steady BI for a few hours is a good predictor of the polar cap potential drop, which in turn is a predictor of magnetospheric activity [Boyle et al., 1997]. During quiet times, the BI can drop below 20 kV, and can reach over 400 kV before or during severe storms. A 3-hour average of the logarithm of BI compared to the following 3-hour Kp has a good correlation (r = 0.74) as shown in figure 1. Our analysis is facilitated by choosing a logarithmic transformation to scale the BI. Although the BI is an overestimation of the PCP’s if it is above 160kV, nevertheless and surprisingly, the correlation is still valid, possibly because Kp is logarithmic as well but also because perhaps, even though the polar cap velocity may saturate, the overall magnetospheric response may not. In figure 1, the vertical and horizontal lines within the plot represents one possible BI cutoff (discriminator level, here 110 kV) and corresponding Kp index (5) cutoff. The cut-off shown here has been deliberately chosen to illustrate the result that the likelihood of a storm having a Kp index of 5 or higher exceeds 95% when the average BI over the previous three hours is over 110 kV.

However, a fair number of “misses” occur with that discriminator level. By reducing the discriminator level to 100 kV, for example, the number of “misses” decreases, but the “false positives” increase. One can set these cut-offs arbitrarily by trading a few hits for misses, right rejections for false alarms, and vice versa depending on the kind of forecast desired. During 2003 and 2004, when the BI fell below 110 kV, the magnetosphere was quiet (Kp index < 4) as shown in figure 2. During active periods of solar maxima, as shown in figure 3, hourly averages of the BI and Kp are correlated, where the Kp has been oversampled to one hour resolution. During some very active periods, an hourly averaged BI can exceed 250 kV, in which case the geomagnetic Kp index could be over 7, causing major geomagnetic storms and low-latitude auroras to form within the next few hours.

3.1. Skill Scores

Predictions that require “Yes/No” answers, for example, “Will Kp exceed 6?” are needed for certain applications, for example, protection of hardware resources or mobilizing an observer network. For some of these applications (e.g. alerting observer networks), one wants to minimize “false alarms”; for other applications, such as protecting delicate equipment, one wants to minimize “misses” while allowing a few false alarms. One can determine a discriminator level (in BI) for each trigger level (in Kp), that either minimizes the misses or minimizes the false alarms.

A skill score can be computed that takes all the above-mentioned factors into account. Conventionally, in a categorical forecast the forecast events are conveniently displayed using an i x j contingency table with i x j representing the possible combinations of all possible events. Figure 4
displays a definitive way to quantify the performance scores of the training set using total skill statistics (TSS), a most commonly used skill score to summarize a 2 x 2 contingency table. More specifically, the Heidke Skill Score (HSS) for the 2 x 2 situation used in this study is written as [Wilks, 1995]:

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}$$  

(1)

where a is the number of “hits”, c is the number of “misses”, b the number of “false positives” and d is the number of “correct rejections” for a given sample, thereby constituting the 2 x 2 contingency table. A perfect forecast receives a TSS score of 1 while a random forecast receives a score of 0. Figure 4 shows a steady increase in the TSS and eventually saturating with the BI as Kp threshold is increased. Forecast attributes based on a 2 x 2 contingency table [see Wilks, 1995], for the plots shown in figures 1 and 2 have been summarized in table 1. From an operational standpoint, these results strongly suggest the feasibility of using the BI to make short-term predictions (up to 3-hr ahead) of the magnetospheric activity.

### 3.2. Cross-correlation Analysis

We closely examine the BI and Kp using cross-correlation techniques in order to better understand the time scales of the solar wind and IMF and their influence on the magnetosphere, and hence the resulting Kp. The cross-correlation function offers the best choice in dynamical weather forecasts to explore the temporal correlations involved in a linear time series. In figure 5, we plot the estimated cross-correlation functions of the logarithm of BI with Kp. For a 3-hour average (solid line), the strongest correlation (0.79) occurs at a positive lag of 3 hour i.e., the solar wind data for the 3-hour average most strongly influences the following 3-hour Kp index.

Next we binned the data in 1-hour average bins, and analyzed the 1-hour data set separately since our interest lies in training the network using both 1-hour and 3-hour averages. An hourly averaged (solid line) and Kp have a good correlation while the strongest correlation (0.76) still occurs at a positive time lag of 3 hour. Although there was no significant difference in the two plots, it is worth noting that the lagged correlations are extremely small for large lags and decay rapidly after a few hours. We can therefore infer that with Kp trailing BI, the prediction lead-times are in essence decided by the positive time lags; therefore, forecasts can be made accurately within a time range of 1-4 hours. Therefore, by training the preconditioning events using the BI on time scales imposed by cross-correlation analysis, a reliable Kp proxy can be estimated in advance.

### 3.3. Kp Linear Predictor

Kp can be approximated using the BI with a linear fit via:

$$Kp_{Lin} = 8.93 \log_{10}(BI) - 12.55$$  

(2)

Figure 6 shows the linear predicted Kp versus the official Kp index for 2006 using a 3-hour average BI (linear correlation r = 0.766). Similarly, the linear predicted Kp versus the official Kp index for 2006 using a 1-hour averaged BI has a linear correlation of r = 0.69. Clearly, it can be seen from the above equation that this simple linear-fit model predictions are only valid (i.e., Kp≥0) for BI>25.4 kV, a deficiency that has been overcome by the more efficient ANN models discussed below. Note that this year was not part of the training data used to create the linear fit.

### 3.4. Boyle Index: Effect of Preconditioning Events

Coronal mass ejections (CME) from the sun, large solar flares, cloud-driven storms and high-speed solar wind streams are often responsible for causing dramatic disturbances in the magnetosphere resulting in powerful geomagnetic storms. Recently, several studies have examined the geomagnetic storm drivers in the context of space weather forecasts [e.g., Lavraud et al., 2006; Borovsky et al., 2006; Wu et al., 2002]. A preconditioning event tunes the magnetosphere to a specific state as a function of the preceding solar wind and IMF conditions before the onset of a storm. For example, a prior substorm could increase the ionospheric conductivity and/or provide a seed ring current population that will be injected farther in by a following substorm. Therefore, success of a forecast algorithm largely depends on training the network with the preconditioning events and the magnetospheric response to such events. The following example illustrates the effect of preconditioning.

As an example, figure 7 shows a time series plot of the solar wind, magnetospheric and Kp index values of a storm (BI>200 kV; Kp>6; Dst <-110 nT) that occurred on 14 April 2006. This event has been chosen for its steady high Boyle index lasting over 5 hours. Despite the steadiness of the Boyle Index, the Kp index showed a steady rise, remaining at 6 or higher for a duration of 9 hours, illustrating the non-linearity of the response. The dotted line represents the Kp proxies predicted by our model 1 described below, correctly predicting the rise in Kp despite the flat or slightly falling BI.

### 4. Feedforward Backpropagation Neural Network

The basic architecture of a computer NN replicates the human brain in that both have a common fundamental building block called a neuron. The network’s ability to analyze an unseen problem largely depends on the training it received. Since typical time scales for magnetospheric response to a sudden change in the solar wind or IMF vary from a few minutes to a few hours, it is vital to train the network to memorize the integrated events of the recent past in order to get a reliable estimate out of it. The cross-correlation results lay out a quantitative guideline for constructing a dynamic neural network using time delay inputs.

In this paper, we employ the standard feedforward back propagation multilayer perceptron network as has frequently been used in the literature and proven to be well suited for non-linear time series prediction [Haykin, 1999]. A schematic representation of the network architecture used in this study is shown in figure 8. The network was designed with an input layer for which the number of inputs vary with model, a hidden layer consisting of approximately 10 neurons, a parameter also varying with model, and an output layer with a single neuron. The firing of the neurons in the hidden layer was achieved through a non-linear hyperbolic tangent transfer function while the output neuron was activated using a linear function and finally, the network training and validation was achieved using the Levenberg-Marquardt algorithm. One of the handicaps of using this algorithm is that it requires large memory and storage space and therefore is not feasible for large problems. Nevertheless, it offers faster convergence among the gradient descent algorithms [Haykin, 1999].

Training and validation has been done simultaneously by combining them in to a single operation using a cross validation technique. Here, the available data is divided into two distinct subsets of training and validation set. The validation error is monitored at the end of each iteration and
the training set is cross validated against a validation set. A cross validation technique is effective in that it prevents the network from over fitting the training set and helps generalization of any data outside of the training set. Also, it requires stopping criteria to be set as the errors ripple back and forth after every epoch. Training is terminated if the desired number of iterations has been reached or when the training and validation converges to a rms error, as specified by the user, prior to the specified number of iterations. In this study, the network training parameters were determined by trial and error. To ensure better training and generalization the number of hidden nodes were cautiously chosen as the network is very sensitive to the number of neurons in the hidden layers.

5. Magnetospheric Activity Index: Kp Prediction

Prediction accuracy varies both with the integrated time and with the delay time used in the inputs. We trained the network with different input granularities using 30-minute, 1-hour and 3-hour averages and found that the best linear correlation coefficient between the true and predicted Kp was observed while using a 1-hour average and looking back 6 hours into the past. However, when the look back time was increased from 5 hours to 10 hours or to a much higher value by keeping the integration time constant, the prediction efficiency barely changed, a finding consistent with Johnson and Wing [2005] and Wing et al. [2005]. Therefore, introducing a delay line guarantees a drastic change in the network dynamics but does little to the prediction efficiency when the delay time gets sufficiently large. This implies that the algorithm is approaching an optimal solution. Optimization of our function was effectively achieved by adjusting the learning rate, number of iterations, mean square error or by a combination of all the three parameters. The gradient descent algorithm intrinsic to the levenberg-marquardt routine offers the best choice for minimization. Given that official Kps are 3-hour averages, data granularity of an hour or less is obtained by simply oversampling the data points. We have explicitly developed three Kp prediction algorithms to run at different time granularities and using different inputs. The following subsections describe the results and performance of each algorithm individually.

5.1. Network tests and evaluation

In conjunction with the TSS described in section 3.1, the forecast accuracy of the models is also characterized by RMS error as defined by

\[
RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (X_t - O_t)^2}
\]

and, prediction efficiency (PE) defined as

\[
PE = \sqrt{r}
\]

where \( r \), \( X_t \), and \( O_t \) represent the linear correlation coefficient, predictions, and the actual values respectively.

5.2. Kp Prediction with 1-hour Lead-Time using the BI and Kp history - Model 1

One of the models we developed takes the time history of Kp and the derived BI as its inputs, hereafter model 1, to predict Kp approximately one hour ahead. For operational purposes, the time history of Kp is obtained from the estimated 3-hour planetary Kp index derived at the U.S. Air Force Space Forecast Center using several ground-based magnetometers serving in near real-time. There is, however, a 30-40 minute lag before they are publicly available, mainly owing to processing delays but still usable within an hour.

The non-linear functional relationship of the time series connecting the input elements BI and Kp to their targets Kp follows the function below

\[
K_{p_{t+1}} = f(K_{p_{t-3}}, K_{p_{t-4}}, K_{p_{t-5}}, \ldots, K_{p_{t-11}}; BI_t, BI_{t-1}, BI_{t-2}, \ldots, BI_{t-8})
\]

where the time \( t \), \( t-1 \), etc. in each case represents the end of the integration time period. Thus the time \( t \) is the most recent BI measurement, and the 1-hour prediction then covers the time frame of \( t \) to \( t+1 \). Note that the solar wind takes roughly 40 minutes to arrive at Earth. With Kp clearly time shifted behind the BI, puts BI as the precursor of rises or fall in Kp to come and gives an upper hand leading up to the predicted time. The most recent Kp value used is \( t-3 \), assuring that the network is not trained to predict a known value of Kp.

The network is designed such that it looks 9 hours into the past (equation 5). It was trained, validated and tested using hourly averages of BI and Kp covering a 13 year period from 1995-2007 with data sets classified to avoid overlap. In all cases the month of April 2001, and all the 2006 and 2007 data, were used for testing only and were not part of the training set. In dealing with the null values of Kp or the BI, we have either completely rejected or interpolated the missing data in order to minimize their statistical impact. The data are rejected when there is a long streak of null data lasting several hours but, if null values occur at some isolated instances, the data is interpolated using the two adjacent points. A network’s prediction efficiency is characterized by its performance on a test data that is completely new to the network along with a long time period over which it is tested.

In figure 9, we plot the ANN Kp predictions as a function of the official Kp (linear correlation coefficient of \( r = 0.863 \)). The network has in fact, learned to reproduce over 74% of the variance of the data presented to it through a mock test intended to simulate a real-time operation over the whole test set (April 2001, all of the 2006 & 2007). This test strictly follows our training in that we do not consider the most recent Kp. Also, as a reminder, this test set was not included in our training. However, caution must be used while running model 1 in real time. The realtime model uses the NOAA nowcasted Kp, which is only estimated to the nearest unit, and since the network training and model evaluations were based on the official Kp record, the network performance is likely to diminish slightly.

Forecast attributes, at selected Kp thresholds, of the test results have been tabulated in table 2. Figure 10 offers a different perspective of the model performance by displaying the predicted and official values over-plotted for randomly chosen 30 day periods from 2006 (Dec) and 2007 (Aug). It is seen that a handful of storms with Kp > 5 have been well predicted both during the sudden commencements of the storm and the recovery phases. A shorter time resolution also helps to warn users of imminent storms owing to rapidly changing conditions in the magnetosphere without having to wait until the next three hour conventional Kp.

5.3. Kp Prediction with 2-hour Lead-Time using the BI and Kp history - Model 2

A more practical approach to any advanced warning systems is to look for a longer lead-time without straying too far from accuracy. Our second model has similar input as
model 1, but predicts Kp approximately 2 hours ahead, here-
after called model 2. Recall from section 3 that prediction
from a 3-hour average of the BI, although with a lower pre-
diction accuracy than a one-hour average, is nevertheless
reasonable. In addition, we saw from the cross-correlation
analyses that the statistical significance of the correlation
coefficient lasts a few hours before decaying rapidly. The
volatile nature of magnetospheric dynamics makes it diffi-
cult to capture and forecast any impending changes shorter
than 3-hour duration. Another potential downside of this
model is its time resolution. Moreover, the feedforward net-
work has a better prediction efficiency while using a 1-hour
lead-time as opposed to a 2-hour lead-time. Nevertheless,
it is still a viable and reasonable option to train the network
to predict Kp approximately 2 hours ahead, hereafter model
2.

The network inputs and training are similar to the de-
scription in section 5.2 and involves the following equation:

\[ K_{t+2} = f(K_{t-1}, K_{t-2}, K_{t-3}, \ldots, K_{t-9}; BI_t, BI_{t-1}, BI_{t-2}, \ldots, BI_{t-8}) \]  

where \( t \) denotes the current epoch. Given the availability of
3-hour nowcast Kp and live updates of the BI, we have de-
signed our network such that it can deliver a prediction for
the next upcoming 3-hour Kp. Assuming a delay of approxi-
ately 40 minutes due to processing time for the preceding
Kp, the lead-up time is, therefore, only slightly more than
2 hours in real-time, i.e., at a time 04:00 UT we predict the
Kp which will cover the time period 03:30-05:30 UT, using the BI
up to 03:55 and the previous Kps up through 03:00 UT. If for
any reason the 0-3 UT Kp is not available, we duplicate the
previous Kp; however, this has not been necessary in the
two years we have been using the realtime Kp.

A direct comparison of the model outputs with the true
Kp shows a drop in the prediction efficiency when the lead-
time is extended. The linear correlation coefficient between
the real and ANN Kp is found to be 0.855 with a prediction
efficiency of 0.73 (figure 11), outperforming the prediction
efficiency of a self predicting Kp. As expected, the network
estimates did not improve with longer time delays. The net-
work also does well both during the initial phases of storm
commencement and during the recovery phases. The dis-
tribution of predicted and official Kps for models 1 and 3
are depicted in figure 12. Clearly, there is a drop off in the
predicted Kp for Kp < 1 which is attributed to the net-
work’s inability to learn the minimum and maximum states
[Wasserman, 1989]. As a supplement to table 2, the fore-
cast skill scores based on models 1 and 2 is shown in figure
13 where the HSS is plotted as function of Kp threshold
although seemingly, it under reflects the performance levels
shown in figures 9 and 11 i.e., forecasts obtained from 1-hour
and 2-hour lead-time models. A similar caution to the one
noted in model 1, must be used while running model 2 in
real time.

5.4. Kp Prediction with 1-hour Lead-Time using the
BI alone - Model 3

Our recent success, with only a few false alarms, in pro-
viding space weather alerts using BI derived from the solar
wind and IMF measurements at the L1 point (also available
in real time from http://space.rice.edu/ISTP/wind.html)
provides the means to develop this model. Although the
use of time history of Kp in addition to solar wind inputs
provides good results, from a forecast standpoint, if a model
using solar and data alone is essentially as effective as the mod-
els which also require nowcast Kp, then simpler functions
should be used which avoid the concerns about the availabil-
ity or quantization of the nowcast Kp. The network training
and validation was done in a manner similar to model 1 and
the outputs have a lead-time of 1 hour.

Recently, Newell et al. [2008] have shown that the use of a
viscous term in addition to a merging term dramatically im-
proves the predictability of geomagnetic indices such as Kp
up to 75% (r = 0.866) without prior knowledge of the target
index. Our algorithm uses \( 10^{-4} \nu^2 \) and 11.7Bsin(\( \theta/2 \)) as the
viscous and IMF-dependent terms, with no significant
contribution from the solar wind pressure, to predict Kp ap-
proximately one hour ahead. The prediction efficiency of the
network using the BI as a stand alone to estimate the next
upcoming Kp is 73% with a linear correlation coefficient of
0.852 between the real and predicted Kp (figure 14). Since
its prediction efficiency is virtually identical to that of Model
1 (at least for this test set), it offers a practical alternative
to provide Kp proxies in a timely manner, without concerns
about nowcast Kp availability.

5.5. Kp Prediction with 3-hour Lead-Time using the
BI alone - Model 4

Our fourth model inputs solar wind parameters to de-
vote the BI and predicts Kp 3 hours ahead, hereafter called
model 4. Therefore, it not only extends the forecast range
but also offers a full 3-hour Kp prediction capability every
one hour since the BI is generally available at near-real time.
Figure 15 displays the scatter plot of the official Kp versus
the ANN Kp, based on April 2001, 2006 and 2007 test re-
sults, with a prediction efficiency of 71% and \( r = 0.845 \). We
would like to remind the reader that these results are almost
as good as those obtained from model 2 which uses the BI
and known Kps (PE = 73%, \( r = 0.855 \)). The network was trained
using hourly averages of the BI to forecast Kp a full
3 hours ahead.

6. Post-test cross correlation analysis

Using the time history of Kp in the algorithm to predict
future values of Kp can suffer from "persistence". That is,
one reasonable prediction is: "the next Kp will be the same
as the previous one". If that if your algorithm, the pre-
diction you come up with will have a tendency to lag the
actual Kp by 3 three hours, and will have a correlation coef-
ficient which is just the autocorrelation of Kp at +3 hour lag.
Thus, in order to get a true "prediction", one must have a
correlation coefficient which is significantly above the "per-
sistence" value, and have a prediction time series that peaks
at lag = 0 (i.e. the prediction time = the measurement time)
and not lag = -3 (the prediction just repeats the previous
data point).

Figure 16 shows the auto-correlation of the 3-hour official
Kp, shown as the solid line. It of course has a 100% efficiency
in predicting itself at zero lag. More importantly, though, it
shows a a high (0.805) auto-correlation at 3 hours, showing
the persistence of Kp from one 3-hour measurement to the
next, setting the standard for effectiveness of prediction al-
gorithms. The results of the cross correlation of our model
2 and model 4 predictions with the official Kp are shown
respectively as a dashed and thick lines. For those curves, a
lag=0 means that we are correctly predicting the following
Kp value at the proper time. Both of our model predictions
at lag=0 are significantly above 0.805, showing effectiveness
in beating the persistence hypothesis. A Z-test [Kanji, 1999]
performed on model 4 to investigate the significance of its
correlation coefficient (\( r = 0.845 \)) from the auto-correlated
Kp at \( r = 0.805 \) shows a difference which is significant at the
1% level. However, for our model which uses Kp his-
tory (model 2), the best correlation is not for the predicted
time but actually is for \( t=-3 \), that is, the prediction lags the
real data. Our neural network model which includes Kp
history (model 2) appears to overly on the previous Kp
value. Thus a prediction using Kp history significantly lags
a true prediction. This effect can be seen in other papers which use Kp history (e.g., see Fig 5 (g) and (h) of Wing et al. [2005]). However, the APL model 3, which depends only on solar wind and not Kp, does not exhibit this lag. We can now further claim that in spite of a small (not significant) drop in linear correlation coefficient and the prediction efficiency, the model 4 really does forecast (not just duplicate) Kp, as effectively using BI (and its history) alone as do models which include Kp history. Note that a similar decisive trend was hard to notice in our other models.

7. Comparison with existing Kp models

Among some of the existing and fully functional Kp models, the APL Model 3 [Wing et al. 2005], the Boberg et al. NN [Boberg et al. 2000], and the Costello NN [Costello 1997] all use the solar wind to predict Kp approximately one hour ahead and therefore, easily compares with our own model 3 described in section 5.4. In this paper, we did not attempt to re-evaluate them model by model, although we did conduct a head-to-head test with the Costello Kp model over a limited range of data available to both the models. As far as the other models, we use their own evaluations reported in the literature for comparison. Of these, the APL model 3 is by far the best (r = 0.84) while the Boberg et al. NN Kp model reports r = 0.768 for a RMSE = 0.985 tested over a 11 year period. Clearly, our model 3 outperforms the Costello NN and the Boberg et al. NN models. But the slight edge our model 3 has over the APL model 3 could be attributed to the fact that the latter was evaluated over a sufficiently longer time period. On the other hand, our model 1 has a lower correlation coefficient (r = 0.863) than the APL model 1 (r = 0.92). Further, the APL models 1 and 3 and Costello NN Kp model predicts Kp approximately 1 hour ahead using a 15-minute granularity, which partly explains the differences in the prediction performance.

7.1. Head-to-head Performance: Model 1 vs Costello NN Kp

The Costello ANN Kp model [Costello, 1997] was developed primarily to feed the Rice Magnetospheric Specification and Forecast Model (MSFM) by providing Kp as one of its basic input parameters. Currently, NOAA uses the model (http://www.swpc.noaa.gov/ftp/costello/index.html) to provide Kp proxies once every 15 minutes with a lead-time of approximately one hour through solar wind speed, IMF Bz, and [B] inputs from ACE. We have performed a head-to-head test of our model 3 with Costello’s ANN Kp towards a true comparison. In order to conduct a fair test, we have chosen the data set such that it includes both a severe (April 2001) and a benign period (Jan-Feb 2007) of solar activity and, to further facilitate the test by providing a large dynamic range of Kp values necessary for comparison. Also, prior to running the head-to-head test, we reran our model to ensure that the training set does not include the test data. Figure 17 (a) shows the results of our model 3 predicted Kp versus the real Kp (r = 0.84) while figure 17 (b) shows the Costello ANN predicted Kp versus the real Kp (r = 0.81) for the same time frame. However, possibly because of the smaller size of our test data, our evaluation of the Costello NN Kp is not consistent with the correlation coefficient (r = 0.75) obtained by Wing et al. [2005] through their evaluation of the same model and, inconsistent with the trend observed by Detman and Joselyn [1999].

8. Discussion and Conclusions

In this study, we have used the ACE level 2 data of SWEPAM, IMP-8 and WIND extending from 1995 to 2006 to derive BI and thus investigated its statistical correlation with Kp. We have shown that both 3-hour and 1-hour averages of the natural logarithm of BI and Kp are strongly correlated. In addition, using cross-correlation analysis, we have also shown that Kp and BI are strongly correlated at lead-times of 1 and 3 hour, depending up on the integration time with the optimum time lead time around 3 hours for both the one-hour and three-hour predictions. One of the key characterizations of a correlation analysis is to obtain quantitative information about the time scales involved in the magnetosphere’s response to the changing solar wind and IMF conditions. We have further analyzed a few pre-conditioning mechanisms and the magnetosphere’s response to such mechanisms by using a time delay neural network. Incorporating delay time into the training process helps the network learn the signatures of extended activity. In accordance with the time scales suggested by the cross-correlation analyses, we have now successfully constructed and evaluated simple computer algorithms based on an artificial neural network to forecast the geomagnetic activity index in Kp in real-time.

We have developed a real-time Kp forecast model operational in four different modes: (1) a model that takes solar wind and magnetospheric data from ACE and Kp from the National Weather Service (http://sec.noaa.gov/) and predicts Kp 1 hour ahead; (2) a model that takes the same input as above but predicts Kp 2 hours ahead; (3) a model that takes only solar wind and magnetospheric data from ACE to derive the Boyle index and predicts Kp 1 hour ahead; and (4) a model that takes only the solar wind data and predicts Kp for a full 3 hours ahead. Our analyses indicate a run-down in the prediction accuracy with increasing lead-time; the best prediction efficiency was achieved for an 1-hour lead-time. Our retrospective tests and network performance (see Table 4) indicate that our algorithm can give a reliable Kp forecast with a lead-time of 1-hour with a prediction efficiency of over 74% when using the past information of the BI and Kp. For predictions with a lead-time of 3 hour, the correlation coefficient between the predicted and official Kp is found to be 0.845 with a prediction efficiency of 71%. We have also shown that we can predict up to 73% of the true Kp variance for a lead-time of an hour while using the solar wind alone. Finally, running our models 1 and 2 in real time may not yield the accuracies reported because the NOAA nowcasted Kp, at times, may or may not mimic the official Kp. However, the models which do not use the history of Kp, but only use the history of the Boyle Index, are nearly as good a predictor of future Kp, and do not suffer the lag problem of models which include the history of Kp.

Acknowledgments. This paper is based upon work supported by the Center for Integrated Space Weather Modeling, which is funded by the STC Program of the National Science Foundation under Agreement ATM-0120950, and by NASA under the Immersive Earth Program, NCCS-316.

We thank the ACE SWEPAM and MAG instrument teams and the ACE Science Center for providing the ACE data. We also thank the GeoForschungszentrum (GFZ), Potsdam, Germany for maintaining and providing the official Kp records.

Ramkumar Bala thanks Simon Wing for his assistance in providing some of the data and Erzsébet Merényi for her insights on Artificial Neural Networks.
References


Ramkumar Bala and Patricia Reiff, Department of Physics and Astronomy, Rice University, 6100 Main Street, Houston, TX 77005, USA. (ramkumar@rice.edu)

J. E. Landivar, Department of Physics, The University of Texas at Arlington, Arlington, TX 76019, USA.
Table 1. 3-hour averaged forecast summary based on a 2x2 contingency for an eight year period between 1998 and 2005.

<table>
<thead>
<tr>
<th>Forecast Attributes</th>
<th>BI &gt; 200 kV; Kp &gt; 6</th>
<th>BI &gt; 160 kV; Kp &gt; 5</th>
<th>BI &gt; 110 kV; Kp &gt; 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Skill Score</td>
<td>0.867</td>
<td>0.887</td>
<td>0.761</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.984</td>
<td>0.957</td>
<td>0.886</td>
</tr>
<tr>
<td>False Alarm</td>
<td>0.118</td>
<td>0.070</td>
<td>0.125</td>
</tr>
<tr>
<td>Probability of False Alarm</td>
<td>0.0004</td>
<td>0.0006</td>
<td>0.0052</td>
</tr>
</tbody>
</table>

Table 2. Skill score statistics (HSS) of NN Kp with a lead time of 1 hour for the test set April 2001, 2006 & 2007. Similar to Table 1, but now the discriminator is the predicted Kp value (e.g. 4) and the forecast cutoff is the measured Kp, with the same numerical value. Thus, if the model (in this case Model 1) predicts Kp of 6 or greater, the true skill score is 0.774, with only a false alarm probability of less than 1%.

<table>
<thead>
<tr>
<th>Forecast Attributes</th>
<th>Kp &lt; 4</th>
<th>Kp &gt; 4</th>
<th>Kp &gt; 6</th>
<th>Kp &gt; 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill Score</td>
<td>0.707</td>
<td>0.600</td>
<td>0.715</td>
<td>0.648</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.978</td>
<td>0.994</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>False Alarm</td>
<td>0.012</td>
<td>0.400</td>
<td>0.293</td>
<td>0.333</td>
</tr>
<tr>
<td>Probability of False Alarm</td>
<td>0.281</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 3. Table showing the prediction summary of the two linear and the four ANN Kp models. All the test results are based on data set covering April 2001, all of the 2006 & 2007. Costello’s ANN predictor results are also listed.

<table>
<thead>
<tr>
<th>Forecast Model</th>
<th>Linear Correlation</th>
<th>RMS Error</th>
<th>Slope</th>
<th>Prediction Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear 1-hour predictor using BI</td>
<td>0.712</td>
<td>1.44</td>
<td>1.1</td>
<td>0.60</td>
</tr>
<tr>
<td>Linear 3-hour predictor using BI</td>
<td>0.770</td>
<td>1.21</td>
<td>1.1</td>
<td>0.59</td>
</tr>
<tr>
<td>Model 1: ANN 1-hour predictor using BI &amp; Kp</td>
<td>0.863</td>
<td>0.71</td>
<td>0.75</td>
<td>0.74</td>
</tr>
<tr>
<td>Model 2: ANN 2-hour predictor using BI &amp; Kp</td>
<td>0.854</td>
<td>0.82</td>
<td>0.79</td>
<td>0.73</td>
</tr>
<tr>
<td>Model 3: ANN 1-hour predictor using BI</td>
<td>0.852</td>
<td>1.12</td>
<td>0.88</td>
<td>0.73</td>
</tr>
<tr>
<td>Model 4: ANN 3-hour predictor using BI</td>
<td>0.845</td>
<td>1.12</td>
<td>0.82</td>
<td>0.71</td>
</tr>
<tr>
<td>Costello ANN Predictor</td>
<td>0.760</td>
<td>1.5-2*</td>
<td>0.94</td>
<td>0.57</td>
</tr>
</tbody>
</table>

* Largest RMS error for cases of very low and very high Kp respectively.

Figure 1. 3-hour averaged log(BI) versus the following 3-hour Kp of a complete data set between 1998 and 2005 (r=0.74). Each quadrant represents either a hit, miss, false positive or a correct rejection, given a cutoff value of BI (110 kV here) and a target value of Kp (here 5).

Figure 2. Plot shows the 3-hour averaged log(BI) versus the following Kp for 2003 and 2004 with a linear correlation coefficient of 0.785. Note that the BI cut-off in this case is 100kV while the trigger level is reduced to Kp = 4.
Figure 3. 1-hour averaged log(BI) versus the following Kp (r=0.71) for an active period during 2000 and 2001 is shown here. BI cut-off in this case is 150kV.

Figure 4. True skill score statistic of input data (1998-2005) covering a solar maximum.

Figure 5. Cross-correlation coefficient of log(BI) and Kp versus the time lag. Both 3-hour averages (smooth curve) and 1-hour averages (dash-dot) display a strong correlation at 3 hour with the BI leading Kp.

Figure 6. Linear predictor: Official Kp vs Predicted Kp for April 2001, 2006 & 2007 using 3-hour averages of the BI.
Figure 7. Time series plots of the ACE and official Kp index of an event in April 2006. (a) Derived Boyle index. (b) Solar wind velocity (SWEPAM). (c) BZ (IMF). (d) Official Kp values (smooth curve) and Model 1 ANN predicted Kp (dotted curve).

Figure 8. Network architecture: A feedforward backpropagation network.

Figure 9. One hour ahead ANN predicted Kp vs Official Kp (Model 1): Simulated Kp using the BI and time history of official Kp covering April 2001, Jan-Dec 2006 and Jan-Dec 2007 ($r = 0.863$).

Figure 10. One hour ahead Kp prediction using Model 1: A 30-day interval in Dec 2006 (top) and Aug 2007 (bottom) is shown with Official Kp over-plotted on ANN Kp (dotted curve).

Figure 11. 2-hour ahead ANN predicted Kp vs Official Kp (Model 2): Simulated 2-hour Kp using the derived BI and time history of official Kp from April 2001, Jan-Dec 2006 and Jan-Dec 2007 ($r = 0.854$).

Figure 12. Kp distribution plot: Jan 2006-Dec 2007 & April 2001 with official Kp (dark histogram) and ANN Kp (pale histogram) is shown here.

Figure 13. Forecast skill score as function of Kp threshold for test years is shown here. The data covers April 2001, all of the 2006 & 2007.
Figure 14. One hour ahead ANN predicted Kp vs Official Kp (Model 3): Simulated Kp using the derived BI covering April 2001, Jan-Dec 2006 and Jan-Dec 2007 shows a linear correlation of 0.852.

Figure 15. Three hour ahead ANN predicted Kp vs Official Kp (Model 4): Simulated Kp using the derived BI covering April 2001, Jan-Dec 2006 and Jan-Dec 2007 shows a linear correlation of 0.845.

Figure 16. Auto-correlation for 3-hour Kp and cross-correlation for the official Kp vs ANN predicted 3-hour Kp (Model 4) is shown here. Note that the model which includes Kp history (Model 2) appears to lag the real data.

Figure 17. Head-to-head performance of Model 3 and Costello ANN Kp on the same data set (April 2001 & Jan-Feb 2007): (a) Model 3 predicted Kp vs measured Kp (r = 0.84) and, (b) Costello ANN predicted Kp vs measured Kp (r = 0.81).