

# Kp forecasting with a recurrent neural network

Ernest Scott Sexton\*, Katariina Nykyri, and Xuanye Ma

Center for Space and Atmospheric Research, Space Weather, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114-3900, USA

Received 16 June 2018 / Accepted 8 May 2019

**Abstract**—In an effort to forecast the planetary Kp-index beyond the current 1-hour and 4-hour predictions, a recurrent neural network is trained on three decades of historical data from NASA’s Omni virtual observatory and forecasts Kp with a prediction horizon of up to 24 h. Using Matlab’s neural network toolbox, the multilayer perceptron model is trained on inputs comprised of Kp for a given time step as well as from different sets of the following six solar wind parameters,  $B_z$ ,  $n$ ,  $V$ ,  $|\mathbf{B}|$ ,  $\sigma_B$  and  $\sigma_{B_z}$ . The purpose of this study was to test which combination of the solar wind and Interplanetary Magnetic Field (IMF) parameters used for training gives the best performance as defined by correlation coefficient,  $C$ , between the predicted and actually measured Kp values and Root Mean Square Error (RMSE). The model consists of an input layer, a single nonlinear hidden layer with 28 neurons, and a linear output layer that predicts Kp up to 24 h in advance. For 24 h prediction, the network trained on  $B_z$ ,  $n$ ,  $V$ ,  $|\mathbf{B}|$ ,  $\sigma_B$  performs the best giving  $C$  in the range from 0.8189 (for 31 predictions) to 0.8211 (for 9 months of predictions), with the smallest RMSE.

**Keywords:** neural network / space weather / machine learning

## 1 Introduction

### 1.1 Artificial Neural Networks (ANNs)

The concept of machine learning was first introduced by Samuel (1959), however, it was not until the early 2000s that the private and scientific communities were able to leverage this powerful idea to solve increasingly difficult problems, largely due in part to the explosion of both computing power and our ability to record, store, and process large amounts of data. One subset of the overarching field of machine learning is time series prediction, in which a computer’s ability to forecast future events based on prior data is tested. Chaotic time series are notoriously difficult to predict and are seen in fields ranging from signal processing (Lapedes & Farber, 1987) to economics (Kaastra & Boyd, 1996).

Inherently nonlinear in nature, forecasting is a prime candidate for the application of neural networks, which have been shown to be effective in modeling the behavior of highly complex systems. In 2006, Karunasinghe & Liong (2006) showed that the Multilayer Perceptron (MLP) model was able to outperform two nonlinear prediction methods: a local approximation using nearest neighbors, and a second order nonlinear polynomial. The Artificial Neural Network (ANN) achieved an error

rate half that of the best nonlinear predictor when forecasting a noise-free Lorentz series (Karunasinghe & Liong, 2006). More recently, Ayala Solares et al. (2016) showed that a network similar to the one proposed here performed best in predicting Kp for some horizon  $h$  when trained on inputs with a time lag equal to  $h$ .

A full description of how the ANN operates is beyond the scope of this study, but at a high level it can be viewed as a massively parallel computational system with a very large number of tunable parameters. It consists of basic computational elements called neurons, which share connections with every other neuron in the layers preceding and following the layer in which a given neuron resides. The influence of the connections, and thus the individual neurons, are controlled via scalar weights whose values are adjusted during the training process. Each neuron is modeled as a summation of weights  $w_{kj}$  for inputs  $x_k$  (Stearns, 1985) whose output is a function the sigmoid activation function  $\phi$ .

$$y_k = \phi \left( \sum_{j=0}^N w_{kj} x_j - \mu \right) \quad (1)$$

Training the network is accomplished via the backpropagation algorithm (Werbos, 1990), a gradient descent method which follows a simple error-correction scheme for updating

\*Corresponding author: [essexton@protonmail.com](mailto:essexton@protonmail.com)

weights until a minimal error  $\varepsilon$  is achieved or a desired number of iterations is completed. For a forward-pass of inputs  $\mathbf{x}$  through the network, with desired output  $Y$  and learning rate  $\gamma$ :

**Algorithm 1:** Backpropagation

**while**  $E > \varepsilon$  **do**

$$\left\{ \begin{array}{l} y = \phi(\sum \mathbf{w}\mathbf{x} - \mu); \\ \mathbf{E} = \frac{1}{2} \sum ||Y - y||^2; \\ \nabla \mathbf{E} = \frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \dots, \frac{\partial E}{\partial w_m}; \\ \Delta w_i = -\gamma \frac{\partial E}{\partial w_i}, i = 1, \dots, m; \end{array} \right.$$

**end**

Training the network results in an optimized  $\mathbf{w}$  unique to the network architecture and desired output. Predictions are then performed with a single forward pass of a new set of inputs.

## 1.2 Space weather and the planetary K-index

The cumulation of Sun driven physical phenomena taking place in the solar corona, solar wind, Earth's magnetosphere, ionosphere, and thermosphere, is called Space Weather (SW). Space weather affects the operation of space and ground-based technological systems and services (e.g., GPS technology) and can present a threat to life (Wrenn, 1995; Hellweg & Baumstark-Khan, 2007). The majority of air and space transport activities are at risk during active space weather, in particular navigation, communications, and avionics. Improved forecasting of geomagnetic activity conditions resulting from space weather is therefore important.

The time series considered for forecasting here is the planetary K-index, Kp, a measure of geomagnetic activity arising from the interaction of the solar wind with the Earth's magnetosphere, and which was first introduced by Bartels (1949). Geomagnetic disturbances in the Earth's magnetosphere caused by the interaction of the solar wind are continuously measured by ground stations across the globe and are used to derive K-index at each station. The planetary K-index, the Kp index, is then derived from all the K-indices around the Earth. Kp values range from 0 to 9, with a higher value corresponding to a higher level of geomagnetic activity: one being a calm, five or more indicating a geomagnetic storm, and indices of 8–9 corresponding to very major geomagnetic storms.

The higher values of Kp index, characterizing stronger geomagnetic activity, result from enhanced solar forcing by various solar wind drivers such as Coronal Mass Ejections (CMEs) (Gonzalez & Tsurutani, 1987; Wilson, 1987), Co-rotating Interaction Regions (CIRs) (Alves et al., 2006), solar energetic particles and solar flares (Shibata & Magara, 2011). Solar wind speed typically varies between  $350 \text{ km s}^{-1}$  and  $800 \text{ km s}^{-1}$ , but CMEs can travel as fast as  $3000 \text{ km s}^{-1}$ . When the CME speed is higher than the speed of the ambient solar wind, a shock wave can form in front of the CME and accelerate particles. The interplanetary CMEs and their sheath regions, characterized by magnetic and velocity field fluctuations, can drive intense geomagnetic storms (Kilpua et al., 2017 and references therein) and lead to enhanced geomagnetic activity. The magnetic field of the Sun is frozen into solar wind flow. The direction of this

Interplanetary Magnetic Field (IMF) and SW properties (e.g., density, speed, and temperature) at the Earth's orbit vary. IMF orientation at Earth's location determines where the SW energy can most effectively access and penetrate through the Earth's magnetic shield, the magnetopause. When the solar magnetic field lines are anti-parallel to the Earth's magnetic field ( $B_z < 0$ ), the solar wind magnetic field lines can connect to the Earth's magnetic field lines and the magnetopause gets penetrated by the solar wind plasma. Magnetic reconnection was first proposed by Dungey (1961) and observational evidence of electron-scale microphysics of reconnection was recently found by NASA's MMS mission (Burch et al., 2016). Similarly, if the magnetic field direction of the CME is opposite to that of the Earth's dipolar magnetic field, the resulting geomagnetic storm will be larger than if the fields are in the same direction. The enhanced magnetic perturbations, captured by ground magnetometers and that are used in determining the Kp index, arise due to various current systems that get enhanced and modified during enhanced solar wind driving.

In an effort to prepare for times of high solar activity, NASA currently uses the Wing Kp model (Wing et al., 2005) to provide 1-hour and 4-hour predictions, updated every 15 min. The operational Wing model is a recurrent neural network, originally trained on solar wind and IMF data from 1975 to 2001. It takes as inputs solar wind  $|V_x|$ ,  $n$ , IMF  $|\mathbf{B}|$ ,  $B_z$ , and the current nowcast Kp prediction. A comprehensive study comparing this model to other Kp models showed better performance across all values of Kp, but the Wing model showed particularly increased performance for moderate and active times of  $Kp > 4$ . Previous work using a single neural network and hybrid models of multiple specialized neural networks has produced models that predict Kp (Boberg et al., 2000; Kutiev et al., 2009) up to 6 h in advance, but a reliable, even longer forecast model would be desirable for allowing for a better preparation during times of high solar activity.

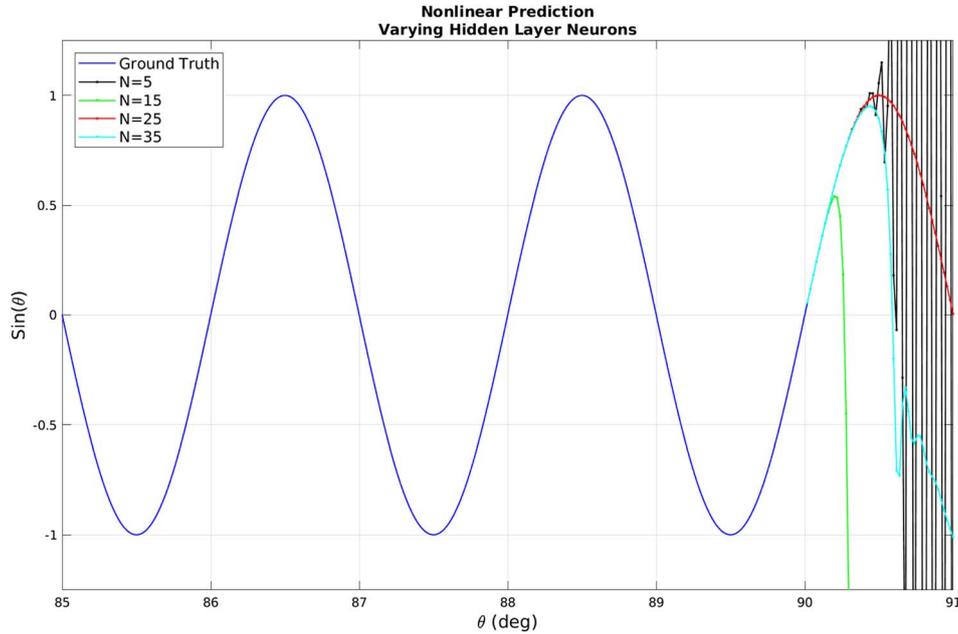
A technique that has proven successful in modeling highly nonlinear systems known as NARMAX (nonlinear autoregressive moving average model with exogenous inputs) has been studied as a means of predicting Kp (Balikhin et al., 2011; Wei et al., 2011; Rasttter et al., 2013; Gu et al., 2017). Similar to the network proposed here, NARMAX models are fit to a lagged time series and make predictions for some time  $t$  in the future. NARMAX models differ, however, in that the system noise is used as an input to the model. The ANN proposed here uses only solar wind parameters as input and effectively learns how to model noise within the system to produce a prediction.

In this paper we discuss forecasting the Kp index using a neural network trained on five solar wind and IMF parameters:  $B_z$ , solar wind proton density  $n$ , plasma speed  $V$ , IMF magnitude  $|\mathbf{B}|$ , and the RMS standard deviation of the IMF  $\sigma_B$ .

## 2 Methods

### 2.1 Parameter tuning

High resolution (1 min) and low resolution (1 h averaged) data were obtained from NASA's Omniweb service. Both the test and validation sets were comprised of data from 1986 to 2016, and the test set from 2017. This date range was chosen as it contained two full solar cycles, and previous work found



**Fig. 1.** Prediction using different sizes of the hidden layer for a toy nonlinear model. The ground truth is shown in blue line, while the black, green, red, and cyan lines represent the prediction results using 5, 15, 25, and 35 hidden layers, respectively.

that models shown to be accurate during times of low activity were not necessarily accurate for predicting times of high activity (Boberg et al., 2000). Training on two full solar cycles ensured that our model was exposed to periods of high and low activity.

Matlab's Neural Network Toolbox was used to build the MLP model, starting with a single hidden layer of 10 neurons and a linear output layer. Prior to defining the full set of features and final network architecture, the network was trained on the full data set with  $B_z$  as the only input, and its ability to predict a single time step ahead was evaluated. Initial tests were conducted to determine the maximum viable input lag time,  $\tau$ , for this study, and with  $B_z$  as the only input  $\tau > 36$  yielded diminishing returns in prediction accuracy.

A parametric study was conducted to determine the ideal number of neurons in the hidden layer, as too few neurons leads to a poor fit of the model, and too large of a hidden layer increases computational costs with diminishing returns as well as introduces a vulnerability to overfitting, as indicated by the cyan line in Figure 1.

A toy model is shown in Figure 1, illustrating the effect of adding more neurons to the hidden layer. An insufficient hidden layer number is likely to cause unstable predictions (see, black and green lines), while an overly large hidden layer number will cause overfitting (see cyan line). To a limit, an increased number of neurons allows the network to more effectively map the inputs to the desired output value by increasing the number of available weights that are optimized during training. For the simple nonlinear case in Figure 1, the ideal number of neurons was 25 (red line), however, this parameter is unique to each data set and can be constrained by the computational ability of the machine. The parametric study on the dataset showed that the best compromise between accuracy and efficiency was achieved

with a hidden layer size of 28 neurons. Adding more neurons to the hidden layer showed increasingly expensive computation with little effect on accuracy. The final MLP model is shown in Figure 2, with a hidden layer consisting of 28 neurons.

The generic MLP model is a time-lagged recurrent neural network, whose output at  $t + 1$  is a function of the most recent output and input, as well as a tunable number of previous outputs and inputs, and can be modeled as

$$y_{t+1} = F(y_t, y_{t-1}, \dots, y_{t-\tau}, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-\tau}) \quad (2)$$

where  $\tau$  is the earliest time used by the network to predict a given  $y_{t+1}$ . For the purposes of this study, a given  $y$  is the Kp value at the corresponding hourly time step, and  $\mathbf{x}$  the full feature vector for the same  $t$ . Determining the best set of time-lagged features in  $\mathbf{x}$  was a large part of this study and a more thorough explanation is provided in Section 2.3.

Similar to the size of the hidden layer, tuning for the optimal value of  $\tau$  is a tradeoff between accuracy and computational expense, and is unique to the training data. The network was trained with  $\tau = [0, 6, 12, 18, 24, 30, 36]$ h and predicted  $Kp_{t+12}$  at 1 h intervals. The average prediction error for each  $\tau$  can be found in Table 1.

Both average prediction error and model complexity scale with  $\tau$ ; a simpler network looking back 1 h cannot capture the system dynamics as well as one looking back 24 h, but this comes at the cost of more expensive computation. Following the trials in Table 1,  $\tau$  values of 24 and 30 were both tested when making final predictions. All training was completed on a Cray CS400 supercomputer and the trained models were deployed for prediction on an Intel i5-3570k.

Broadly speaking, more training data results in a model that can provide more accurate and precise results, assuming that the training set is diverse enough to cover the entire solution space.

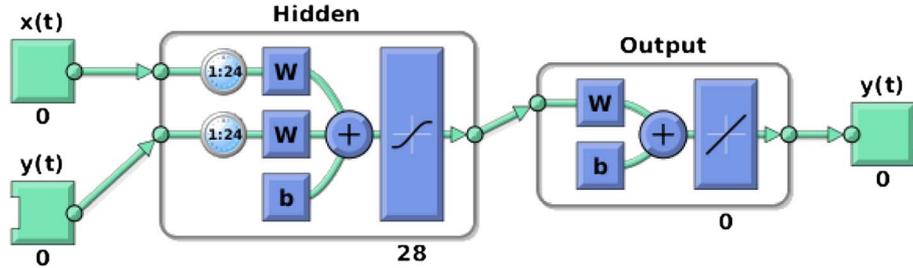


Fig. 2. MLP Architecture. Image credit: Mathworks MATLAB R2014b (MATLAB, 2014).

Table 1. Average prediction errors.

$\tau$ (h)	Avg error
1	0.424
6	2.596
12	0.512
18	0.376
24	0.316
30	0.120
36	0.168

Attempts were made at training networks with smaller data sets, but the resulting models proved to be less stable and provide less accurate results when predicting more than an hour or two in the future. The high nonlinearity of the data is also a driver of the requisite amount of training data required to get an accurate model. A linear model needs only a few data points, and the toy model shown in Figure 1 was trained on just a couple thousand data points and successfully modeled the nonlinear function, but the periodicity of the toy example means it is an inherently easier function to model when compared to the more complex Kp index.

## 2.2 Closed loop algorithm

Initial Kp predictions showed low error rates for timesteps 1 or 2 h ahead, but quickly diverged for a longer prediction horizon. During training, the nonlinear autoregressive neural network (NARXnet) uses as a target for prediction the true Kp values. However, in a fully closed-loop architecture post training, the true Kp value is missing as a target. Using the standard closed-loop network where the current prediction of  $Kp(t+1)$  is used as an input to predict  $Kp(t+2)$ , we successfully predicted 1 h ahead, and 2 h predictions were less accurate. Initial attempts at forecasting times greater than 2 h resulted in the network reaching a steady state output where all successive predictions were the same value. A guided closed-loop architecture was proposed, where we assume that the Kp value 1 h in the future will not largely deviate from the previous 10 h. A single time step in the future is predicted and then averaged with the preceding 10 predictions. This single averaged value is then used as the target for the next prediction. The process is repeated in a closed loop to obtain a full day's worth of hourly predictions. During this time, the network is never exposed to the true value of Kp, but uses  $y'_{t+1}$  in its place. For a given prediction  $y_t$  and a target  $y'_{t+1}$ :

Table 2. Feature vector correlation values 31 24-hour predictions.

Feature vector	Correlation
$B_z, \sigma_B$	0.7637
$B_z, \sigma_{Bz}$	0.7618
$B_z, n, \sigma_B$	0.8065
$B_z, n, \sigma_{Bz}$	0.8015
$B_z, n, V$	0.8056
$B_z, n, V,  B $	0.8090
$B_z, n, V,  B , \sigma_{Bz}$	0.6402
$B_z, n, V,  B , \sigma_B$	0.8189

**Algorithm 2:** Guided Closed Loop Prediction for Prediction Horizon =  $t_f$

**Result:** Multi-step Prediction

Train on  $N$  samples;

$$y_{t+1} = F(y_t, y_{t-1}, \dots, y_{t-\tau}, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-\tau});$$

**while**  $t < t_f$  **do**

$$y'_{t+1} = \text{avg}(y_t, y_{t-1}, \dots, y_{t-10});$$

$$y_{t+1} = F(y'_{t+1}, y_t, y_{t-1}, \dots, y_{t-\tau}, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-\tau});$$

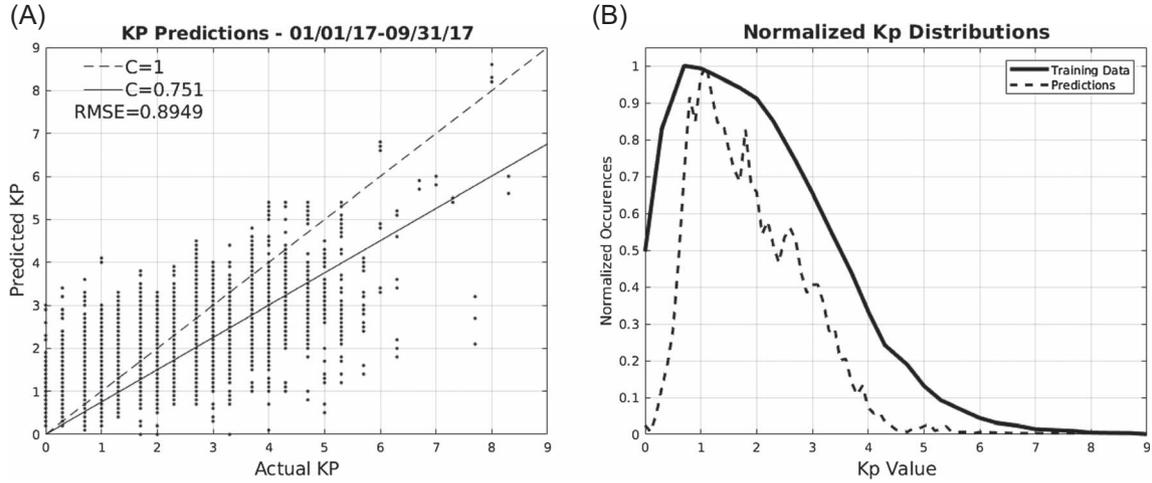
**end**

Prior to training on the full data set, the guided closed loop network's ability to predict nonlinear data was tested on the simple nonlinear function:  $Y = \sin(\theta)$ . Training consisted of 4500 values of  $\theta$  and  $\sin(\theta)$  and predictions were made 50 steps forward. With a hidden layer size of 25 neurons and  $\tau = 2$ , the average prediction error was  $9.8 \times 10^{-4}$ . A plot of the validation model is shown in Figure 1.

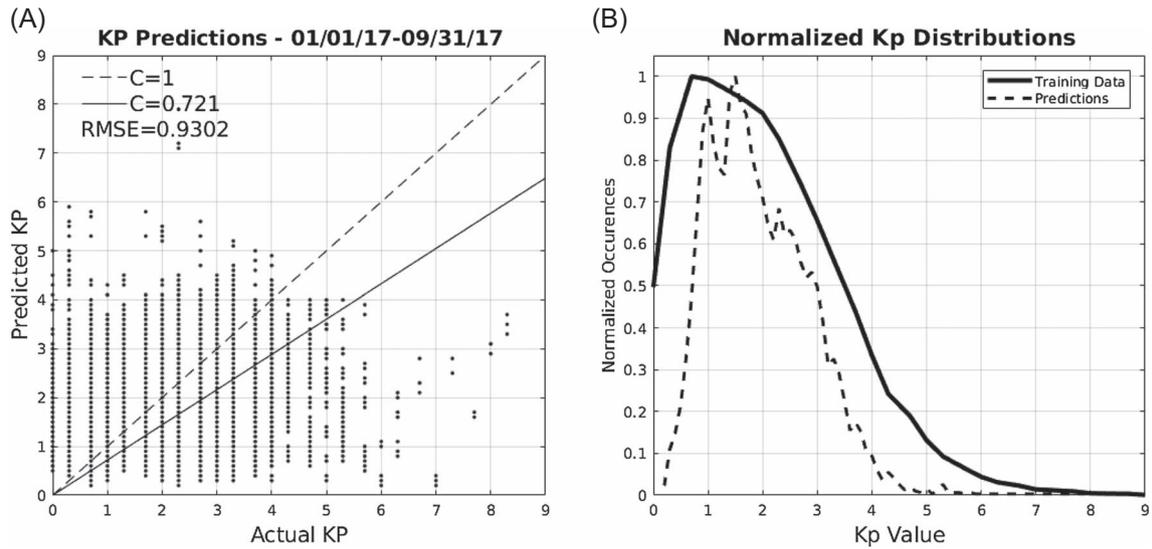
## 2.3 Feature selection

While the NARXNet was shown to be accurate in predicting a simple nonlinear time series, its accuracy is highly dependent on the training data. Previous work using an ANN to predict Kp showed the  $B_z$  component of the IMF (Boberg et al., 2000) to be a viable input to the network, and  $\sigma_B$ , RMS deviation of the IMF, has been shown to correlate with Kp (Pudovkin et al., 1980). This initial feature set [ $B_z, \sigma_B$ ] was used as an input and the network was trained on low resolution data from 1/1/1986 to 12/31/2016.

The data were preprocessed prior to both training and testing, as the inputs to the network must be normalized to [0,1]



**Fig. 3.** (A) 6528 Kp predictions with  $\tau = 24$ , 01/01/2017–09/30/17 with inputs  $B_z$ ,  $n$ ,  $\sigma_B$ . Each dot is a prediction plotted against its corresponding true Kp value, and the solid line indicates the line of best fit with a slope of  $C = 0.751$ . Also shown is a line of slope 1, corresponding to an ideal correlation value. (B) Frequency of Kp values for full data set 1986–2016 and 2017 predictions. The imbalance of low Kp to high Kp in both sets is evident.



**Fig. 4.** (A) 6528 Kp predictions with  $\tau = 30$ , 01/01/2017–09/30/17 with inputs  $B_z$ ,  $n$ ,  $\sigma_B$ . (B) Frequency of Kp values for full data set 1986–2016 and 2017 predictions. Compared with [Figure 3](#), a smaller range of predicted Kp values is seen in both plots.

due to clamping of the sigmoid activation function for the neurons in the hidden layers. Saturated sensor values were replaced with NaN placeholders to ensure they would be withheld from training, and each input was normalized. The normalization factor  $\alpha$  applied to Kp was stored and  $\alpha^{-1}$  was applied to the forecasts prior to calculating prediction error or correlation values to return the forecasted Kp to its original range. The rescaled prediction value was then rounded to the nearest integer value to more accurately reflect the true Kp scale.

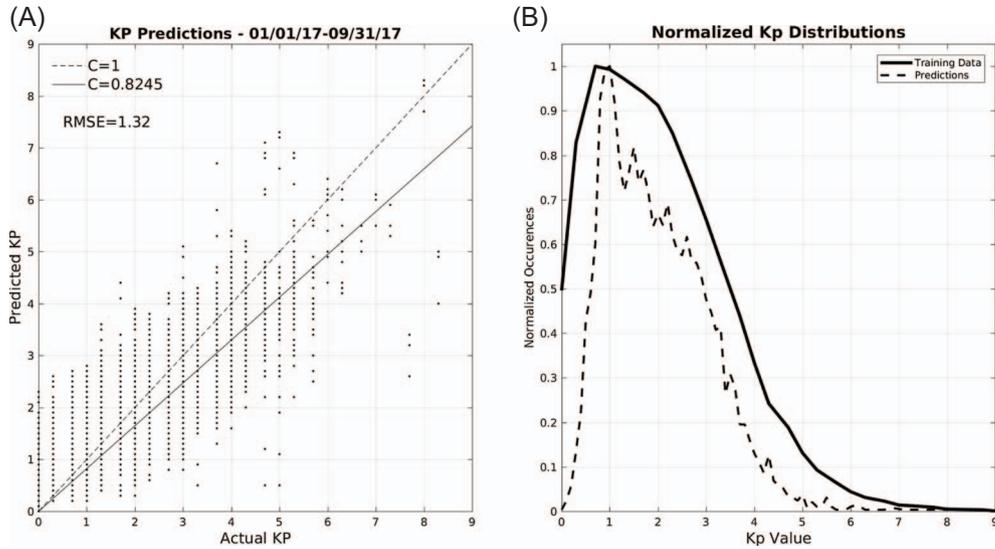
$$\alpha = \frac{Y - \min(Y)}{\max(Y) - \min(Y)} \quad (3)$$

Twenty-four hour predictions made with the initial feature vector showed an average correlation of 0.7637 when

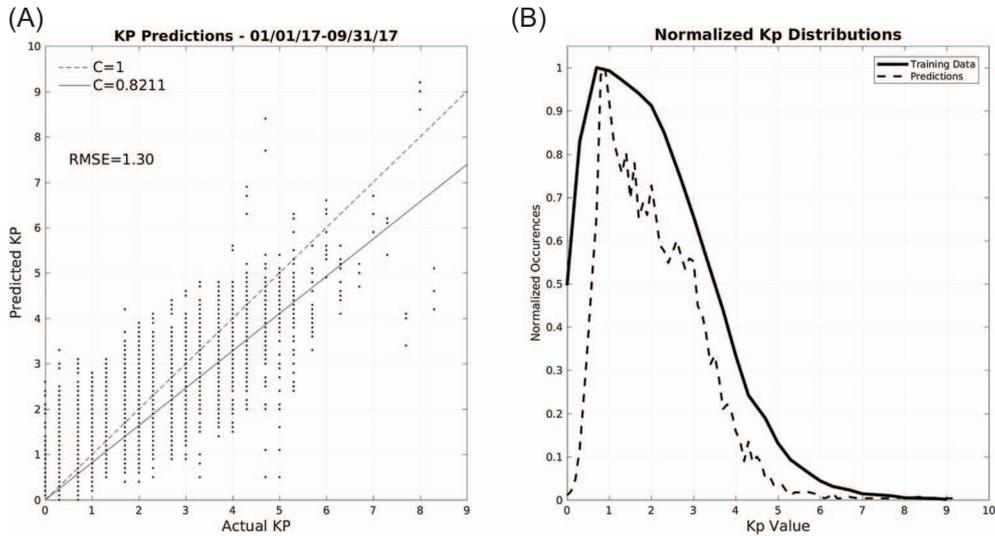
tested against the 24 h data from 01/01/2017. Correlation between the set of predicted  $y_p$  and actual  $Y_a$  Kp values was calculated as:

$$C = \frac{1}{N\sigma_p\sigma_a} \sum_{j=0}^N (y_j - \langle y_p \rangle)(Y_j - \langle Y_a \rangle) \quad (4)$$

The network was also trained on a downsampled (15 min) version of the high resolution (1 min) data which showed high accuracy in predicting multiple time steps ahead. However, due to the relatively short interval between each step, the prediction horizon was too short to be useful and all further testing was conducted with the low resolution 1 h data.



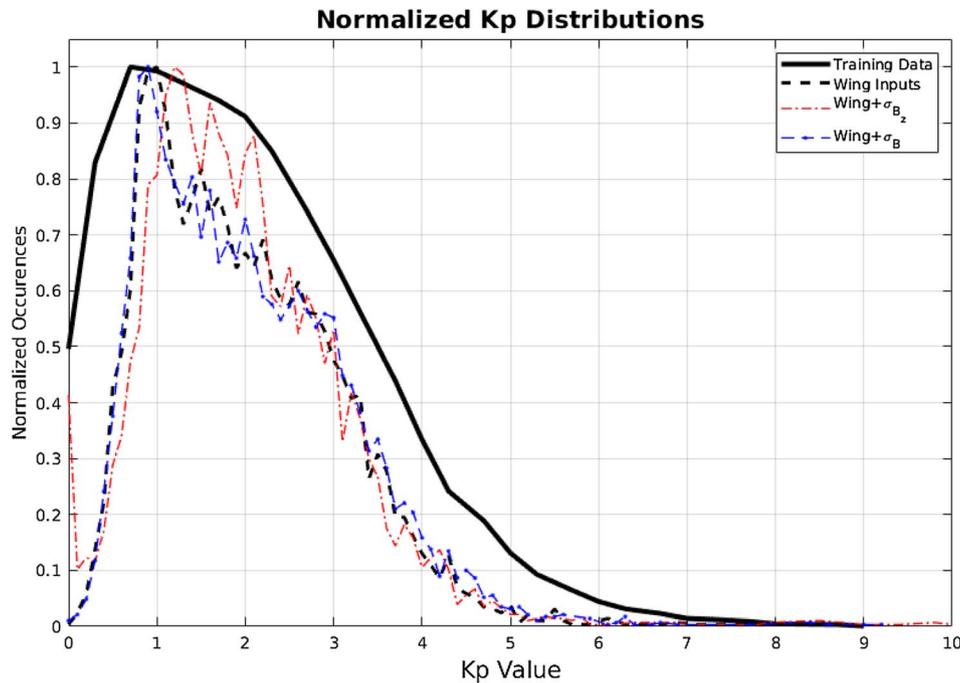
**Fig. 5.** (A) 6528 Kp predictions 01/01/2017–09/30/17 from a network trained on the original Wing model inputs:  $B_z$ ,  $n$ ,  $V$ ,  $|\mathbf{B}|$ . (B) Frequency of Kp values for full data set 1986–2016 and 2017 predictions. The model consistently predicts Kp values lower than the actual measured values.



**Fig. 6.** (A) 6528 Kp predictions 01/01/2017–09/30/17 from the network trained on the original Wing model inputs with an additional parameter:  $B_z$ ,  $n$ ,  $V$ ,  $|\mathbf{B}|$ ,  $\sigma_B$ . (B) Frequency of Kp values for full data set 1986–2016 and 2017 predictions. Compared with Figure 5, the distribution of predictions more closely matches that of the training set.

Measured by the OMNI virtual observatory, NASA makes available 48 attributes from the IMF, solar wind plasma, solar indices, and plasma particles. Testing all possible combinations of these features as inputs would be a noble yet time consuming experiment. For the purposes of this study we selected  $B_z$ ,  $\sigma_{B_z}$ ,  $\sigma_B$ , proton number density  $n$ , and proton speed  $V$  as potential inputs to the network. The selection of  $n$  and  $V$  naturally represent the solar wind property, and  $B_z$  is directly related to the dayside reconnection.  $\sigma_{B_z}$  and  $\sigma_B$  are used to reflect the variation of IMF, which plays an important role for magnetosphere (e.g., triggering reconnection, substorm). In principle, there are endless combinations that can be used as inputs.

Here we only present the best results based on trial and error. Also tested were the inputs provided to the model in Wing et al. (2005) but using  $|V|$  instead of  $V_x$  in training, as well as the addition of  $\sigma_B$  and  $\sigma_{B_z}$ . Table 2 provides a summary of the most successful feature vectors, with the correlation value for a 24 h prediction horizon, repeated for the entire month of January, 2017. The addition of  $\sigma_{B_z}$  to the feature vectors consistently decreases prediction accuracy, and the effect is most pronounced on the Wing et al. (2005) features. We attribute the loss of accuracy to the model reaching a local minimum in training that it avoids when  $\sigma_{B_z}$  is withheld from the training data.



**Fig. 7.** Frequency of Kp values for full data set 1986–2016 and 2017 predictions of three networks trained on the inputs from Wing et al. (2005) with the addition of  $\sigma_{B_z}$  and  $\sigma_B$ .

### 3 Results

This study generated 272 24 h forecasts for January, 2017 through September, 2017 based on the set of inputs  $[B_z, n, V, |B|, \sigma_B]$  from 1986 to 2016. The network with  $\tau = 24$  showed a correlation value  $C = 0.7510$  with RMSE = 0.8949, and the network with  $\tau = 30$  showed correlation of  $C = 0.7211$  with RMSE = 0.9302. This is somewhat consistent with Wing et al. (2005) and Johnson & Wing (2005) who found that the self significance in the Kp time series peaks at around 30–50 h, suggesting an optimal  $\tau$  for 24 h ahead prediction should be at least around 26 h. The small difference between this study and their results may be explained by differing network architectures or distributions of the data chosen as training sets. Data from 12/31/16 were initialized and the model predicted 24 h in advance. These were replaced with data from 1/1/16 and the process repeated for 272 days. A plot of the set of predicted vs actual values is found in Figure 3. Predicted values are plotted against the actual Kp value for each hour of the 272 days. The true Kp value was obtained from NASA’s OMNIweb service. A solid line indicating the correlation value is shown, as well as a line with a slope of 1, corresponding to an ideal correlation. The model shows relatively high accuracy for  $Kp \leq 3$  but consistently makes low predictions during times of higher solar activity. Normalized total occurrences of each value of Kp for both the entire data set as well as the predictions are shown in the right panel of Figures 3 and 4. A decline in the number of  $Kp > 4$  available for training is evident, and is reflected by the sharp decrease in predictions  $\geq 4$ .

Initial testing of the inputs indicated that  $\sigma_B$  increased prediction accuracy, as indicated in Table 2. To test whether this was unique to the feature vector used in this study, our network

was trained on the features used for the Wing model, but using data from the same date range as our inputs from 1986 to 2016. The model was then trained with the same inputs, but with the addition of  $\sigma_B$  and  $\sigma_{B_z}$ . Predictions made for January 2017–September 2017 from the network trained using the original Wing inputs showed a correlation of 0.8245, with RMSE = 1.32. The addition of  $\sigma_{B_z}$  during training reduced the prediction correlation to 0.6939 with RMSE = 1.75, and  $\sigma_B$  performed similarly to Wing inputs with correlation of 0.8211 with slightly smaller RMSE = 1.30. Figures 5 and 6 show the predictions and distributions for the networks, respectively, and Figure 7 plots the distributions of the original Wing inputs against our modified inputs. Given the results from our input tests in Table 2, as well as the increase in correlation from the original Wing model inputs,  $\sigma_B$  appears to positively influence the ability of the network to make accurate predictions for extended horizons.

### 4 Conclusions

We have shown that the proposed neural network architecture is a viable means of forecasting the Kp-index during times of low solar activity. Inherently reliant on large data sets, the ANN proposed here benefited from the abundance of historical data, but an imbalanced data set lead to imbalanced predictions. The operational Wing model trained on  $[|V_x|, n, B, B_z, Kp]$  achieved a correlation coefficient,  $C$ , of 0.92 when predicting 1 h ahead, and 0.79 when predicting 4 h ahead. Here our Wing-like model but using  $|V|$  instead of  $V_x$  in training, gave a  $C$  of 0.809–0.825 with RMSE of 1.32 for 24 h prediction.

When adding  $\sigma_B$  to the Wing-like inputs performed similarly giving  $C$  of 0.819–0.821. [Boberg et al. \(2000\)](#) encountered similar difficulties in predicting high Kp values and proposed a hybrid model consisting of two networks, one trained on low Kp values and the other on high values. Their hybrid model was able to achieve a correlation of 0.76 for 3 and 6 h predictions, comparable to the accuracy presented here but over a much shorter prediction horizon; our non-hybrid model has shown a similar accuracy for a full 24 h period.

For the current model, the lag time  $\tau$ , in principle, represents the typical time scale of the system to respond to a perturbation, which plays an important role. Our results show that the optimal value of  $\tau = 24$  and 30, which is mostly consistent with the [Johnson & Wing \(2005\)](#) results. The difference between our results and their results may come from the closed loop algorithm used in this study, difference in network architectures, or the selection of data sets used for training; a further comprehensive comparison is required. It is also important to notice that [Johnson & Wing \(2005\)](#) also found that the significance in the solar wind driver variables,  $VB_s$ , and Kp time series peaks at  $\tau = 3$  h and decays very rapidly, suggesting that  $\tau$  for the solar wind can be much smaller than the  $\tau$  for Kp. Thus, it is interesting to investigate the performance of the model by using different lag time  $\tau$  for solar wind input and Kp, which is planned as a future study.

Techniques exist for training with highly imbalanced data sets such as oversampling ([Han et al., 2005](#)) and rebalancing, wherein over-represented samples are intentionally withheld to reduce bias and further work might investigate the effect of these methods on accuracy. Also worth investigating is the effect of adding additional hidden layers to the network, and the application of deep learning to time series prediction, as it has recently proven to be highly effective at solving notoriously difficult computational problems.

**Acknowledgements.** The work of E. Scott Sexton, Katariina Kykyri, and Xuanye Ma was supported by NASA grant NNX16AF89G. Computational work was supported by the Vega super computer at Embry-Riddle Aeronautic University in Daytona Beach, FL. The editor thanks two anonymous referees for their assistance in evaluating this paper.

## References

- Alves MV, Echer E, Gonzalez WD. 2006. Geoeffectiveness of corotating interaction regions as measured by Dst index. *J Geophys Res (Space Phys)* **111**: A07S05. DOI: [10.1029/2005JA011379](https://doi.org/10.1029/2005JA011379).
- Ayala Solares JR, Wei H-L, Boynton RJ, Walker SN, Billings SA. 2016. Modeling and prediction of global magnetic disturbance in near-Earth space: A case study for Kp index using NARX models. *Space Weather* **14**(10): 899. DOI: [10.1002/2016sw001463](https://doi.org/10.1002/2016sw001463).
- Balikhin MA, Boynton RJ, Walker SN, Borovsky JE, Billings SA, Wei HL. 2011. Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit. *Geophys Res Lett* **38**(18): L18105. DOI: [10.1029/2011gl048980](https://doi.org/10.1029/2011gl048980).
- Bartels J. 1949. The standardized index, Ks, and the planetary index, Kp. *IATME Bull 12b* **97**: 2021.
- Boberg F, Wintoft P, Lundstedt H. 2000. Real time Kp predictions from solar wind data using neural networks. *Phys Chem Earth Part C: Solar Terr Planet Sci* **25**(4): 275–280.
- Burch JL, Torbert RB, Phan TD, Chen L-J, Moore TE, et al. 2016. Electron-scale measurements of magnetic reconnection in space. *Science* **352**(6290): L03803. DOI: [10.1126/science.aaf2939](https://doi.org/10.1126/science.aaf2939), URL <http://science.sciencemag.org/content/352/6290/aaf2939>.
- Dungey JW. 1961. Interplanetary magnetic field and the auroral zones. *Phys Rev Lett* **6**(2): 47.
- Gonzalez WD, Tsurutani BT. 1987. Criteria of interplanetary parameters causing intense magnetic storms ( $Dst < -100$  nT). *Planet Space Sci* **35**(9): 1101–1109. DOI: [10.1016/0032-0633\(87\)90015-8](https://doi.org/10.1016/0032-0633(87)90015-8), URL <http://www.sciencedirect.com/science/article/pii/0032063387900158>.
- Gu Y, Wei H-L, Boynton RJ, Walker SN, Balikhin MA. 2017. Prediction of Kp index using NARMAX models with a robust model structure selection method. In: 2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAD), Targoviste, Romania, 29 June–1 July 2017. DOI: [10.1109/ecai.2017.8166414](https://doi.org/10.1109/ecai.2017.8166414).
- Han H, Wang W-Y, Mao B-H. 2005. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: International Conference on Intelligent Computing, Hefei, China, August 23–26, 2005.
- Hellweg CE, Baumstark-Khan C. 2007. Getting ready for the manned mission to Mars: the astronauts risk from space radiation. *Naturwissenschaften* **94**(7): 517–526.
- Johnson JR, Wing S. 2005. A solar cycle dependence of nonlinearity in magnetospheric activity. *J Geophys Res: Space Phys* **110**(A4): A04211. DOI: [10.1029/2004JA010638](https://doi.org/10.1029/2004JA010638), URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JA010638>.
- Kaasra I, Boyd M. 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* **10**(3): 215–236. DOI: [10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9).
- Karunasinghe DS, Liong S-Y. 2006. Chaotic time series prediction with a global model: Artificial neural network. *J Hydrol* **323**(1–4): 92–105. DOI: [10.1016/j.jhydrol.2005.07.048](https://doi.org/10.1016/j.jhydrol.2005.07.048).
- Kilpua E, Koskinen HEJ, Pulkkinen TI. 2017. Coronal mass ejections and their sheath regions in interplanetary space. *Living Rev Solar Phys* **14**(1): 5. DOI: [10.1007/s41116-017-0009-6](https://doi.org/10.1007/s41116-017-0009-6).
- Kutiev I, Muhtarov P, Andonov B, Warnant R. 2009. Hybrid model for nowcasting and forecasting the K index. *J Atmos Solar-Terr Phys* **71**(5): 589–596.
- Lapedes A, Farber R. 1987. *Nonlinear signal processing using neural networks: Prediction and system modelling*. Los Alamos National Laboratory Tech Report LA-UR87-2662.
- MATLAB. 2014. *Version 8.4 (R2014b)*, The MathWorks Inc, Natick, MA.
- Pudovkin MI, Shukhtina MA, Poniavin DI, Zaitseva SA, Ivanov OU. 1980. On the geoefficiency of the solar wind parameters. *Ann Geophys* **36**: 549–553.
- Rasttter L, Kuznetsova MM, Glocher A, Welling D, Meng X, et al. 2013. Geospace environment modeling 2008–2009 challenge: Dst index. *Space Weather* **11**(4): 187–205. DOI: [10.1002/swe.20036](https://doi.org/10.1002/swe.20036).
- Samuel AL. 1959. Some studies in machine learning using the game of checkers. *IBM J Res Dev* **3**(3): 210–229. DOI: [10.1147/rd.33.0210](https://doi.org/10.1147/rd.33.0210).
- Shibata K, Magara T. 2011. Solar flares: Magnetohydrodynamic processes. *Living Rev Solar Phys* **8**(6). DOI: [10.1007/lrsp-2011-6](https://doi.org/10.1007/lrsp-2011-6), URL <http://www.livingreviews.org/lrsp-2011-6>.
- Stearns SD. 1985. *Adaptive signal processing*.
- Wei H-L, Billings SA, Sharma AS, Wing S, Boynton RJ, Walker SN. 2011. Forecasting relativistic electron flux using dynamic multiple regression models. *Ann Geophys* **29**(2): 415–420. DOI: [10.5194/angeo-29-415-2011](https://doi.org/10.5194/angeo-29-415-2011).

- Werbos PJ. 1990. Backpropagation through time: What it does and how to do it. *Proc IEEE* **78(10)**: 1550–1560.
- Wilson RM. 1987. Geomagnetic response to magnetic clouds. *Planet Space Sci* **35(3)**: 329–335. DOI: [10.1016/0032-0633\(87\)90159-0](https://doi.org/10.1016/0032-0633(87)90159-0), URL <http://www.sciencedirect.com/science/article/pii/0032063387901590>.
- Wing S, Johnson JR, Jen J, Meng C-I, Sibeck DG, et al. 2005. Kp forecast models. *J Geophys Res: Space Phys* **110(A4)**. DOI: [10.1029/2004ja010500](https://doi.org/10.1029/2004ja010500).
- Wrenn GL. 1995. Conclusive evidence for internal dielectric charging anomalies on geosynchronous communications spacecraft. *J Spacecraft Rockets* **32(3)**: 514–520.

**Cite this article as:** Sexton ES, Nykyri K & Ma X 2019. Kp forecasting with a recurrent neural network. *J. Space Weather Space Clim.* **9**, A19.