AGU PUBLICATIONS

Space Weather

RESEARCH ARTICLE
10.1002/2014SW001075

Special Section:
NASA/NSF Space Weather Modeling Collaboration: Advancing Space Weather Modeling for Improved Specification, Forecasting and Mitigation

Key Points:
- Real-time forecasting
- Model validation
- *Kp, Dst*, and *AE* forecast

Correspondence to:
R. Bala,
ramkumar@rice.edu

Citation:
Bala, R., and P. Reiff (2014), Validating the Rice neural network and the Wing *Kp* real-time models, *Space Weather*, *12*, 417–425, doi:10.1002/2014SW001075.

Received 21 APR 2014
Accepted 28 MAY 2014
Accepted article online 2 JUN 2014
Published online 19 JUN 2014

Validating the Rice neural network and the Wing *Kp* real-time models

Ramkumar Bala[1] and Patricia Reiff[1]

[1] Department of Physics and Astronomy, William Marsh Rice University, Houston, Texas, USA

Abstract The Rice neural network models of *Kp* have been running in real time at http://mms.rice.edu/realtime/forecast.html since October 2007; *Dst* and *AE* models were added to our operations in May 2010. All these models use the Boyle index as basis functions computed from ACE real time inputs. Later, two more driving functions were included in November 2012: (a) the "Ram" functions that had dynamic pressure term added to the Boyle index and (b) the Newell functions. The Wing models are a set of neural network-based *Kp* forecast models adopted by NOAA/Space Weather Prediction Center in March 2011 to supersede the Costello *Kp* model. This study indicates that any of the three Rice neural net predictors had a better success rate than the Wing model in predicting *Kp* ($r = 0.828$ with Boyle, $r = 0.843$ with Ram, and $r = 0.820$ with Newell for 1 h predictions; similarly, $r = 0.739$, 0.769, and 0.755 for 3 h predictions) in real time. In a head-to-head challenge using harvested real-time outputs between April 2011 and February 2013, the Rice Boyle *Kp* models predicted better than the Wing models (0.771 versus 0.714 for 1 h predictions and 0.770 versus 0.744 for 3 h predictions). In addition, Wing's prediction was missing more often than the Rice prediction (≈6% versus 4.6%), meaning it had less reliability. The Rice models also predict *AE* ($r = 0.811$ with Boyle; 0.806 with Ram; 0.765 with Newell, and 0.743 with Boyle; 0.747 with Ram for 1 h and 3 h predictions) and pressure-corrected *Dst* ($r = 0.790$; 0.767, and 0.704, and $r = 0.795$; 0.797 and 0.707 for 1 h and 3 h predictions).

1. Introduction

Forecast models exist to predict actual physical parameters (or their proxies) and to predict the rate of changes over a specific time range. They complement other real-time measurements to augment the process of decision-making for forecasters. The level of confidence a forecaster places on predictive tools is critical for a user (e.g., power company and airline industry) to invest their resources where such tools should be updated and augmented through improved understanding as needed, have false alarms minimized, and offer better reliability and redundancy at critical times. The most useful test of models is to assess the success of their actual real-time outputs. This work will test several of the real-time Rice neural network models against the Wing *Kp* models (operated by NOAA/Space Weather Prediction Center (SWPC)). Since both have been running for a good long time, it provides a good opportunity to validate their performance and to perform a head-to-head test over different metrics, e.g., accuracy and model availability.

The Rice artificial neural network (ANN) models (all running at Rice University, Houston, Texas) are empirical space weather prediction models that are capable of giving *Kp*, *Dst*, and *AE* forecasts up to 3 h ahead in near real time [*Bala et al.*, 2009; *Bala and Reiff*, 2012]; the models are also capable of providing forecasts up to 6 h ahead but with slightly less certainty. ANNs are nonlinear models trained using historical data sets on specific input-target patterns that are then capable of generalizing the mapping over a new set of unknown inputs to deliver an output. The Rice ANN models can be classified based on the base functions that they use: the "Boyle model," uses the Boyle Index [*Boyle et al.*, 1997] as the base function, the "Ram model" is similar to the Boyle model but adds a pressure term to the base function, and the "Newell model" shares a similar neural network architecture but uses the Newell functions [*Newell et al.*, 2007, 2008] as its base function. Base functions are solar wind-magnetosphere coupling functions which represent the dayside merging rates in terms of solar wind parameters [e.g., *Wygant et al.*, 1983; *Reiff and Luhmann*, 1986; *Gonzalez et al.*, 1989]. The coupling functions are good representations of the global state of the magnetosphere over a variety of activity (e.g., *Dst*, *AE*, *Kp*, and Auroral Power). All Rice models are neural network-based models yielding 1 h and 3 h ahead predictions.

The Rice ANN Boyle *Kp* forecast models (delivering separate 1 h and 3 h predictions) started out their real-time operations in October 2007, with improvements posted as new solar wind and interplanetary magnetic field (IMF) data became available to train the neural network, and therefore, have evolved over time. The Rice ANN Boyle *Dst* and *AE* models were added in May 2010 and have also been running in real time, while the Ram and Newell models of *Kp*, *Dst*, and *AE* have only been running for a year and a half now (November 2012). Space weather alerts and warnings are delivered in real time based on the models' output. The Wing models [*Wing et al.*, 2005], presently running at SWPC in Boulder, Colorado, since March 2011, are neural network models driven by individual solar wind and IMF parameters giving 1 h and 4 h predictions, with the prediction lead time variable with the solar wind velocity; the Costello *Kp* model [*Costello*, 1997] was run by NOAA/SWPC prior to that. The Wing model is also capable of giving out forecasts up to 4 h ahead, unlike the Costello *Kp* model. The Rice models are available in real time at http://mms.rice.edu/realtime/forecast.html and the Wing model at http://www.swpc.noaa.gov/wingkp/.

The purpose of this work is to test four real-time empirical space weather prediction functions to see which has been the most successful running in real time. We will test their effectiveness in predicting *Kp* and also their "up time," by using their actual predictions posted in real time against the final version *Kp* values. In the real-time plots, we however use the United States Air Force (USAF) real-time estimated *Kp* (http://www.swpc.noaa.gov/rt_plots/kp_3d.html) that have only one-step granularity (taking either a zero or a positive integer value from 1 to 9, whereas the official *Kp* values vary from 0 to 9 in steps of 0.33), and so do the Wing models. We will also validate our models using the true observed data available up until now. This paper will also report on the other Rice forecast models of the *Dst* and *AE* index though the use of different base functions.

2. Model Progression and Test Setup

The Boyle Index has been actively computed in near real time since September 2003, as 10 min averages. It is an actual predictor of the Earth's polar cap potential computed using the solar wind and interplanetary magnetic field parameters. *Bala et al.* [2009] and *Bala and Reiff* [2012] provide more insights to the Boyle Index. We have been running it with a great deal of success, both in terms of serving as a precursor to an impending activity and also for being actively up and running for almost ≈95% of the time. The "downtimes" can be attributed to three factors: (a) ACE instruments going down, perhaps as a result of a space weather event, (b) latency at NOAA/SWPCs end before the operational data are made publicly available, and (c) other issues related to software or hardware glitches and power outages experienced during operations locally. In the last 4 years alone, we have had over 35,000 visitors to our website from all around the world, of which nearly 65% are returning visitors. In the meantime, we have also added more than 1300 individual subscribers to our free e-mail alert system.

On the other hand, the Rice Boyle ANN *Kp* forecasts started out in October 2007 as 1 h and 3 h predictions of the *Kp* index. Since these forecasts are made using time histories of the Boyle Index, so long as we have at least several hours of uninterrupted data, a valid prediction can be made. The Rice Boyle ANN *Dst* and *AE* forecast models were added to the system in May 2010, again many hours of reasonable estimates of the Boyle Index gives a usable forecast. Finally, we introduced a few more 1 h and 3 h models of *Kp*, *Dst*, and *AE* in November 2012, and in this case, the neural networks were trained using two more base functions: the "Ram" functions same as the Boyle but adds a pressure term and the "Newell" functions. We will address the models in greater detail in section 3.

The Rice models (1 h or 3 h) are automated to predict the magnetic indices at 1 h granularity. For example, a forecast made at 0100 UT by 1 h models is valid for the hour between 0100 and 0200 UT, with the actual predicted time dictated by prevalent solar wind conditions. Similarly, for 3 h models the predicted time will be 0300–0400 UT. The Wing *Kp* models, on the other hand, run every 15 min (at 00, 15, 30, and 45 min UT) doing 1 h and 4 h predictions, and the predicted lead time varies with the solar wind velocity. The Wing 1 h model is based on solar wind inputs alone, but the Wing 4 h *Kp* model incorporates the nowcast *Kp* from SWPC in its input stream in addition to solar wind parameters; the Rice models use only solar wind data as inputs.

We have tested the Wing *Kp* models and the Rice Boyle *Kp* models (1 h and 3 h prediction models) to the official *Kp* records over a 22 month period between April 2011 and February 2013; the Wing *Kp* values were obtained through personal communication with NOAA/SWPC, while official *Kp* values were obtained from

NASA CDAWeb (http://cdaweb.gsfc.nasa.gov). Given the real-time Wing *Kp* model, which is being run by NOAA/SWPC since March 2011, and the time of this work, we have about 3 years of data to compare the two models. However, we were able to obtain about 2 years worth of continuous data through personal communications with the concerned personnel. Though there were minor interruptions, with over 15,000 data points to compare, we identified this period to be sufficient to test the models. Notably, this period also happened to be at a time when the Sun was very active, with several resulting geomagnetic activities reaching *Kp* 6 or higher. In the event of an unavailable forecast, all Rice models use the previously recorded value (duplicating the values) as a "placeholder" while the Wing models marks them with "−1." We have chosen different events and apply conventional metrics to test their performances in detail.

3. The Models

All the Rice models generally need up to 9 h (1 h models) or 21 h (3 h models) of look-back time, meaning that they need time histories of solar wind and IMF data. The modeling techniques, the research methodology and the basic neural network architecture representing our models have been addressed in detail in *Bala et al.* [2009] and *Bala and Reiff* [2012]. The currently operational models to forecast *Kp*, *Dst*, and *AE* over 1 h or 3 h range are broken down in terms of their base functions and input time histories are as follows:

1. using the Boyle Index (BI) = $10^{-4}(\frac{v}{\text{km/s}})^2 + 11.7(\frac{B}{\text{nT}})\sin^3(\theta/2)$ kV, where v is the solar wind velocity in km/s, B is the magnitude of the IMF in nanoteslas, and θ is the clock angle:

$$\left.\begin{aligned} Kp_{t+1} &\equiv f(BI_t, BI_{t-1}, \ldots, BI_{t-8});\\ Kp_{t+3} &\equiv f(\overline{BI}_t, \overline{BI}_{t-3}, \ldots, \overline{BI}_{t-21});\\ Dst_{t+1;t+3} &\equiv f(BI_t, BI_{t-1}, \ldots, BI_{t-9});\\ AE_{t+1;t+3} &\equiv f(BI_t, BI_{t-1}, \ldots, BI_{t-9}); \end{aligned}\right\} \tag{1}$$

   where t represents the epoch in question while $t-1$, $t+1$ means 1 hour behind and 1 h ahead of t, respectively. Kp_{t+1} and Kp_{t+3} are the forecasted values, and each BI_t, BI_{t-1}, BI_{t-8}, etc. are hourly averages or 3 h averages of the BI (denoted by $\overline{BI}_{t-3}$ $\overline{BI}_{t-6}$, ...), as the case may be;

2. the Ram functions: using the sample inputs as above plus a dynamic pressure term

$$\left.\begin{aligned} Kp^{\text{Ram}}_{t+1} &\equiv f(BI_t, BI_{t-1}, \ldots, BI_{t-8};\\ &\quad \sqrt[3]{Dp_t}, \sqrt[3]{Dp_{t-1}}, \ldots, \sqrt[3]{Dp_{t-8}})\\ Kp^{\text{Ram}}_{t+3} &\equiv f(\overline{BI}_t, \overline{BI}_{t-3}, \ldots, \overline{BI}_{t-18};\\ &\quad \sqrt[3]{\overline{Dp}_t}, \sqrt[3]{\overline{Dp}_{t-3}}, \ldots, \sqrt[3]{\overline{Dp}_{t-18}})\\ [AE; Dst]^{\text{Ram}}_{t+1;t+3} &\equiv f(BI_t, BI_{t-1}, \ldots, BI_{t-9};\\ &\quad \sqrt[3]{Dp_t}, \sqrt[3]{Dp_{t-1}}, \ldots, \sqrt[3]{Dp_{t-9}}) \end{aligned}\right\} \tag{2}$$

   here *Dynamic pressure* (Dp) : $P_{\text{sw}} = m_p n_p v^2_{\text{sw}}\,(1 + 4n_a/n_p)$, where n_p is the number density of the protons, n_a/n_p is the alpha to proton ratio, and m_p is the proton mass;

3. the Newell functions

$$\left.\begin{aligned} Kp^{\text{Newell}}_{t+1} &\equiv f(C_t, C_{t-1}, \ldots, C_{t-8};\\ &\quad V_t, V_{t-1}, \ldots, V_{t-8})\\ Kp^{\text{Newell}}_{t+3} &\equiv f(\overline{C}_t, \overline{C}_{t-3}, \ldots, \overline{C}_{t-18};\\ &\quad \overline{V}_t, \overline{V}_{t-3}, \ldots, \overline{V}_{t-18})\\ [AE; Dst]^{\text{Newell}}_{t+1;t+3} &\equiv f(C_t, C_{t-1}, \ldots, C_{t-9};\\ &\quad V_t, V_{t-1}, \ldots, V_{t-9}) \end{aligned}\right\} \tag{3}$$

   where C and V denote the coupling ($\mathrm{d}\Phi_{\text{MP}}/\mathrm{d}t = v^{4/3}_{\text{sw}} B^{2/3}_T \sin^{8/3}(\theta/2)$) and viscous components ($n^{1/2}v^2_{\text{sw}}$) respectively for 1 h averages (3 h averages are explicitly denoted by $\overline{C}$ and $\overline{V}$).

Finally, we have the Wing *Kp* models, also known as APL *Kp* models [*Wing et al.*, 2005]. The models currently run by NOAA/SWPC are two of the three APL models: APL 3 model that uses as input only the solar wind parameters *n*, |Vx|, IMF |B|, and B_Z and predicts *Kp* 1 h ahead and the APL model 2 that also includes the nowcast *Kp* in addition to the solar wind parameters and predicts *Kp* 4 h ahead. Though both the Wing and Rice models are fundamentally the same in the machine-learning technique they employ: neural network, they are dissimilar, however, in terms of how the input-output mappings are paired. Wing models use "recurrent network" with at most 20 nodes, while the Rice models use a "standard two-layer feedforward back propagation" architecture with roughly 70 nodes. The Rice models "preprocess" the inputs, i.e., their input data are scaled to fall within the range [−1,1], before feeding them to the neural networks. At the output, the values from the neural network are then scaled back to their original units. Similarly, the Wing models also apply a "postprocessor" before obtaining the final *Kp* values. However, these are intrinsic to their own design implementation and approach than an advantage.

4. Metrics for Observation-Model Comparison

A few common metrics to be used here to compare models are the linear correlation coefficient and RMSE (root-mean-square error). They are given by

$$\text{correlation coefficient } r = \frac{\sum_{t=1}^{N}\left[X_t{}'Y_t{}'\right]}{\left[\sum_{t=1}^{N} X_t{}'^2\right]^{1/2}\left[\sum_{t=1}^{N} Y_t{}'^2\right]^{1/2}} \tag{4}$$

where X_t, Y_t represent the predictions and the actual values, respectively, and $X' = X_t - \overline{X}$ and $Y' = Y_t - \overline{Y}$ represents the deviation from the mean.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} e_t^2}\Bigg\} \tag{5}$$

where e_t = measured$_t$ − observed$_t$ with measured$_t$ is the actual desired value and observed$_t$ is the model output. *N* is the total number of samples.

An autocorrelation function or a temporal autocorrelation computes the correlation of a variable with its past and future values of a time series or waveform. It is given by

$$r_k = \frac{\sum_{t=1}^{N-k}(X_tX_{t+k}) - (N-k)\overline{X}^2}{\sum_{t=1}^{N} X_t{}'^2 - N\overline{X}^2}\Bigg\} \tag{6}$$

where r_k denotes the autocorrelation coefficient and *k* denotes the lagged time step.

We also use the Heidke Skill score (HSS), probability of detection (POD), and the false alarm rate (FAR) for verification purposes. They are defined as follows:

$$\text{HSS} = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}\Bigg\} \tag{7}$$

where *a* is the number of "hits," *c* is the number of "misses," *b* is the number of "false positives," and *d* is the number of "correct rejections" for a given sample, thereby constituting the 2 × 2 contingency table. A perfect forecast receives a HSS score of 1, while a random forecast receives a score of 0.

$$\text{POD} = \frac{a}{a+c}\Bigg\} \tag{8}$$

For a perfect forecast the POD is 1 and 0 for the worst. Similarly, the false alarm rate (FAR) is expressed as

$$\text{FAR} = \frac{b}{a+b}\Bigg\} \tag{9}$$

Perfect forecasts results in a FAR of 0 (no false alarms), and the worst FAR is 1. Additionally, as a reliability measure, we have also chosen to convey the percentage times the models were actively running in real time, referred to as the downtime.

Table 1. Table Showing the Prediction Summary of the Rice Boyle *Kp* and Wing *Kp* Models[a]

Forecast Model	r	RMSE	Downtime	Cross Correlation
Rice Boyle Kp_{t+1}	0.771	0.833	4.6%	"−" Lag
Wing Kp_{t+1}	0.714	0.974	6.0%	"−" Lag
Rice Boyle Kp_{t+3}	0.770	0.836	3.3%	"+" Lag
Wing Kp_{t+4}	0.744	1.049	2.9%	"−" Lag

[a]The linear correlation coefficient is r. The results cover data from April 2011 to February 2013.

5. Head-to-Head Test: Rice and Wing *Kp* Prediction Models

We have conducted a thorough head-to-head test of two space weather models of Rice Boyle *Kp* and Wing *Kp* using outputs posted in real time between April 2011 and February 2013. We summarize the overall test results in Table 1. We eliminated the "null" or invalid points with appropriate placeholders on both the models. Overall, not only are the Rice models better in terms of prediction accuracy ($r = 0.771$ and 0.770 and RMSE = 0.833 and 0.836 versus $r = 0.714$ and 0.744 and RMSE = 0.974 and 1.049 for the 1 h and 3 h models, respectively), but they are also better in terms of the amount of downtime (at the most 4.6% versus Wing's 6.0%). Here, the Wing results are "time matched" with the Boyle *Kp* values as they run every 15 min. Therefore, only those values that are predicted at the "00" UT minute are compared with the Rice *Kp* values. For the Wing's 4 h model, we compared the predicted *Kp* with true *Kp* values at the predicted time that is 4 h ahead and also with true *Kp* values 3 h ahead. We found the latter to correlate better than the former, which is what is reported in Tables 1–3. Here the Wing model has a slight better downtime compared to the Rice model.

However, an undesirable aspect of the Rice models is perhaps that they need longer time inputs and, as a result, a few hours of interruption in the solar wind data can result in bad predictions. We recognize this as a design issue than anything else. Similarly, the 4 h Wing model uses nowcast *Kp* as one of their inputs. Again, any delays in obtaining the nowcast data can slightly hurt their predictions.

In Tables 2 and 3 we present contingency tables (or confusion matrix) for forecast verification; Table 2 displays the 1 h models, while Table 3 displays the 3 h models (unlike Table 1, the invalid points were included here when computing the skill scores). For this purpose, we define three discriminant values for *Kp* (>= 3, 4, and 6) to compute the combinations based on yes's and no's. These results reflect the overall trend seen in Table 1. The Rice 1 h *Kp* models result in significantly higher skill scores (0.934, 0.949, and 0.964) and lower false alarm probabilities (0.10, 0.09, and 0.07) than the corresponding Wing *Kp* models (0.828, 0.829, and 0.764 and 0.248, 0.274, and 0.379). We also observe a similar outcome in the 3 h models (Rice *Kp* with HSS of 0.901, 0.934, and 0.866 and FAR of 0.152, 0.117, and 0.234 versus Wing's HSS of 0.773, 0.683, and 0.528 and FAR of 0.294, 0.441, and 0.636). We also note that in the *Kp* 6+ range, the Rice predictions are far better than the Wing *Kp* models. The zeroes in Tables 2 and 3 mean that the Rice models have not missed a single *Kp* 6+ storm. One can say that carefully guided solar wind-magnetosphere coupling functions, in addition to longer time inputs, in the case of the Rice models, prove to be better than the Wing models' implementation of "raw" solar wind inputs.

Table 2. Contingency Table for Boyle *Kp* 1 h Model[a]

	Observation (*Kp* >= 3[b])		Observation (*Kp* >= 4[c])		Observation (*Kp* >= 6[d])	
Forecast	Yes	No	Yes	No	Yes	No
Yes	2392 (2387)	279 (789)	887 (887)	88 (335)	108 (108)	8 (66)
No	2 (7)	13287 (12777)	1 (1)	14984 (14737)	0 (0)	15844 (15786)

[a]Wing *Kp* 1 h model is shown within parentheses.
[b]HSS = 0.934 (0.828); POD = 0.999 (0.997); FAR = 0.10 (0.248).
[c]HSS = 0.949 (0.829); POD = 0.999 (0.999); FAR = 0.09 (0.274).
[d]HSS = 0.964 (0.764); POD = 1.000 (1.000); FAR = 0.07 (0.379).

Table 3. Contingency Table: Boyle *Kp* 3 h Model (*Kp* >= 4)[a]

	Observation (*Kp* >= 3[b])		Observation (*Kp* >= 4[c])		Observation (*Kp* >= 6[d])	
Forecast	Yes	No	Yes	No	Yes	No
Yes	2,393 (2,289)	430 (952)	888 (850)	118 (673)	108 (107)	33 (187)
No	1 (105)	13,136 (12,614)	0 (38)	14,954 (14,399)	0 (1)	15,819 (15,665)

[a]Wing *Kp* 4 h model is shown within parentheses.
[b]HSS = 0.901 (0.773); POD = 0.999 (0.956); FAR = 0.152 (0.294).
[c]HSS = 0.934 (0.683); POD = 1.000 (0.957); FAR = 0.117 (0.441).
[d]HSS = 0.866 (0.528); POD = 1.000 (0.990); FAR = 0.234 (0.636).

A recent major solar eruption on 29 September 2013 produced a magnificent coronal mass ejection (CME) that glazed past Earth on 2 October 2013. The first impact happened around 00 UT, 2 October 2013 sending the *Kp* index to almost 7 (*Kp* reached 6.67 at 0500 UT) within a matter of hours (Figure 1). During this event, a "Yellow" *Kp* alert based on our prediction went out to subscribers at 02:04:00 UT, 2 October followed by a "Red" alert at 02:24:00 UT, 2 October based on the Boyle Index. The figure shows two panels comparing Rice and Wing 1 h predictions (blue curves) against the USAF *Kp* (red curves) from NOAA/SWPC. The Rice predictions were clearly better than the Wing model (0.877 versus 0.790). In addition, the Rice prediction peaks (green dashed lines) slightly ahead of the Wing model (red dashed line).

6. Validating Rice Neural Network *Kp*, *Dst*, and *AE* Models

The Rice models also predict *Dst* and *AE* in near real time in addition to *Kp*. The real time "harvested" values (the final values that were published in real time based on which any warnings or alert e-mails were sent out) were compared with the final values of *Kp*, *Dst*, and *AE*; these are the values that were posted in real time that are have been used for this analysis. In this case, we had more than 16 months of data to validate the *Kp* and *Dst* models, while we had 13 months to compare the *AE* models with the official *AE* record. We tested 17 models in total, and the overall results are summarized in Table 4, which is similar to Table 1, showing the linear correlation, RMS error, time of coverage, their downtimes, and cross correlations. The results, however, cover a different time range as shown in column 4. It can be seen that all the models are running

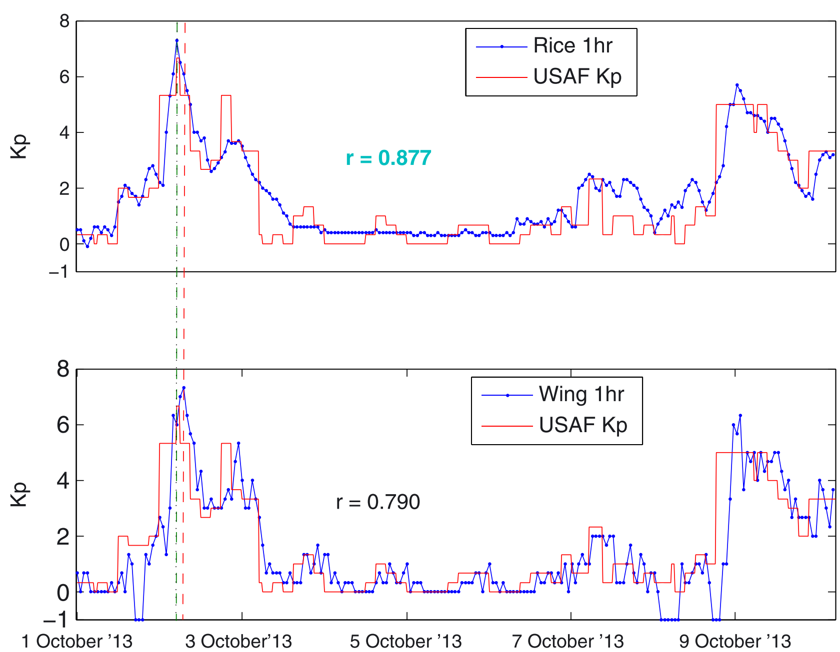

Figure 1. Rice ($r = 0.877$) versus Wing (0.790) predictions during a recent storm. Though we did not get a direct impact, it was large enough to send *Kp* to almost 7. The predictions are compared against USAF ground magnetometer *Kp*. Rice prediction peaks slightly ahead (dashed green line) of the Wing model (dashed red line).

Table 4. Table Showing the Prediction Summary of Rice Models[a]

Forecast Model	*r*	RMSE	Time Covered	Downtime	Cross Correlation
Boyle Kp_{t+1}	0.828	0.749	11/22/2012 to 03/31/2014	< 0.1%	"−" Lag
Ram Kp_{t+1}	0.843	0.655	11/22/2012 to 03/31/2014	<0.1%	"−" Lag
Newell Kp_{t+1}	0.820	0.708	11/22/2012 to 03/31/2014	< 0.1%	"−" Lag
Boyle Kp_{t+3}	0.739	0.881	11/22/2012 to 03/31/2014	< 0.1%	"+" Lag
Ram Kp_{t+3}	0.769	0.788	11/22/2012 to 03/31/2014	< 0.1%	"+" Lag
Newell Kp_{t+3}	0.755	0.810	11/22/2012 to 03/31/2014	< 0.1%	"+" Lag
Boyle Dst_{t+1}	0.790	15.54 nT	11/22/2012 to 03/31/2014	<0.1%	"+" Lag
Ram Dst_{t+1}	0.767	13.87 nT	11/22/2012 to 03/31/2014	<0.1%	"+" Lag
Newell Dst_{t+1}	0.704	15.20 nT	11/22/2012 to 03/31/2014	<0.1%	"+" Lag
Boyle Dst_{t+3}	0.795	15.41 nT	11/22/2012 to 03/31/2014	< 0.1%	"+" Lag
Ram Dst_{t+3}	0.797	12.64 nT	11/22/2012 to 03/31/2014	< 0.1%	"+" Lag
Newell Dst_{t+3}	0.707	14.21 nT	11/22/2012 to 03/31/2014	< 0.1%	"+" Lag
Boyle AE_{t+1}	0.811	98.88 nT	01/01/2013 to 01/31/2014	< 0.1%	"+" Lag
Ram AE_{t+1}	0.806	99.03 nT	01/01/2013 to 01/31/2014	< 0.1%	"+" Lag
Newell AE_{t+1}	0.765	122.37 nT	01/01/2013 to 01/31/2014	< 0.1%	"+" Lag
Boyle AE_{t+3}	0.743	114.77 nT	01/01/2013 to 01/31/2014	< 0.1%	"+" Lag
Ram AE_{t+3}	0.747	113.12 nT	01/01/2013 to 01/31/2014	< 0.1%	"+" Lag

[a]The linear correlation coefficient is *r*.

perfectly with downtimes that are negligibly small, less than 0.1% in all cases. The models are driven using three different driving functions with predictions that are either 1 h or 3 h ahead. The *Kp* values are better predicted using the Ram functions, while *Dst* is modeled better through the use of the Boyle Index. *AE*, on the other hand, is predicted equally well by both the Boyle and Ram functions, where it is statistically too close to differentiate. As can be seen, their cross correlation functions exhibit positive lag except in the case of 1 h *Kp* models which by its very definition is a 3 h average. It should also be noted that the *Dst* values obtained from the models were compared against the final *Dst* values after applying for pressure correction described in *Bala and Reiff* [2012].

The last column (Tables 1 and 4) represents the time correlation (autocorrelation) between the two series: predicted and observed. A positive lag means the predicted series lead the observed series and vice versa. The results show a positive lag for all but the *Kp* 1 h models. The "negative lag" trend seen in our 1 h *Kp* models is because strictly the observed *Kp* is a 3 h average and inevitably they are oversampled over that

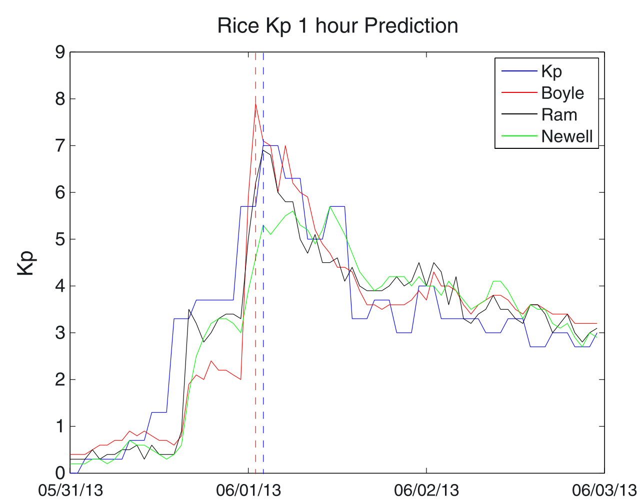

Figure 2. Time series plot of the Rice *Kp* 1 h predictions of an event in June 2013. Here blue is observed, red is the Boyle *Kp*, black is the Ram *Kp*, and green is the Newell *Kp*. Note the Boyle *Kp* peaks slightly ahead of the true values.

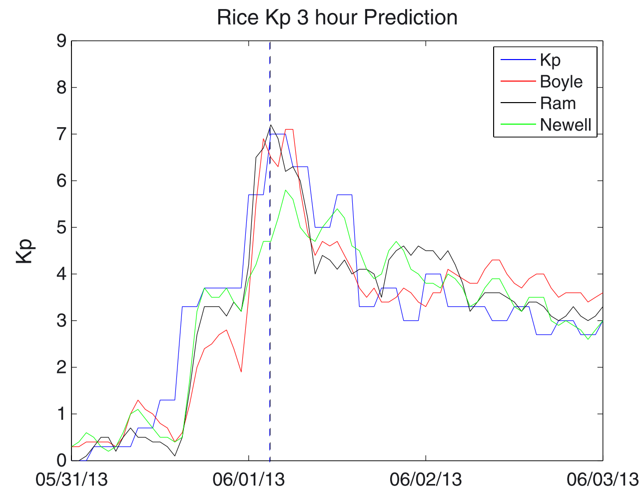

Figure 3. Time series plot of the Rice *Kp* 3 h predictions of an event in June 2013. Observed *Kp* is shown in blue. Note in this case that Ram *Kp* and the true *Kp* peak at the same time.

duration for the purpose of data-model comparison. The Wing 4 h *Kp* model is also subject to debate on the issue of whether it delivers a "true" forecast. Autocorrelation function performed on the Wing *Kp* 4 h model shows a negative lag peaking at the "−5" h. We believe the use of nowcast *Kp* as an input term is a factor that contributes to this deficiency. Whereas the Boyle *Kp* 3 h model, which uses solar wind data alone, does not show this trend, the negative lag at "−3" hours peaks at the predicted time. Persistence-based forecast has been discussed earlier in *Bala et al.* [2009].

A CME hit Earth during the late hours of 31 May 2013, triggering a major geomagnetic activity (measured *Kp* peaked at 7 for 0300–0600 UT, 1 June 2013 and *Dst*, had a minimum value of −119 nT at 0800–0900 UT, 1 June 2013). During this time, a ANN prediction-based *Kp* Yellow alert was triggered at 0104 UT, 1 June 2013. Figures 2–5 show the Rice *Kp* and *Dst* 1 h and 3 h models during this event. The blue curve represents the actual observed response, while the predicted values are plotted using different colors (Boyle in red, Ram in black, and Newell in green). The Boyle function captures the 1 h *Kp* well "peaking" ahead of the rest, whereas the Ram function does better on the 3 h predictions. The *Dst* index, on the other hand, is modeled well by the Ram functions.

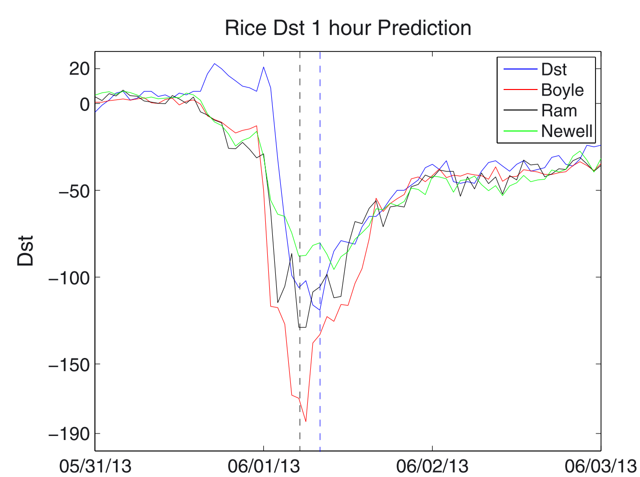

Figure 4. Time series plot of the Rice *Dst* 1 h predictions of an event in June 2013. Observed *Dst* is shown in blue. The predicted Ram *Dst* peaks slightly earlier than the observed *Dst*.

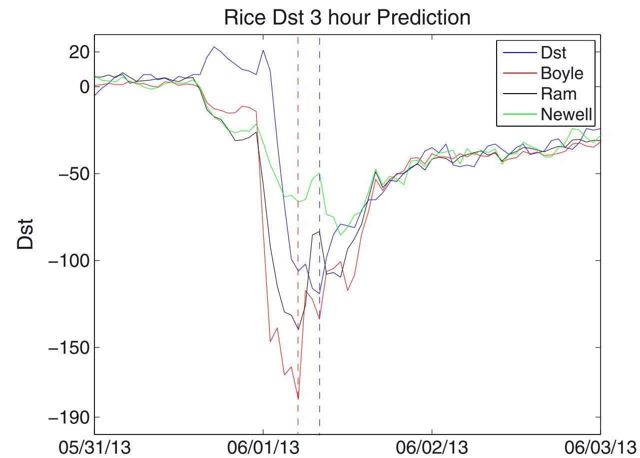

Figure 5. Time series plot of the Rice *Kp* 1 h predictions of an event in June 2013. Observed *Dst* is shown in blue. Note here that the Ram and Boyle *Dst*s peak ahead of the true *Dst* index.

7. Conclusions

Through this study, we are able to validate the forecasting capabilities of the Rice neural network prediction models of *Kp*, *Dst*, and *AE* and the Wing models of *Kp* so that adjustments and improvements can be made in the future. The tests were conducted using actual real-time outputs. Models that have been running in real time, relying on telemetry of semioperational in situ measurements, for some time do provide good insights to the aspects of operational forecasting and specification. We are able to demonstrate that the Rice Boyle *Kp* models clearly challenges and outperforms the Wing *Kp* models on the 22 months of data tested using different verification measures. The Rice models have slightly better "up times" than the Wing models in the 1 h range, but the Wing 4 h prediction model has a better "up" time than the Rice model. Also, the Wing models have higher false alarm probabilities. The Rice models are more accurate and reliable in delivering true forecasts and also have a wider scope, through the use of different basis functions and in their ability to forecast *Dst* and *AE* in addition to the *Kp* index. These models were validated using up to 16 months or real-time harvested outputs. Furthermore, the Rice models are incorporated to send our space weather "alerts" that we have a growing database of subscribers (1300+ individual subscriptions so far) each day.

Acknowledgments
The authors would like to thank the USAF for providing Wing *Kp* data. This study is partly funded by NASA under the Magnetospheric Multiscale (MMS) Mission.

References

Bala, R., and P. H. Reiff (2012), Improvements in short-term forecasting of geomagnetic activity, *Space Weather*, *10*(6), S000779, doi:10.1029/2012SW000779.

Bala, R., P. H. Reiff, and J. E. Landivar (2009), Real-time prediction of magnetospheric activity using the Boyle Index, *Space Weather*, *7*(4), S04003, doi:10.1029/2008SW000407.

Boyle, C. B., P. H. Reiff, and M. R. Hairston (1997), Empirical polar cap potentials, *J. Geophys. Res.*, *102*, 111–125, doi:10.1029/96JA01742.

Costello, K. A. (1997), Moving the Rice MSFM into a real-time forecast mode using solar wind driven forecast models, PhD dissertation, Rice university, Houston, Tex.

Gonzalez, W. D., B. T. Tsurutani, A. L. C. Gonzalez, E. J. Smith, F. Tang, and S.-I. Akasofu (1989), Solar wind-magnetosphere coupling during intense magnetic storms (1978–1979), *J. Geophys. Res.*, *94*(A7), 8835–8851.

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, *J. Geophys. Res.*, *112*, A01206, doi:10.1029/2006JA012015.

Newell, P. T., T. Sotirelis, K. Liou, and F. J. Rich (2008), Pairs of solar wind-magnetosphere coupling functions: Combining a merging term with a viscous term works best, *J. Geophys. Res.*, *A4*, A04218, doi:10.1029/2007JA012825.

Reiff, P. H., and J. G. Luhmann (1986), Solar wind control of the polar-cap voltage, in *Solar Wind-Magnetosphere Coupling*, edited by Y. Kamide and J. A. Slavin, pp. 453–476, D. Reidel Publishing, Tokyo.

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, *J. Geophys. Res.*, *110*, A04203, doi:10.1029/2004JA010500.

Wygant, J. R., R. B. Torbert, and F. S. Mozer (1983), Comparison of S3-3 polar cap potential drops with the interplanetary magnetic field and models of magnetopause reconnection, *J. Geophys. Res.*, *88*, 5727–5735, doi:10.1029/JA088iA07p05727.

