

RICE UNIVERSITY

Forecasting Geomagnetic Activity Indices using
the Boyle Index through Artificial Neural
Networks

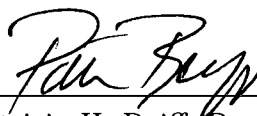
by

Ramkumar Balasubramanian

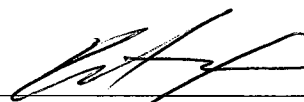
A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

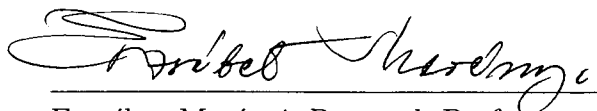
APPROVED, THESIS COMMITTEE:



Patricia H. Reiff, Professor
Physics & Astronomy
Thesis Advisor



David Alexander, Professor
Physics & Astronomy



Erzsébet Merényi, Research Professor
Electrical & Computer Engineering

Houston, Texas

April, 2010

UMI Number: 3421158

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

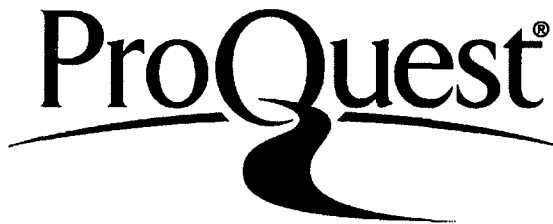
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3421158

Copyright 2010 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Forecasting Geomagnetic Activity Indices using the Boyle Index
through Artificial Neural Networks

by

Ramkumar Balasubramanian

Adverse space weather conditions affect various sectors making both human lives and technologies highly susceptible. This dissertation introduces a new set of algorithms suitable for short term space weather forecasts with an enhanced lead-time and better accuracy in predicting Kp, Dst and the AE index over some leading models. Kp is a 3-hour averaged global geomagnetic activity index good for midlatitude regions. The Dst index, an hourly index calculated using four ground based magnetic field measurements near the equator, measures the energy of the Earth's ring current. The Auroral Electrojet indices or AE indices are hourly indices used to characterize the global geomagnetic activity in the auroral zone. Our algorithms can predict these indices purely from the solar wind data with lead times up to 6 hours.

We have trained and tested an ANN (Artificial Neural Network) over a complete solar cycle to serve this purpose. Over the last couple of decades, ANNs have been successful for temporal prediction problems amongst other advanced non-linear techniques. Our ANN-based algorithms receive near-real-time inputs either from ACE (Advanced Composition Explorer), located at L1, and a handful of ground-based magnetometers or only from ACE.

The Boyle potential, $\Phi = 10^{-4}(\frac{v}{km/sec})^2 + 11.7\frac{B}{nT} \sin^3(\theta/2)$ kV, or the Boyle Index (BI) is an empirically-derived formula that approximates the Earth's polar cap potential and is easily derivable in real time using the solar wind data from ACE. The

logarithms of both 3-hour and 1-hour averages of the Boyle Index correlate well with the subsequent Kp, Dst and AE: $K_p = 8.93 \log_{10}\langle BI \rangle - 12.55$, $Dst = 0.355\langle BI \rangle - 6.48$, and $AE = 5.87\langle BI \rangle - 83.46$. Inputs to our ANN models have greatly benefitted from the BI and its proven record as a forecasting parameter since its initiation in October, 2003.

A preconditioning event tunes the magnetosphere to a specific state before an impending geomagnetic storm. The neural net not only improves the predictions but also helps the prediction by capturing the influence of preconditioning. Two of our models have been running in near-real-time forecast mode already, and the BI and Kp predictions can be obtained from <http://space.rice.edu/ISTP/wind.html>.

Acknowledgments

Dr. Reiff, it was in my eighth grade when I first told my math teacher, with a lot of shyness, that I wanted a doctorate in physics. Here I am. Thank you for all the education, financial support and motivation. Part of my success comes from the freedom you offered and uninhibited approach I was allowed to take during the course of my PhD.

Dr. David Alexander, I thank you for your time, tips and advice. It is truly humbling to have known you.

Dr. Erzsébet Merényi, I thank you for your formal education and insights on artificial neural networks. All the neuro computing I know today is all what you taught me.

Dr. Will Burgett, I thank you for your support and constant encouragement. Rice University would have just been a dream for me without you. Yes, you are right, it is a tremendous institution!

Dr. Chris Johns-Krull, I thank for your financial support and introduction to Astrophysics.

I wish to thank the CISM and the William and Elva Gordon Fund for partially supporting my research.

I thank my parents, sisters, friends and relatives for their constant support and motivation.

Finally, to Srini and Srimathy, I dedicate this thesis to you both.

Contents

Acknowledgments	iii
List of Illustrations	ix
List of Tables	xxii
Symbols and Notations	xxiv
1 Introduction	1
1.1 Geomagnetic Activity Indices	3
1.1.1 The Kp Index	5
1.1.2 The AE index	8
1.1.3 The Dst index	12
1.2 Motivation	14
1.2.1 Solar Wind Coupling Functions	17
1.3 Linear Predictor	20
1.4 Persistence forecasting	23
1.5 ANN and forecasting	24
1.6 Related Work	25
1.6.1 Takahashi Kp Nowcast Model	26
1.6.2 Costello NN Kp Model	27
1.6.3 Boberg NN Kp Model	27
1.6.4 APL Kp Models	28
1.6.5 Wu and Lundstedt NN Dst model	29

1.6.6	Temerin and Li Dst Model	31
1.6.7	Gleisner and Lundstedt NN AE model	31
1.7	Scientific Objective	31
1.8	Thesis Organization	33
2	Artificial Neural Networks	35
2.1	Artificial Neural Networks	35
2.1.1	Biological Neurons	35
2.1.2	Artificial Neurons	36
2.2	A Non-linear Neuron Model	38
2.3	The Backpropagation Algorithm	41
2.3.1	Learning through Gradient-descent Technique	44
2.3.2	Backpropagation algorithm for a two-layer network	49
2.3.3	Notes on Backprop	50
2.4	Conjugate Gradient Algorithm	51
2.4.1	Newton's Method	52
2.4.2	Conjugate Directions	53
2.4.3	The CG Algorithm	54
2.4.4	Levenberg-Marquardt Method	55
3	Research Methodology	58
3.1	Data and Instruments	58
3.1.1	WIND	59
3.1.2	IMP-8	59
3.1.3	ACE	59
3.1.4	Kp	60

3.1.5	Dst and AE	60
3.1.6	Real-time data	61
3.2	Problem Definition	61
3.3	Hypotheses to Test	63
3.4	Proposed Models	63
3.4.1	Training Parameters	64
3.4.2	Model 1: Kp prediction with 1-hour lead time using only the BI	66
3.4.3	Model 2: Kp prediction with 3-hour lead time using only the BI	67
3.4.4	Model 3: Kp Prediction with 1-hour Lead-Time using the BI and Kp history	72
3.4.5	Model 4: Kp Prediction with 2-hour Lead-Time using the BI and Kp history	78
3.4.6	Model 5: 1-hour lead time Dst predictions from BI	82
3.4.7	Model 6: 3-hour lead time Dst predictions from BI	84
3.4.8	Model 7: 1-hour lead time AE predictions from BI	85
3.4.9	Model 8: 3-hour lead time AE predictions from BI	88
3.5	Model Validation	88
3.5.1	Skill Scores	93
3.5.2	Tests of Significance	95
3.5.3	Autocorrelation Function	97
4	Solar Wind-Magnetosphere Coupling	99
4.1	Solar wind-Magnetospheric Interactions	99
4.2	Magnetospheric Convection	102
4.2.1	Open and Closed Models	102
4.3	Role of the Ionosphere	107

4.4	Polar Cap Potential	109
4.5	The Boyle Index: A Solar wind-Magnetosphere Coupling Function . .	113
4.6	Polar Cap Saturation	115
4.7	Newell functions: another coupling formula	117
4.8	Boyle Index: Effect of Preconditioning Events	118
4.9	The BI and space weather	119
5	The Algorithms	122
5.1	Linear Correlations	123
5.2	Cross-correlation Analysis	127
5.3	Prediction Algorithms	130
5.3.1	Model 1: 1-hour lead time Kp predictions from BI	131
5.3.2	Model 2: 3-hour lead time Kp predictions using only the BI .	136
5.3.3	Model 3: 1-hour lead time Kp predictions using the BI and Kp history	140
5.3.4	Model 4: 3-hour lead time Kp predictions using the BI and Kp history	150
5.3.5	Kp models: Understanding accuracy	156
5.3.6	“Persistence” vs. “true” forecasting	162
5.3.7	Post-test Correlation Analysis	163
5.3.8	Kp models compared	166
5.3.9	“Spacalrt” real-time warning system	168
5.3.10	Model 5: 1-hour lead time Dst predictions from BI	170
5.3.11	Model 6: 3-hour lead time Dst predictions from BI	173
5.3.12	Model 7: 1-hour lead time AE predictions from BI	176
5.3.13	Model 8: 3-hour lead time AE predictions from BI	178

5.3.14	Green's function test	183
5.4	Pressure term inclusion	187
5.4.1	Network Training	187
5.4.2	The "Ram" functions	189
5.5	Longer range predictions	195
6	Discussion and Conclusion	199
6.1	Summary	201
6.2	Kp models: Possible applications	203
6.3	Future Possibilities	204
6.3.1	Borovsky Function	205
6.3.2	Improving long term forecasts	206
	References	209

Illustrations

1.1	Kp Network Stations	5
1.2	An illustration of the typical diurnal variation of the horizontal component of the Earth's magnetic field as seen at a Kp station [McPherron, 1997].	6
1.3	AE Network Stations (Figure courtesy: World Data Center C2 for Geomagnetism, Kyoto, Japan)	9
1.4	A severe geomagnetic activity is shown here as an example indicating the values of various solar wind parameters and the geomagnetic indices.	14
1.5	Plot shows the 3-hour averaged log(BI) versus the subsequent 3-hour Kp for 2003 and 2004 with a linear correlation coefficient of 0.785 [Bala et al., 2009].	20
1.6	Linear predictor: Official Kp vs Predicted Kp for April 2001, 2006 & 2007 using 3-hour averages of the BI. Solid line represents the best linear fit ($y = 1.1 x - 0.17$).	21
1.7	1-hour averaged BI versus the Dst (left panel) and AE (right panel). Note the value of Dst is deliberately multiplied by -1 for easy plot comparison.	22
1.8	Linear predictor: Measured Dst vs Predicted Dst for 2006 & 2007 using 1-hour averages of the BI. Solid line is the equality line and the dashed line is the linear fit.	22

1.9 AE Linear predictor: Measured AE vs Predicted AE for 2006 & 2007 using 1-hour averages of the BI. Solid line is the equality line and the dashed line is the linear fit. 23

1.10 Plot showing Predicted Kp vs Official Kp for the APL model 3, $r = 0.84$. [Adapted from Wing et al., 2005] 28

1.11 The performance of the APL models are shown here for comparison. Note that the models 1 and 2 which includes Kp history as inputs appears to lag the real data (black curve) [Adapted from Wing et al., [2005]]. 30

1.12 Shown here is a typical BI plot as seen in real time, reporting the most recent 72 hours of its history. 33

2.1 A biological neuron (dotted oval) is composed of a cell *body*, a multitude of *dendrites*, *synapses* and an *axon*. 36

2.2 A non-linear neuron Model. Σ and φ are defined in equation 2.1. . . . 39

2.3 A sigmoid function $\tanh(x)$ and its derivative (dashed curve). 40

2.4 An example of a two-class classification (black and open circles) : (a) is linearly separable by a simple perceptron, and (b) is not separable by a simple perceptron. 41

2.5 Shown here is a generic architecture of a two-layer *feedforward* network which may be extended to any number of hidden layers. . . . 43

2.6 An example of a learning curve for a function approximation problem ($y = \frac{1}{x}$). Shown here is the RMSE Error for the training (thin curve) and test data (thick curve), recalled at the end of each time step. . . . 46

2.7	Cross-validation: recalling the error on the training and validation data simultaneously as training proceeds. Learning is terminated as soon as the error on the validation set begins to rise.	47
2.8	A comparison of the speed of convergence using different methods. . .	56
2.9	Learning history for the function approximation problem performed using the LM routine. Training RMSE (thin curve) is shown along with the test RMSE (thick curve).	57
3.1	Shown here are the 3-hr Kp and 1-hour Kp smoothed quadratically. .	61
3.2	Each input to the ANN is weighted according to equation 3.1. The most recent value of the solar wind/target index is given the maximum weight of unity.	64
3.3	Diagram showing an input vector and its corresponding output, constituting a pattern. Here $n=6$ is chosen for example.	69
3.4	Model 1 learning histories to predict Kp_{t+1}^* using hourly averages of BI, corresponding to table 3.2 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 0, 1, \dots, 7.$	70
3.5	Model 1 learning histories to predict Kp_{t+1}^* using hourly averages of BI, corresponding to table 3.2 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 8, 9.$	71
3.6	Diagram showing an input vector and its corresponding output, constituting a labeled pattern. Here $n = 6.$	72
3.7	Model 2 learning histories using 3-hour averages of the BI, corresponding to table 3.3 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), n = 0, 1, \dots, 7.$	73

- 3.8 Model 2 learning histories using 3-hour averages of the BI, corresponding to table 3.3 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), n = 8, 9. \dots$ 74
- 3.9 Diagram showing an input vector and its corresponding output, constituting a pattern. Here $n = 6. \dots$ 75
- 3.10 Model 3 learning histories using hourly averages of BI and Kp, corresponding to table 3.4 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-n-3}), n = 1, 2, \dots, 5. \dots$ 76
- 3.11 Model 3 learning histories using hourly averages of BI and Kp, corresponding to table 3.4 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-n-3}), n = 6, 7, 8, 9. \dots$ 78
- 3.12 Diagram showing an input vector and its corresponding output, constituting a pattern. Here $n = 6. \dots$ 79
- 3.13 Model 4 learning histories using 3-hour averages of BI and Kp, corresponding to table 3.5 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-(n+1)*3}), n = 0, 1, \dots, 7. \dots$ 80
- 3.14 Model 4 learning histories using 3-hour averages of BI and Kp, corresponding to table 3.5 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-(n+1)*3}), n = 8, 9. \dots$ 82
- 3.15 Model 5 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $Dst_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 0, 1, \dots, 7. \dots$ 83

3.16 Model 5 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data).
 $Dst_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 8, 9. \dots \dots \dots 85$

3.17 Model 6 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data).
 $Dst_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), n = 0, 1, \dots, 7. \dots \dots \dots 87$

3.18 Model 6 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data).
 $Dst_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), n = 8, 9. \dots \dots \dots 88$

3.19 Model 7 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data).
 $AE_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 0, 1, \dots, 7. \dots \dots \dots 89$

3.20 Model 7 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data).
 $AE_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 8, 9. \dots \dots \dots 91$

3.21 Model 8 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data).
 $AE_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), n = 0, 1, \dots, 7. \dots \dots \dots 92$

3.22 Model 8 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data).
 $AE_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), n = 8, 9. \dots \dots \dots 93$

4.1 Pressure balance between solar wind dynamic pressure and magnetic pressure of the magnetosphere. Plasma pressure is normal to the magnetopause. $\dots \dots \dots 100$

4.2	Equatorial convection pattern in a closed magnetosphere due to viscous interaction between the solar wind (Sun is to the left) and the magnetosphere, as envisioned by Axford and Hines (1961).	103
4.3	Schematic view of the two-cell convection pattern over the polar cap shown in the geomagnetic latitude coordinate system [Kivelson and Russell, 1995].	105
4.4	Dungey's model of the reconnecting magnetosphere for a purely southward IMF [Adapted from Kivelson and Russell, 2007]. N denotes the neutral point ($\mathbf{B} = \mathbf{0}$), formed near the equatorial plane.	106
4.5	A schematic view toward the sun and over the north polar cap showing the region of field-aligned currents. Region 1 and region 2 currents are labeled as R_1 and R_2 respectively [Adapted from Russell, 2007].	108
4.6	The Geocentric Solar Magnetospheric (GSM) coordinate system. X-axis points to the Sun and the dipole is in XZ plane. Y-axis is perpendicular to the Sun-Earth line and point towards the dusk side.	111
4.7	Hourly averages of PCP are plotted against $ \mathbf{B} $ for 2006-2007 using the Hill-Siscoe formulation [Siscoe et al., 2002a]. PCP asymptotes around 160 kV.	112
4.8	A 11-year data (using 1-hour averages) comparing the Boyle index (black diamonds) with the Hill-Siscoe model (open circles) as a function of solar wind electric field. While the former shows a linear rise, the latter asymptotes to a constant value for large electric fields.	116
4.9	K_p predicted using equation 4.6, Newell et al., [2008].	118

4.10	Time series plots of the ACE and a few geomagnetic indices following a CME in April 2006 (3-hour averages). From top to bottom: (a) derived Boyle index, (b) B_z (IMF), (c) official Kp values, (d) the Dst, and (e) the AE index.	120
4.11	Real-time “snapshot” of the BI (1-hour average) and the NOAA Kp during a recent magnetospheric activity on 22 July 2009.	121
5.1	3-hour averaged $\log(\text{BI})$ versus the following 3-hour Kp of a complete data set between 1998 and 2005 ($r=0.74$) [Bala et al., 2009].	123
5.2	Plot shows the 3-hour averaged $\log(\text{BI})$ versus the following Kp for 2003 and 2004 with a linear correlation coefficient of 0.785. Note that the chosen BI cut-off in this case is 100kV while the trigger level is reduced to $K_p = 4$ [Bala et al., 2009].	124
5.3	1-hour averaged $\log(\text{BI})$ 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 [Bala et al., 2009].	125
5.4	3-hour averages of the BI is compared against the Dst index for 2006-07.	126
5.5	1-hour averages of the BI is compared against the Dst index for 2006-07.	126
5.6	1-hour (left) and 3-hour (right) averages of the BI is compared against the AE index for 2006-07.	127
5.7	Cross-correlation function of $\log(\text{BI})$ and Kp (oversampled) versus the time lag. Both 3-hour averages (solid curve) and 1-hour averages (dash-dot) display a strong correlation at 3 hour with the BI leading Kp [Bala et al., 2009].	128
5.8	Cross-correlation function of BI and Dst: 1-hour averages (solid curve) and 3-hour averages (dashed).	129

5.9	Cross-correlation function of BI and AE: 1-hour averages (solid curve) and 3-hour averages (dashed).	129
5.10	Model 1: ANN Kp vs measured Kp, $r = 0.865$. The dashed lines are the prediction bounds with 95% confidence limit and the thick line is the linear fit.	133
5.11	Kp distribution for model 1 using the best function, $n=9$	134
5.12	Model 1 test results : Measured (black) and predicted (white) 1-hour Kp distribution with each number representing the number of inputs.	135
5.13	Model 1 test results contd.: Measured (black) and predicted (white) 1-hour Kp distribution with each number representing the number of inputs.	137
5.14	Results of our model 1 predictions from a recent activity.	137
5.15	Model 2: 3-hr ANN Kp vs official Kp, $r = 0.819$. The dashed lines are the prediction bounds with 95% confidence limit and the thick line is the linear fit. The model uses 21 hours of solar wind history to make a prediction.	139
5.16	Kp distribution for model 2 using the best function, $n=6$	140
5.17	Model 2 test results: Measured (black) and predicted (white) 3-hour Kp distribution with each number representing the number of inputs. Note: since these are 3-hour averages, the distribution has been multiplied by a factor of 3 for easy plot comparison.	141
5.18	Model 2 test results: Measured (black) and predicted (white) 3-hour Kp distribution with each number representing the number of inputs. Note: since these are 3-hour averages, the distribution is multiplied by a factor of 3 for easy plot comparison.	142

5.19	Results of our model 2, 3-hour ahead predictions, from a recent activity. The bottom panels compare the NOAA Kp and the official Kp with the model 2 predictions (white histograms).	142
5.20	Model 3: 1-hour ahead ANN predicted Kp vs measured Kp. Linear correlation, $r = 0.960$. The network uses 6 hours of input history of each BI and Kp. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	144
5.21	Kp distribution for model 3 using the best function, $n=5$	145
5.22	Model 3 test results: Measured (black) and predicted (white) 1-hour Kp distribution. Each number here denotes the number of inputs each of BI and Kp.	148
5.23	Model 3 test results cont.: Measured (black) and predicted (white) 1-hour Kp distribution. Each number here denotes the number of inputs each of BI and Kp.	149
5.24	Model 4: 3-hour ahead ANN predicted Kp vs measured Kp. Linear correlation, $r = 0.839$. The ANN uses 27 hours of input history of each BI and Kp. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	151
5.25	Kp distribution for model 3 using the best function, $n=8$	151
5.26	Model 4 test results: Measured and predicted 3-hour Kp distribution with each number here denotes the number of inputs each of BI and Kp. Note: the distribution is multiplied by 3 for easy of plot comparisons.	154
5.27	Model 4 test results contd.: Measured and predicted 3-hour Kp distribution. Note: the distribution is multiplied by 3 for the sake of plot comparisons.	155

5.28	Summary of the best results from the 4 Kp models. Model 1: $r=0.865$, Model 2: $r=0.819$, Model 3: $r=0.960$, Model 4: $r=0.839$. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	155
5.29	Model vs. data comparison: Predictions from all the Kp models are overplotted with the measured Kp. Data shown here covers all of April 2001.	159
5.30	Model vs. data comparison: Predictions from all the Kp models are overplotted with the measured Kp. Data shown here covers all of December 2006. See appendix (page 208) for a closer view of the data within the dotted lines.	160
5.31	The “bulls-eye” plot: Shown here are the relative errors, measured for every measured $Kp > 6$. The results from the models are shown in separate sectors.	161
5.32	Auto-correlation for 3-hour BI (left panel) and 3-hour Kp (right panel) and cross-correlation for the self predicting BI and Kp models are shown here. Note that both the models (circles) lag the auto-correlation curves (smooth curve).	162
5.33	Auto-correlation for 3-hour Kp and cross-correlation for the official Kp vs ANN predicted 3-hour Kp is shown here. Note that the model which includes Kp history as an input (Model 4) appears to lag the real data.	165
5.34	Plot to demonstrate the effect of persistence visible in the models using Kp inputs. Prediction “peaks” are clearly shifted relative to the measured values in the bottom panel (model 3 which uses Kp as an input).	167

5.35	Head-to-head performance of Model 1 and Costello ANN Kp: (a) Model 1 predicted Kp versus measured Kp ($r = 0.84$), and (b) Costello ANN predicted Kp versus measured Kp ($r = 0.81$) [Bala et al., 2009].	169
5.36	Distribution of minimum Dst value due to the passage of various CMEs or related sheath regions versus the time (Adopted from Cane and Richardson, [2003]).	171
5.37	Model 5: ANN predicted Dst vs measured Dst, $r = 0.802$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	171
5.38	Model 6: ANN predicted Dst vs measured Dst, $r = 0.853$. This model gives a lead time of 3 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	174
5.39	Predictions from models 5 and 6 are overplotted with the measured Dst. Data shown here covers all of December 2006.	174
5.40	Predictions from models 5 and 6 are overplotted with the measured Dst. Data shown here covers all of April 2001.	176
5.41	Model 7: ANN predicted AE vs measured AE, $r = 0.810$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	178
5.42	Model 8: ANN predicted AE vs measured AE, $r = 0.74$. This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	180
5.43	Predictions from models 7 and 8 are overplotted with the measured AE. Data shown here covers all of April 2001.	180

5.44	Predictions from models 7 and 8 are overplotted with the measured AE. Data shown here covers all of December 2006.	182
5.45	A Green's function test reveals the network responding instantly, and decaying slowly thereafter, to the short solar wind impulse.	184
5.46	An example: an input-output training pattern for the 1-hour Kp model that includes the pressure term. The same idea may be extended for different inputs and time cadences and, for the Dst and AE models.	188
5.47	BI and Dp included: ANN predicted Kp vs measured Kp, $r = 0.884$ (left panel). This model gives a lead time of 1 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit. Kp histogram is shown to the right.	191
5.48	BI and Dp included: ANN predicted Kp vs measured Kp, $r = 0.841$ (left panel). This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit. Kp histogram is shown to the right.	191
5.49	BI and Dp included: ANN predicted AE vs measured AE, $r = 0.810$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	192
5.50	BI and Dp included: ANN predicted AE vs measured AE, $r = 0.744$. This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	192
5.51	BI and Dp included: ANN predicted Dst vs measured Dst, $r = 0.814$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	193

5.52	BI and Dp included: ANN predicted Dst vs measured Dst, $r = 0.811$. This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.	193
5.53	An example: an input-output training pattern for the 6-hour ahead Kp, AE and Dst models that includes the pressure term.	195
5.54	Predictions from 6-hour models are overplotted with the measured values. Data shown here covers all of December 2006.	197
6.1	Scatter plots of the 3-hour averages of the BI and the CF versus the Kp is shown here.	206
6.2	Supplement plot to figure 5.30 (page 160). Model vs. data comparison: Predictions from all the Kp models are overplotted with the measured Kp.	208

Tables

3.1	Summary of Data	62
3.2	Predict Kp_{t+1}^* from BI.	68
3.3	Predict Kp_{t+3}^* from BI	71
3.4	Predict Kp_{t+1}^* from BI and Kp history	77
3.5	Predict Kp_{t+3}^* from BI and Kp	81
3.6	Predict Dst_{t+1}^* from BI	84
3.7	Predict Dst_{t+3}^* from BI	86
3.8	Predict AE_{t+1}^* from BI	90
3.9	Predict AE_{t+3}^* from BI	91
3.10	Contingency Table	94
5.1	Model 1 summary of results: Kp_{t+1}^* from BI	136
5.2	Model 2 summary of results: Predict Kp_{t+3}^* from BI	143
5.3	Model 3 summary of results: Predict Kp_{t+1}^* from BI and Kp	147
5.4	Model 4 summary of results: Predict Kp_{t+3}^* from BI and Kp	152
5.5	Geomagnetic Storm Alerts ($Kp > 6$).	157
5.6	Geomagnetic Storm Alerts ($Kp > 6$).	158
5.7	Predict Dst_{t+1}^* from solar wind	172
5.8	Predict Dst_{t+3}^* from solar wind	175
5.9	Geomagnetic Storm Alerts (Dst threshold: < -40 nT).	177

5.10	Predict AE_{t+1}^* from solar wind	179
5.11	Predict AE_{t+3}^* from solar wind	181
5.12	Geomagnetic Storm Alerts (AE threshold: > 500 nT).	182
5.13	Table showing the prediction summary. Thresholds: $K_p > 4$; Dst $<$ -40 nT; AE > 500 nT.	186
5.14	Table showing the prediction summary of the models including dynamic pressure term. The numbers inside the parenthesis are the original BI only models. Thresholds: $K_p > 4$; Dst < -40 nT; AE $>$ 500 nT.	194
5.15	Summary of models with 6 hour lead time from BI and Dp.	196
5.16	Geomagnetic Storm Alerts (Thresholds: $K_p > 6$; Dst < -40 nT; AE > 500 nT).	198

Symbols and Notations

k	Boltzmann's constant	$1.38 \times 10^{-23} \text{ J K}^{-1}$
μ_0	permeability of free space	$4\pi \times 10^{-7} \text{ H m}^{-1}$
T	temperature	$^{\circ}\text{K}$
$ B $	IMF magnitude	
B_{MS}	geomagnetic field	
H	Horizontal component of Earth's magnetic field	
P_{sw}	solar wind dynamic pressure	
ρ_{sw}	solar wind density	
v_{sw}	solar wind velocity	km sec^{-1}
m_p	mass of Proton	
B_Z	the southward component of the IMF	
θ	IMF clock angle	
R_E	equatorial radius of the Earth	$6.38 \times 10^6 \text{ m}$
UTC	Coordinated Universal Time	

Chapter 1

Introduction

“Space weather” forecasting commonly refers to efforts to monitor the prevailing conditions at the Sun, the solar wind, energetic particles of solar-origin, and to predict changes in fields and particles in the Earth’s magnetosphere and ionosphere. Researchers using ground and space-based imagers monitor the Sun for active solar structures that are likely to erupt with a solar flare and/or Coronal Mass Ejection (CME); predicting the time of occurrence of a CME before it erupts is extremely difficult today. These eruptions can now be observed by the Extreme-ultraviolet Imaging Telescope (EIT) onboard SOHO, and reconstructed in 3D by imagers aboard the STEREO spacecraft. SOHO orbits around the Lagrangian point L1 which is locked to the Sun-Earth line to provide uninterrupted observations of the Sun. These initial signatures are crucial for space weather forecasters as they can provide input into magnetohydrodynamic (MHD) models of the solar wind, which allows an approximate prediction of the timing and intensity of the CME as it approaches Earth. Such a Sun-to-Earth model is the principal output of the Center for Integrated Space Weather Modeling (CISM) [Baker et al., 2004], and the Michigan Center for Space Environment Modeling (CSEM) [<http://csem.engin.umich.edu/>].

However, these state-of-the-art models do not yet run routinely in *real time*. Furthermore, one key unknown even in these models is the magnetic field polarity and strength at the leading edge of the CME as it nears Earth. The strength, and even more importantly, the direction of the Interplanetary Magnetic Field (IMF) are cru-

cial parameters for the interaction between the CME and the Earth's magnetosphere. If the field at the leading edge of the CME (or magnetic cloud) is large and southward, a strong electrical connection between the Earth and the Sun is effected, triggering magnetospheric storms and accompanying auroral substorms. On the other hand, if the leading edge has a northward field, the electrical connection and associated geomagnetic activity is significantly less, although plasma can still be loaded on the dayside [Oieroset et al., 2008], leading to a more energetic substorm later when the southward field at the trailing edge of the cloud hits the magnetosphere. Thus, there is a critical need for an accurate upstream solar wind monitor that can measure the speed of the solar wind and its magnetic field direction. Although the strongest effect is within the first hour of the impact of the solar wind with the magnetosphere, the effect of mass loading during periods of northward IMF and the effects of previous injections of plasma into the ring current also make the time history of the interaction important, and introduce crucial non-linearities into the system. At present, the ACE (Advanced Composition Explorer) spacecraft is the farthest upstream monitor, lying 1.5 million km from the Earth on the Sun-Earth line, that reliably provides the critical data on the in-situ solar wind and IMF conditions.

Acute space weather and geomagnetic conditions can impact the precise operation of various civil- or defense-related communication and navigation systems, power grids, and the health and safety of astronauts venturing out in space, thereby exposing them to radiation hazards. In general, just as in the case of regional terrestrial weather and climate forecasting, space weather forecasting is a coordinated attempt to acquire data by placing instruments both on the ground and in space and combine them with science-based empirical, semi-empirical or numerical models to generate accurate and uninterrupted forecasts of the highly variable conditions in the near-Earth space environment; e.g., solar X-ray flux, auroral activity. Today, space weather fore-

casters rely on a variety of forecasting tools ranging from simple non-linear regression techniques to empirical, semi-empirical and physics-based models which are primarily physical approximations of the system at the electrodynamic, MHD or kinetic level [Vassiliadis, 2007]; forecasters themselves come from various government, academic or private sectors. A major milestone in today's space weather forecasting is the technological capability to have data acquisition systems at strategic points in near-Earth surroundings that provide data in real time for an instant evaluation of the geospace environment. Unfortunately, such technological capabilities are only as good as their tolerance to potential radiation hazards in space; technical glitches are also not uncommon, if not widespread, causing operational delays. While accurate long-term end-to-end forecast models of the solar-terrestrial system are being developed by major research facilities such as the CISM, it becomes more of a necessity than a matter of interest within the forecasting framework to have short-term predictions of specific parameters to satisfy the needs of various end-users (e.g., satellite, electric power grid operators, and manned space flight missions). This provides the focus and motivation for the work reported in this thesis.

1.1 Geomagnetic Activity Indices

An index can be defined as a global number aimed at representing the amplitude of a physical parameter [Menvielle and Berthelier, 1991] or as a local number representing the state an isolated system (e.g., medical indices: temperature, blood pressure; money market indices: Dow Jones). Geomagnetic activities are succinctly characterized by a variety of magnetic indices. Geomagnetic activity indices provide simple yet physical characterizations of global geomagnetic activity. They can also be used as input parameters to various models. The Earth's magnetic field is usually de-

defined by seven parameters: the declination (D), the inclination (I), the horizontal intensity (H), the vertical intensity (Z), the total intensity (F) and the north (X) and east (Y) components of the horizontal intensity. The geomagnetic field near the surface of the Earth is known to vary under the influence of the solar wind, and its variation, when measured, offers an estimate of the total solar wind energy input into the magnetosphere. The observed variation in the geomagnetic field comes from two principal sources, namely, secular variations, which are due to internal sources within the Earth, and transient variations such as those generated by atmospheric processes, magnetospheric storms and substorms under the influence of solar wind; secular variations do not contribute as much.

With several decades of ground-based observations of magnetic perturbations available, it is possible to delineate the different morphological signatures observed as being due to regular or irregular variations [e.g., Mayaud, 1978]. Regular variations are mainly related to ionospheric current systems and to the atmospheric dynamo processes [e.g., Kamide, 1988] while irregular variations are related typically to, solar cycle influence on the magnetosphere causing reconnection events and to convection in the polar caps. Irregular variations, on the other hand, produce magnetic perturbations whose morphology and duration varies with latitude [e.g., Berthelier, 1979]. For instance, a specific auroral activity is caused by variations in auroral currents and associated field-aligned currents, triggering substorms. Therefore, different geomagnetic indices are chosen to represent the level of geomagnetic activity at different latitudes owing to contributions from both regular and irregular variations. The most commonly used geomagnetic activity indices are the Kp index, Dst index and the AE index. The Kp index is a 3-hour index derived from magnetometers, located in sub-auroral latitudes, measuring variations in the H-component traces of the geomagnetic field. The Dst index, measured at equatorial latitudes, depends on the intensity of

the ring current, which is a large scale current due to charged particles trapped inside the magnetosphere, while the AE indices, calculated at auroral latitudes, give information about the intensity of auroral currents (or the auroral electrojet).

1.1.1 The Kp Index

Kp is a 3-hour geomagnetic index used to characterize activity at auroral and subauroral latitudes, although mostly relevant to subauroral latitudes. It is derived from the 3-hour K index, originally designed by Bartels et al. [1939] to represent a “planetary” index. The K index (0, 1, 2 ..., 9) is site-specific and thus cannot be regarded as a true global characterization of the observed perturbations at those latitudes. The Kp index, however, is a quasi-logarithmic scale calculated using the weighted averages of 13 ground-based magnetometers situated between 48° and 60° magnetic latitude, making it a standardized “global” index; the network comprises observatories located mostly in Western Europe and North America along with a couple of observatories located in the southern hemisphere, and covers a wide longitudinal range (figure 1.1). The duty of each observatory is to register the maximum variation in the amplitude



Figure 1.1 : Kp Network Stations

of the horizontal component of the Earth's magnetic field (given in nT) over a 3-hour interval (figure 1.2), and because of the latitude dependence of the magnetic activity level, each observatory or station uses its own look-up table where a locally measured range can be scaled to a corrected geomagnetic latitude. In its final form, Kp ranges from 0 (signifying a very quiet period) to 9 (severe activity) in 28 quantized levels and therefore take values from 0, 0+, 1-, 1, 1+, 2-, ... 9. The official keeper of the Kp data is GeoForschungs Zentrum, Potsdam, Germany.

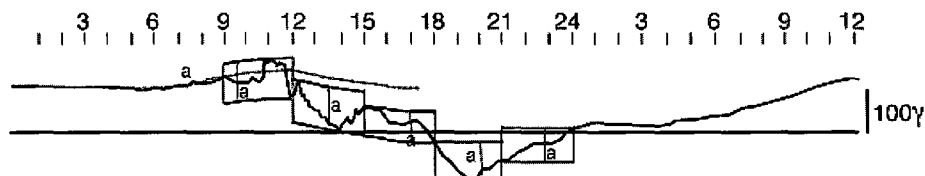


Figure 1.2 : An illustration of the typical diurnal variation of the horizontal component of the Earth's magnetic field as seen at a Kp station [McPherron, 1997].

Kp Proxies

In spite of its crucial role in space weather for being a convenient measure of the geomagnetic activity and as a parameter used in the magnetospheric and ionospheric models [e.g., Carbary, 2004], the Kp index also has several disadvantages. 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 real time applications, such as space weather forecasts; however, it should be noted that some near-real-time proxies are now being made available.

Recently, a large number of researchers have tried to address the issues involving Kp and to bridge the gap so it can be made to serve in real time. Kp proxies are now becoming common and a particularly useful one is the estimated 3-

hour Kp provided by the NOAA (National Oceanic and Atmospheric Administration)/SWPC (Space Weather Prediction Center). It is available through http://www.swpc.noaa.gov/rt_plots/kp_3d.html where the estimated Kp index is derived at the U.S. Air Force 55th Space Weather Squadron using data obtained through tele-links from ground-based magnetometers across the US and Canada (Meanook, Canada; Sitka, Alaska; Glenlea, Canada; Saint Johns, Canada; Ottawa, Canada; Newport, Washington; Fredericksburg, Virginia; Boulder, Colorado; and Fresno, California) along with one European station (Hartland, UK). These are made available through the cooperation of the Geological Survey of Canada (GSC) and the US Geological Survey. They report their final real-time and best estimates of Kp at the end of the prescribed 3-hour interval (0000-0300, 0300-0600,....., 2100-2400). While this provides a significant improvement for space weather forecasting over the monthly official releases, the limitation of this proxy is the lack of homogeneity in the distribution of the observatories which can result in an underestimation of a magnetic storm intensity or in the worse case, may even fail to record a storm. Moreover, these are low-resolution Kp, approximated to the nearest integer values (0, 1, ..., 9) and are therefore, not true depictions of the official Kp.

Kp: A Magnetospheric Activity Index

The Kp index, taken at midlatitudes, responds to both equatorial and auroral disturbances and thus is considered a true global index. This, coupled with the fact that it has the longest history of recorded measurement (continuously since 1932), means that it is one of the most commonly used in geophysical research as a measure of various magnetospheric properties [Menvielle and Berthlier, 1991]; it is also used for the study of the solar wind-magnetosphere interaction [Papitashvili et al., 2000]. One of the earliest studies on the auroral oval demonstrated the dependence

of the equator-ward auroral boundary, λ_m , on Kp: $\lambda_m = 65.2 - 1.04Kp$ [Feldstein and Starkov, 1967]. Among other earlier works, both ground-based whistler observations [e.g., Carpenter, 1967] and space-based measurements [e.g., Binsack, 1967; Gringauz, 1969] have shown a Kp dependence of the size of the plasmasphere. Another magnetospheric phenomenon that is strongly dependent on the Kp index is the Earth's cross polar cap potential drop [e.g., Heppner, 1973; Reiff et al., 1981]. Kp is also viewed as a monitor of the strength of magnetospheric convection (discussed in chapter 4), given its dependence on the latitude of the auroral current [Thomsen, 2004]. All of these studies indicate the importance of the Kp index to the characterization of the state of the geomagnetic environment.

Due to its uniformity and accessibility, Kp is often used as an input parameter to other magnetospheric and ionospheric models. For example, the Costello Kp model [Costello, 1997] developed at Rice University to provide critical input to the Rice Magnetospheric Specification and Forecast Model (MSFM). Another application to the Kp index is the radiation belt model [Fok et al., 2001], which is a comprehensive computational model of the Earth's ring current. Recently, the OVATION (Oval Variation, Assessment, Tracking, Intensity, and Online Nowcasting) model developed by Newell et al. [2002] has been adapted to use Kp as an input parameter for determining the equatorial boundary of the auroral oval.

1.1.2 The AE index

The Auroral Electrojet indices or “AE indices” are used to characterize the global electrojet activity in the auroral zone and are derived from 1-min averages of northward H component traces of a series of ground-based magnetometers. However, similar to Kp, one of the deficiencies in reporting the AE index precisely is due to the lack of uniformity in the location of these observatories along the auroral zone. For example,

universal time and season [Li et al., 2007]. The AE index varies over a free range with minimum at 0 nT. The 12 observatories that contribute to AE are located between 62° and 77° N latitudes (figure 1.3).

AE and magnetospheric activity

There are two main components contributing to the AE index: one that is directly related to the solar wind energy input into the magnetosphere and the other that arises due to reconfiguration of the tail magnetic field, causing internal magnetospheric instability. Geomagnetic storms are temporary magnetic disturbances inside the Earth's magnetosphere (see chapter 4) caused by the Sun. Magnetospheric substorms, another basic class of geomagnetic activity along with the geomagnetic storm, occur over intervals of hours and are common during storms [Russell, 2000]. The AE index responds well to magnetospheric substorms and accompanying auroral displays in the high-latitude ionosphere, because of its sensitivity to the auroral electrojets. AE measures the global activity level of the auroral zone by finding the maximum perturbation around the oval, and it is particularly valuable as an indicator of magnetic substorms [Hargreaves, 1992]. Moreover, the AE index has significant contributions from the AL index because it is driven both directly and indirectly from substorm expansion phase activity [Li et al., 2007]. Although the exact triggering mechanism of the substorm is still unknown, it is generally believed that global substorm onset signatures can be identified using sudden and persistent decreases in the AL index which are characterized by intensification of auroral electric currents.

It is also true, however, that AE indices “alone” cannot provide accurate descriptions of substorm onset, but perhaps with the aid of Pi2 pulsations, which are fluctuations in the geomagnetic field occurring with periods 40-150 seconds and in succession [Saito, 1969], detecting the onset accuracy can be increased [Hsu and

McPherron, 2003]. In another study, Lyons et al [1997] have also suggested the distance of solar wind monitors (e.g., satellites like WIND, IMP-8) perpendicular to the Sun-Earth line as another important parameter in determining whether a possible trigger is detected. They found that spatial structures in the plane perpendicular to the Sun-Earth line critically affects whether or not a trigger is observed from a particular satellite. Though Hsu and McPherron [2003] have argued that the probability for triggered substorms and nontriggered substorms is about the same regardless of the distance to the Sun-Earth line.

AE Availability

The AE index is available in three different forms at present: the quick-look or near-real-time, the provisional, and the final; they are made available at the World Data Center C2 for Geomagnetism, Kyoto, Japan webpage (<http://wdc.kugi.kyoto-u.ac.jp/aedir/index.html>). The first form, or quick-look AE index, is made available in quasi-real time mostly for non-commercial purposes of monitoring and forecasting i.e., it is derived based on the number of stations that are currently reporting at a given point of time, and as more information becomes available for use they are updated periodically. At present, the maximum number of AE observatories reporting in real time is eight. Therefore, due to lack of sufficient longitudinal coverage, the data might lead to an underestimate. Therefore, the quick-look AE index is neither a reliable real-time monitor nor good for scientific analyses. The second form is the provisional index that is published after a period of few months [Source: NOAA]. The problem of provisional AE index is difficulty in obtaining data from stations especially the Russian observatories. The last and the final form is the final AE index updated from the provisional form in about 6 to 12 months, suitable for scientific research.

1.1.3 The Dst index

The ring current and the Van Allen radiation belt consists of high-energy particles trapped inside the Earth's magnetosphere, where the peak energy of the particles in the ring current can reach up to 200 keV while the energy of the particles in the radiation belt can reach relativistic energies [Moldwin, 2008]. The shape and strength of the Earth's dipole field cause these trapped particles to flow eastward (electrons) or westward (ions) depending upon their charge, thereby constituting the ring current. The ring current in turn generates a field that opposes the magnetic field of the Earth which is measurable using instruments at the surface of the Earth. Therefore, the strength of this net magnetic field provides a good measure of the energy content of these circulating particles and the resulting toroidal currents.

The disturbance storm-time index or the Dst index, calculated using four ground based magnetic field measurements near the equator, measures the energy in the ring current as the average depression of the horizontal component of the magnetic field around the Earth at low latitudes [Russell, 2000]; the Dst index is greatly enhanced (more negative) during geomagnetic storms. These electric currents flow about 5.6 Earth radii above the equatorial region. A storm sudden commencement (ssc) is seen as a sharp positive peak in Dst before the main phase of the global storm produces large negative values. However, the conditions in the solar wind and the magnetosphere during a storm are such that the other current systems, such as the magnetopause currents and tail currents, influence the strength of the measured field, and it is therefore customary to apply corrections to the Dst by removing their contribution. Once the corrections have been applied, the Dst index provides a good estimate of the overall energy in the ring current [Dessler, 1967].

Dst Availability

Presently, the Dst index is released every hour by the World Data Center C2 for Geomagnetism, Kyoto, Japan, which is available from their webpage in near real time, and as provisional and final values; the final values are released after about a year and a half because of the delay in obtaining data from all observatories. Each hourly value of the index is the average symmetric disturbance amplitude, projected onto the equator, of the horizontal component recorded at 4 mid-latitude stations. Values are given in nT and are near 0 during quiet times [Geomagnetic Indices Bulletin, January 2009]. Large negative values signify highly disturbed magnetospheric states.

Dst Corrections

During times when the solar wind dynamic pressure is high, the magnetopause moves closer to the Earth causing the currents associated with it to contaminate the Dst. Burton et al. [1975] gave a formal algorithm to estimate the Dst correction, which is written as:

$$Dst^{corrected} = Dst - b\sqrt{P_{sw}} + c, \quad (1.1)$$

called the BMR equation, where b is the measure of the response to dynamic pressure (P_{sw}) changes in the solar wind and c is the measure of quiet day currents. Through empirical analysis, they provided a pressure correction term b of $16 \text{ nT}(\text{nPa})^{-1/2}$ and c of 20 nT . While various values for b and c were given by different authors over the years, disagreements tend to exist over their precise values. Recently, Brien and McPherron [2000] performed an empirical phase space analysis of ring current dynamics to obtain values of $b = 7.26$ and $c = 11$, which I will use throughout this dissertation; their results apply for Dst greater than -150 nT .

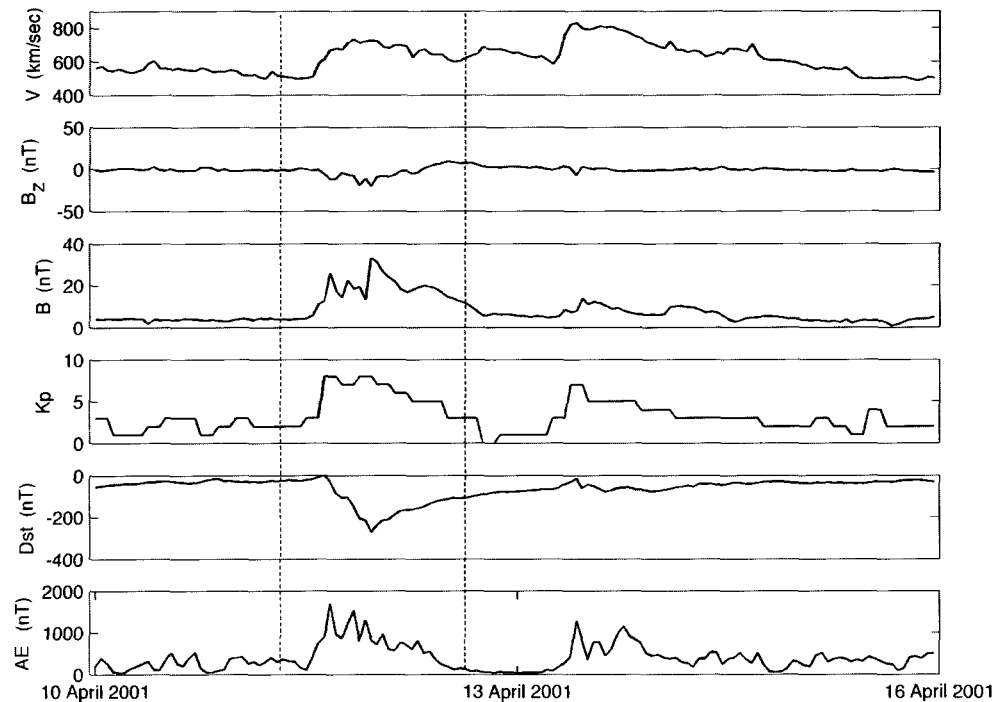


Figure 1.4 : A severe geomagnetic activity is shown here as an example indicating the values of various solar wind parameters and the geomagnetic indices.

1.2 Motivation

In the United States, the SWPC of the National Weather Service (NWS), an entity of NOAA, is the “official source” for space weather alerts and warnings, and is one of the nine National Centers for Environmental Prediction. It provides real-time monitoring and forecasting of solar and geophysical events (figure 1.4). It also explores and evaluates new models and products and transitions them into operations. SWPC is also the primary warning center for the International Space Environment Service and works with many national and international partners with whom data, products, and services are shared [Source: <http://www.swpc.noaa.gov>]. However, given the copi-

ous data available to SWPC for space weather forecasting, the data processing and subsequent scientific analysis can be tackled better through local and international alliances; for instance, sometimes, the SWPC's Kp nowcast model is prone to processing delays. A possible approach to space weather forecasting should include issuing short-term forecasts through event-driven (e.g., solar flares, geomagnetic storms, proton events, electron events) alerts and warnings, offering daily forecasts of solar and geomagnetic activity.

The launch of NASA's ACE spacecraft in 1997 [Vogt et al., 2006] truly revolutionized space weather forecasting, and because of its ability to provide IMF conditions upstream of Earth and solar wind data in real time, it has become possible and relatively straightforward to issue short-term alerts and warnings, of say, a geomagnetic storm. ACE orbits the L1 libration point of the Sun-Earth gravitational equilibrium, which is approximately 1.5 million km from the Earth, and therefore, assuming an average solar wind velocity of 400 km/sec, one can get about 1 hour lead time on existing IMF conditions which will impact the Earth.

Unfortunately, a major limitation to using the geomagnetic indices is that their official values are only available after a fair amount of delay (months in some cases), whereas most real-time space weather applications require instantaneous estimates. Since reliable forecasts are not yet developed, "nowcast" algorithms are becoming common to better serve near-real-time applications by means of ground- or space-based instruments. The term "nowcasting" generally refers to specifying the state of a certain physical parameter in near real time by means of a ground- or space-based instrument. For example, NOAA/SWPC routinely provides instant estimates of 3-hour Kp derived by the US Air Force using magnetometers located mostly in the US and Canada. These are low-resolution Kp proxies, quantized in units of 1, and does not necessarily represent the official ones, which are quantized in units of 1/3, and

because they are measured using magnetometers residing within North America, they are not sufficient to provide accurate descriptions of the worldwide planetary activity. Often, as a result, the intensity of a magnetic storm could be over- or under-estimated depending upon the local time of that storm.

Space weather research can be viewed as a cross-disciplinary scientific effort or as a hybrid of basic space science research and applied science [Behnke, 2008]. A clear distinction between basic science research and space weather research is the predictive nature of the latter. As space weather forecasters and researchers explored the use of advanced techniques of information processing, Artificial Neural Networks (ANNs) were found to be particularly useful for both classification and prediction-type problems. ANNs have been used by several authors for over the last twenty years to develop models in geophysics (e.g., Newell et al. [1990] to classify geospace physical boundaries in the plasma data) and to devise time predictive algorithms for space weather (e.g., Koons and Gorney, [1991] computed relativistic electron flux at geosynchronous orbit and Wu and Lundstedt, [1997] for predicting Dst index). ANN methods have also been suitable to model and predict solar cycle activity using high-time resolution data from SOHO/MDI [Lundstedt, 2001]. Very recently, Valach et al. [2009] have used an ANN-based model to predict geomagnetic activity using solar energetic particle flux measurements. Most importantly, ANN-based models have been successful in making predictive estimates of Kp, Dst and AE ahead of time using real-time solar wind data inputs [e.g., Bala et al., 2009; Boberg et al., 2000; Costello, 1997; Wing et al., 2005; Wu and Lundstedt, 1997].

The NOAA Space Weather Scales were introduced as a way to communicate to the general public the current and future space weather conditions and their possible effects on people and systems (<http://www.swpc.noaa.gov/NOAAscales/>). They describe the space environment for three different event types: geomagnetic storms,

solar radiation storms, and radio blackouts. This dissertation aims to address the category of geomagnetic storms using a physical measure in Kp; the duration of the event influences the severity of the storms, and the space weather advisories are issued using scales ranging from G1 (Minor; Kp = 5) to G5 (Extreme; Kp = 9). Minor events can cause power grid fluctuations and cause minor impacts to satellite operations. Under extreme conditions, power systems could collapse causing blackouts, spacecraft and satellites may experience surface charging along with problems related to orientation, uplink/downlink, and passengers and crew in aircraft traveling across the poles may experience hazardous radiation levels. Therefore, forecasting geomagnetic indices, given the upstream solar wind conditions, has crucial importance from a space weather standpoint, not only as an indicator of the severity of the global magnetic disturbances but also as a relevant parameter currently used in the magnetospheric and ionospheric models, and to study the effects of space weather on satellite drag and help mitigate the risks of ground induced currents for electric power utilities, for example.

1.2.1 Solar Wind Coupling Functions

The ability to predict the geomagnetic response to the solar wind requires an understanding of how the two systems are coupled. One of the ways to characterize and quantify the coupling is through the use of solar wind-magnetosphere coupling functions. These are functions that represent the interaction between the solar wind and the magnetosphere while accounting for much of the energy needed to drive the different physical processes (e.g., energetic ion behavior, auroral power dissipation) in the magnetosphere. Specifically, they represent the dayside reconnection rate in terms of the upstream solar wind parameters, because dayside magnetic reconnection is largely responsible for the solar wind-magnetosphere coupling [e.g., Dungey, 1961;

Kamide and Slavin, 1986]. While some of these parameters are expressed in terms of the solar wind electric field [e.g., Burton et al., 1975 for vB_z , Doyle and Burke, 1983 for vB_T ; Reiff et al., 1981 for $vB_T\sin^2(\theta/2)$], some are expressed in energy rate or power [e.g., Perreault and Akasofu, 1978 for $\epsilon = vL_o^2B^2\sin^4(\theta/2)$; Gonzalez et al., 1989 for $(\rho V^2)^{1.6}vB_T\sin^4(\theta/2)$]. Regardless of how they are quantified, these parameters are basic derivatives of the solar wind and the interplanetary magnetic field.

In several published works on forecast models of magnetic indices, authors have only worked with basic solar wind parameters such as the magnitude of the solar wind velocity, solar wind pressure etc. in a variety of combinations. Recently, Papitashvili et al. [2000] used the 36-year long OMNI* dataset to show the correlation between the solar wind electric field and Kp and found a correlation coefficient of $r = 0.82$ suggesting that Kp can be derived from the solar wind velocity and IMF. More recently, Newell et al. [2007] have reported results of a solar wind coupling function ($d\Phi_{MP}/dt = v_{sw}^{4/3} B_T^{2/3} \sin^{8/3}(\theta/2)$) which appears to have good correlations with a number of geophysical parameters. The work presented in this thesis will test another similar function, the Boyle Index (BI) [Boyle et al., 1997], for its effectiveness in predicting values of various geomagnetic activity indices for use in real-time space weather forecasting.

A major motivation for this study was to use a single optimal solar wind coupling function to predict global magnetic indices. The BI is an empirically-derived scalar

**The OMNI data set contains hourly resolution solar wind magnetic field and plasma data from many spacecraft in geocentric orbit and in orbit about the L1 Lagrangian point. The data set also contains hourly fluxes of energetic protons, geomagnetic activity indices (AE, Dst, Kp) and sunspot numbers [<http://omniweb.gsfc.nasa.gov>].*

functional form given by

$$\Phi = 10^{-4} v_{sw}^2 + 11.7 B \sin^3(\theta/2) \text{ kV}, \quad \theta = \arccos(B_z/|B|) \quad (1.2)$$

where θ is the IMF clock angle, B is the magnitude of the IMF in nT, v_{sw} is the solar wind velocity in km/sec. The first term is a non-magnetic “viscous” term while the second term, called the magnetic “merging” term, stands for the effect of magnetic reconnection (detailed discussion to follow in chapter 4). Through the IMF data, the BI can characterize the asymptotic steady state potential drop across the Earth’s polar cap. The BI was derived by selecting events in which the IMF had a nearly “steady” direction for four hours, e.g. each component was either > 1.5 nT, or < -1.5 nT, or between -1.5 and $+1.5$ nT, over a four-hour long interval before the measurement of the cross polar cap (CPC) potential from the Defense Meteorological Satellite Program (DMSP) spacecraft. Boyle et al. obtained the four-hour steadiness criteria of the IMF as a reasonable representation of the convection pattern over the polar cap, in order to get the best quality of fit between the IMF and DMSP data set. Thus, the BI gives the value of the “asymptotic polar cap potential”. In other words, it is the value that the electric potential across the ionospheric flow would reach if the solar wind is steady for 4 hours. The BI was determined to be a good, statistically constrained, function for predicting the potential from the solar wind measurements, an improvement on previous such estimates [e.g., Reiff et al., 1981, Wygant et al., 1983]. Thus, although the BI was derived only to predict the polar cap potential, it is reasonable to use it as a possible coupling function to predict other measures of geomagnetic activity because geomagnetic indices can be modeled using solar wind derivatives. The following section provides justification for the proposed research.

1.3 Linear Predictor

The simplest way to use the BI, which is a linear index only dependent on the solar wind velocity and IMF, to predict Kp, which is a logarithmic index, is to correlate Kp with the log of BI (figure 1.5). Each point in the figure refers to observed values of 3-hour averages of the BI and the Kp index during the same time interval (e.g., between 0600-0900 Universal Time). The easiest approach to model Kp from the BI

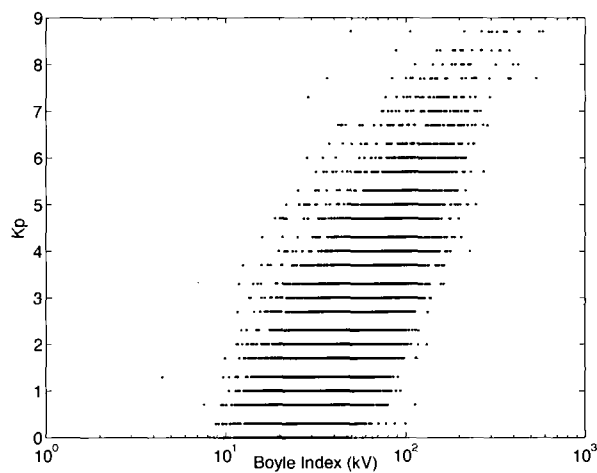


Figure 1.5 : Plot shows the 3-hour averaged $\log(BI)$ versus the subsequent 3-hour Kp for 2003 and 2004 with a linear correlation coefficient of 0.785 [Bala et al., 2009].

is from a linear fit because it uses a single value of the BI at a certain time interval to relate with the subsequent Kp rather than a long time history of the BI. The Kp index can be approximated from the BI by

$$Kp_{Linear} = 8.93 \log_{10}(BI) - 12.55 \quad (1.3)$$

Figure 1.6 shows the linear predicted Kp versus the official Kp index for April 2001, 2006 and 2007 using a 3-hour average BI (linear correlation $r = 0.766$), plotted by applying the fit from equation 1.3. 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., $K_p \geq 0$) for $BI > 25.4$ kV. This deficiency can be overcome by the more efficient ANN models (discussed later in chapter 5).

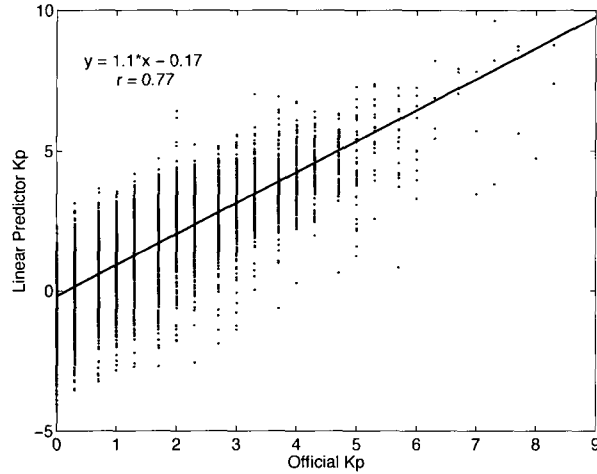


Figure 1.6 : Linear predictor: Official Kp vs Predicted Kp for April 2001, 2006 & 2007 using 3-hour averages of the BI. Solid line represents the best linear fit ($y = 1.1x - 0.17$).

A similar linear relationship can also be established between the BI and the Dst and AE, (figure 1.7) in the following manner:

$$Dst_{Linear} = -0.252 BI + 3.29 \quad (1.4)$$

$$AE_{Linear} = 4.244 BI - 33.11 \quad (1.5)$$

Using these fits from equation 1.4 and 1.5 we can plot the predicted Dst (figure 1.8) and AE (figure 1.9) for 2006 and 2007 using 1-hour averages. These fits becomes a standard against which we will compare the ANN results. The fact that the linear correlation coefficient between the predicted and observed is quite significant, encourages us to use the BI as the basis for the neural network formalism to follow, and sets

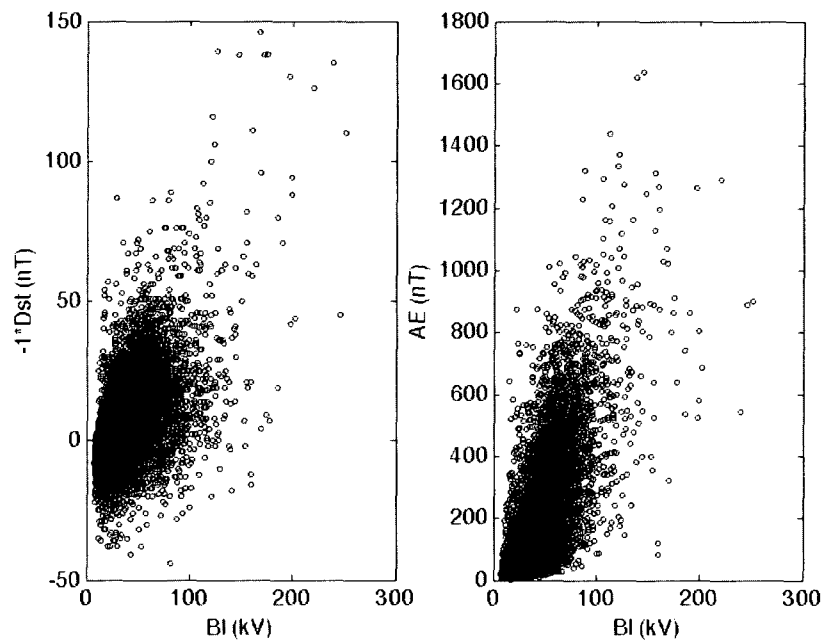


Figure 1.7 : 1-hour averaged BI versus the Dst (left panel) and AE (right panel). Note the value of Dst is deliberately multiplied by -1 for easy plot comparison.

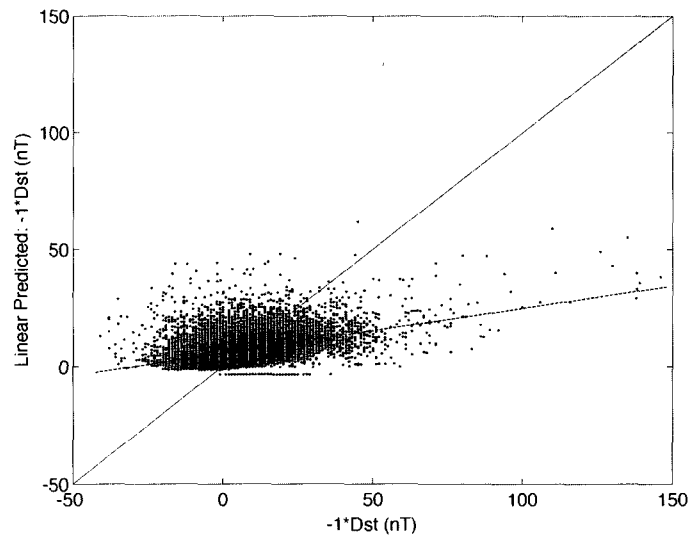


Figure 1.8 : Linear predictor: Measured Dst vs Predicted Dst for 2006 & 2007 using 1-hour averages of the BI. Solid line is the equality line and the dashed line is the linear fit.

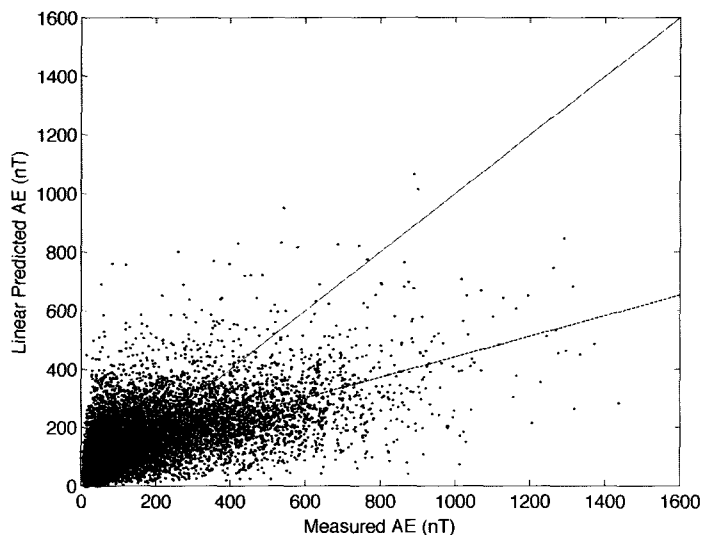


Figure 1.9 : AE Linear predictor: Measured AE vs Predicted AE for 2006 & 2007 using 1-hour averages of the BI. Solid line is the equality line and the dashed line is the linear fit.

a benchmark for judging significant improvement over this baseline. However, the temporal correlations between the BI and the geomagnetic indices still remains to be explored, which will be discussed explicitly in chapter 5 using cross-correlations.

1.4 Persistence forecasting

Persistence forecasting is a useful scheme in which the conditions about the near-future can be characterized using the prevailing or near-past conditions. It is the easiest way to predict the future. In other words, we are forecasting using common sense! Persistence can be defined as the existence of statistical dependence among successive values of the same variable or an event [Wilks, 1995]. It works best for short-term predictions, especially during times when the conditions are fairly quiet. A series of large positive values is likely to be followed by another large value, and the same can be said of a series of small values. This behavior is known as “positive

dependence” or “persistence”.

For example, given that Kp is a 3-hour index, it has a strong tendency for persistence. In other words, for a certain 3-hour interval when the Kp was reported to be above average, the probability that the Kp for the next 3 hours is very likely to be the same or higher is high. However, while it is true that this forecasting scheme works roughly about two-thirds of the time, such an unsophisticated forecasting scheme is neither acceptable nor entirely useful. This is critical to this dissertation because two of the Kp models (to be introduced in chapter 3) will utilize the time history of Kp. Even though it might improve the predictions, it is imperative to distinguish a good “prediction” from a true “forecast”. The only way one can demonstrate this effect of “persistence” or “persistence contamination” is by using autocorrelation arguments (see chapter 3). The statistical consequences of persistence are further explored through examples in chapter 5, since two of my models utilize previously known values for prediction. A more sophisticated and accurate forecasting can be achieved through artificial neural networks. In particular, they are powerful enough that good predictions can be made only using inputs constructed from solar wind.

1.5 ANN and forecasting

In the field of space physics, ANNs have consistently proved effective for temporal prediction problems particularly those which may be described as highly dynamic and non-linear, where the goal is to derive what the future information of a specific parameter of interest will look like given sufficient and necessary information leading up to the current time. We represent the available information using *time series* of measured values which may be continuous or discrete in time e.g., $B(t)$ where $1 \leq t \leq n$ for n instances in time. A network that is trained to make predictions based on

consecutive inputs from $B(t)$ uses a set of r recent values to predict the next time step, $B(t = r + 1)$. This is a short-term *one-lag* prediction problem where we forecast the next value from the recent history.

A magnetospheric “preconditioning” can be described as a mechanism that tunes the magnetosphere to a certain state, prior to a storm over a period of several hours. Since neural networks are inherently learning machines, this gives the algorithms the ability to learn the effects of preconditioning and other nonlinearities residing in the system. In this dissertation, I have explored the possibility of using bivariate time series that are concurrently changing with time. To be successful, forecasting must be based on all available correlations and empirical interdependencies among different temporal sequences [Mehrotra, 1997]. I use *feedforward* networks for forecasting and these will be discussed in chapter 2. The BI, with a long timeline of observations from ACE, thus, provides an excellent platform to develop such an ANN-based forecast model.

1.6 Related Work

In the past, several studies have demonstrated a good correlation between various geomagnetic indices and the IMF and with other parameters of the solar wind and it is now a fairly well-accepted fact that the magnetosphere responds to variations 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. With the presence of the ACE spacecraft and its ability to provide IMF conditions of upstream Earth and solar wind data in real time, it has become possible to estimate geomagnetic indices up to 3 hours in advance [e.g., Wing et al., 2005]. Motivated

by the aforementioned facts, several algorithms have been developed to nowcast Kp [e.g., Takahashi et al., 2001] and forecast Kp, Dst, and AE [e.g., Boberg et al., 2000; Costello, 1997; Pallocchia et al., 2007; Wing et al., 2005; Wu and Lundstedt, 1997]. In fact, one such algorithm based on an ANN, developed by Costello [1997] and currently being used by the US Air Force, takes solar wind data as inputs and generates Kp every 20 minutes with an early warning of up to 30 minutes in advance (see section 1.5.2 for details). Given the accuracy of these estimates and the generally short forecast times, there is an opportunity for significant improvement.

The following sub-sections will describe the evolution of these models over time and some work related to that presented in this thesis, highlighting their important results. These models were built to operate in a variety of modes meant to satisfy different operational constraints using solar wind parameters as their primary data stream. The common denominator in these prediction models is the use of ANNs.

1.6.1 Takahashi Kp Nowcast Model

Takahashi et al. [2001] developed a one-of-a-kind algorithm to derive Kp_{est} (for estimated Kp) using magnetometer data from nine ground observatories which can transmit data in near real time. Their algorithm involve routines to clean the data, estimate and remove a quiet-day-curve (QDC) and eventually convert the data to Kp_{est} . The stations that they chose were the same ones that provide near-real-time data to the NOAA/SWPC through satellite or phone link. Their estimated values are consistent with Kp showing a linear correlation coefficient of 0.93. Even in the event of technical glitches and malfunctions at the stations, their robust design still allows them to derive Kp proxies and offer a good estimate. It has become popular with several forecasters and is routinely in use today (e.g., see the APL models 1 and 2 in section 1.6.4).

1.6.2 Costello NN Kp Model

The Costello ANN Kp model [Costello, 1997] was developed at Rice University primarily to provide input data for the Rice Magnetospheric Specification and Forecast Model (MSFM) which uses Kp as one of its basic input parameters. NOAA routinely uses this model to provide real-time short term Kp forecasts which are available from <http://www.swpc.noaa.gov/rpc/costello/>. It uses the solar wind speed, IMF B_z and $|B|$ inputs from ACE to provide Kp proxies once every 15 minutes with a lead time of approximately one hour. The actual predicted lead times are bound to vary, given variations in the solar wind speed. For comparison purposes, I use the value for the correlation coefficient, i.e., between the Costello NN forecast Kp and official Kp, evaluated by Wing et al., [2005], namely $r = 0.75$ as a base result indicating a strong correlation. However, my own evaluation of this model from a head-to-head test with one of my models will be discussed later in chapter 5.

1.6.3 Boberg NN Kp Model

The Boberg Kp model (Boberg et al., 2000) is also an ANN model which uses the solar wind and IMF data, just like the Costello model. Here, the model inputs the IMF B_z , solar wind density n , and the solar wind velocity v to produce an output that is a predicted 3-hour Kp index and is achieved using multilayer feed-forward neural network. The live results are obtained from <http://rwc.lund.irf.se/rwc/kp/models.php>. With a correlation coefficient of $r = 0.765$, their model performance during both active and quiet times closely resembles that of the Costello Kp model. It must be noted, however, that unlike the Costello NN Kp model, I do not have the means to access the Boberg model or its output for a personal evaluation, and therefore rely on the figures and numbers quoted in their paper.

1.6.4 APL Kp Models

Wing et al. [2005] developed three real-time Kp models called the APL models: (1) a model that take inputs from ACE to nowcast Kp (from Takahashi algorithm) and predicts Kp 1 hour ahead ($r = 0.92$); (2) a model with the same inputs as model 1 and predicts Kp 4 hours ahead ($r = 0.79$); and (3) a model that predicts Kp 1 hour ahead using inputs from solar wind only ($r = 0.84$, see figure 1.10). Their forecasted Kp can be obtained in real time from <http://sd-www.jhuapl.edu/UPOS/ForecastingKP/index.html>. Fundamentally, their models include IMF $|B|$, B_z , $|V_x|$ and the dynamic pressure term, n in their inputs. Unlike models 1 and 2, model 3 best compares to the Costello NN Kp and Boberg NN Kp models since it does not include or input the nowcast Kp, but the Wing et al model clearly outperforms the Costello and Boberg models.

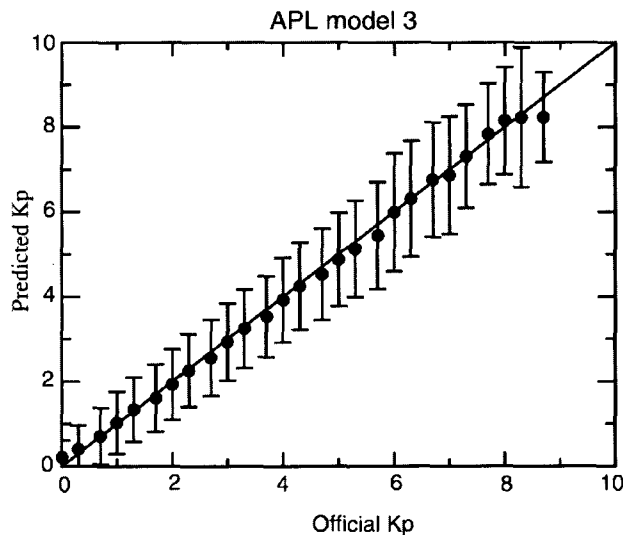


Figure 1.10 : Plot showing Predicted Kp vs Official Kp for the APL model 3, $r = 0.84$. [Adapted from Wing et al., 2005]

The APL model 2 uses the time history of Kp but does not have the luxury of using the last value because of lack of instant availability of any nowcast Kp data, leading

up to the forecast time of four hours, and the maximum solar wind propagation time for any algorithm using ACE is approximately 45 minutes to an hour. Therefore, their model 2 has a good “time” advantage over their own model 1. The success of their models 1 and 2 is largely attributed to the presence of the target index in their input stream. In general, using measured Kp to predict following Kp causes “persistence contamination” which means that predictions over-rely on the previous measured Kp. The effects of that contamination are seen as a “lag” feature in plots 5 (g) and 5 (h) of their paper (shown here in figure 1.11). However, their model 3, which depends only on the solar wind and not on Kp, does not exhibit this lag or show evidence of persistence contamination.

Wing et al. [2005], while describing their model results, plot the predicted Kp as the Kp not directly from the NN, but rather as the best-fit Kp linear function as shown in figures 6,7 and 8 of their paper. It is not clear whether this best-fit function is then fitted back at the end using error bars, and in which case, might possibly skew their own validations. Nevertheless, a real testimony to the success of their models can be described in terms of the wide acceptance they receive.

1.6.5 Wu and Lundstedt NN Dst model

Wu and Lundstedt [1997] studied the solar wind-magnetosphere coupling by predicting geomagnetic storms (Dst) using partially recurrent neural networks; here again, a variety of combinations of the solar wind parameters can be used to give accurate predictions. They applied statistical correlation studies to find the best coupling functions, and concluded that the best combinations giving accurate predictions are B_s (if $B_z < 0$, $B_s = -B_z$), ρ_{sw} , v_{sw} and B_z , ρ_{sw} , v_{sw} . For predictions 1 hour and 2 hour ahead, they claim the linear correlation between the real and predicted Dst is 0.90, and also claim that all phases of geomagnetic storms can be predicted accurately in

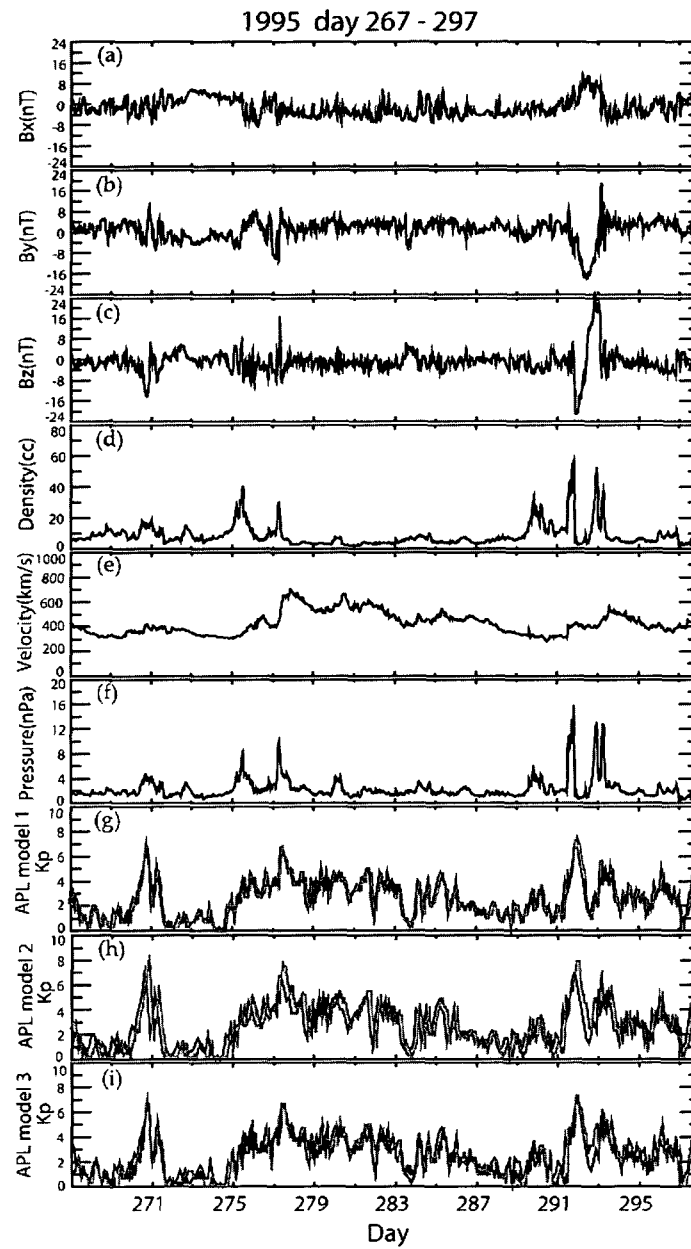


Figure 1.11 : The performance of the APL models are shown here for comparison. Note that the models 1 and 2 which includes K_p history as inputs appears to lag the real data (black curve) [Adapted from Wing et al., [2005]].

that time range. For 3 to 5 hour predictions, the linear correlation varies from 0.88 to 0.84. While for predictions 6-8 hours ahead, the correlation varies from 0.82 to 0.77.

1.6.6 Temerin and Li Dst Model

Following the BMR equation, several attempts, mostly through empirical studies, have been made to test and improve the values of the constants obtained by Burton et al., notably Gonzalez et al., 1994, O'Brien and McPherron, 2000. Among the other models that are not based on ANN is the Temerin and Li [2002; 2006] model. They provided a new model for the Dst based on the solar wind parameters through rigorous empirical studies. Their model predictions are made available in near real time, based on data from ACE, which are currently available at <http://lasp.colorado.edu/~lix>. However, this is largely a nowcast rather than a forecast.

1.6.7 Gleisner and Lundstedt NN AE model

Gleisner and Lundstedt [2001] used an Elman recurrent network [Mehrotra, 1997] for the prediction of the auroral electrojet index AE from solar wind data. They used the solar wind parameters B_z , ρ_{sw} , v_{sw} as inputs to show that an Elman recurrent network can predict around 70% of the observed AE variance using single sample of solar wind density, velocity and magnetic field as input.

1.7 Scientific Objective

ACE data include a broad spectrum of conditions responsible for geomagnetic activity. This study will use the unprecedented quality of solar wind data provided by ACE to determine the best function since 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 timeline of observations from ACE, with up to one minute resolution, and the history of Kp, Dst and AE data, provides an excellent opportunity to develop a prediction algorithm for the next upcoming time intervals. Furthermore, data from a complete solar cycle, and over a variety of solar events will facilitate the ANN training (please see chapter 3 on Research Methodology).

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, correlations between the BI and Dst and AE 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 main content of my thesis and the ways it contribute to short-term space weather forecasting are summarized as follows:

- To show that the logarithm of BI, with its viscous and merging terms, is linearly correlated with Kp and that it can be used to predict Kp when coupled to an Artificial Neural Network. Likewise, Dst and AE can also be predicted from the time history of the BI.
- To successfully demonstrate that it has an enhanced forecasted-time range over the popular forecasting algorithms available today.
- To investigate the effects of “persistence contamination” in some of the existing algorithms, including my own, and that it is indeed possible to issue a reliable forecast from solar wind data alone, just as successfully as algorithms that use the previous time history of the target index.
- To successfully integrate the near-real-time predictions from my models to

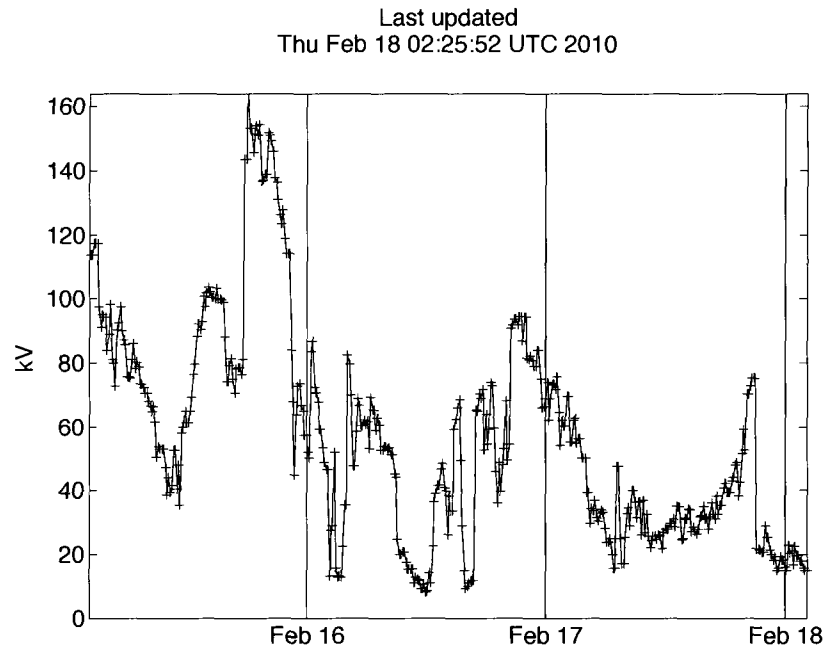


Figure 1.12 : Shown here is a typical BI plot as seen in real time, reporting the most recent 72 hours of its history.

<http://space.rice.edu/ISTP/wind.html> (courtesy of the *Rice Space Institute*), originally dedicated in October, 2003 to show the BI in real time and its past history (figure 1.12). Further, I will improve the “spacalrt” system so that subscribers receive email notices of “red alerts” from well defined thresholds e.g., whenever the 10-minute BI average exceeds 200 kV, or whenever the predicted K_p exceeds 6 .

1.8 Thesis Organization

My thesis introduces a novel, fully-automated time predictive algorithm with prediction capabilities for moderate (e.g., $3 < K_p < 6$) to severe storms (e.g., $Dst < -120$ nT, $K_p > 6$) and to forecast K_p , Dst and AE up to approximately 3 hours ahead in near-real time. In the work chapters, I focus on training the ANN using

solar wind data, drawing insights from correlation analysis; chapter 5 deals with this effort. Chapters 2 and 3 give a thorough overview of the ANN fundamentals and the research methodology to produce to best architecture. Chapter 4 provides us the basic physical processes behind “magnetospheric convection”, and then introduces the Boyle Index. Chapter 5 deals with the algorithms, verification and validation tools applied to the developed models. I also perform a head-to-head test with the Costello ANN Kp model and show the results. A final summary of my results and future research directions are provided in chapter 6.

Although, in colloquial speech, the terms predict and forecast are often used interchangeably, technically the term forecast refers to making a prediction of future events based on the recent past, whereas a numerical prediction can be used in retrospective analyses of historical data [Singer et al., 2001]. Therefore, this thesis uses predictive analyses of such historical data to create a functional form that is now being used in real time to routinely, and accurately forecast space weather.

Chapter 2

Artificial Neural Networks

In this chapter, the primary intent is to elucidate the minimization technique used to train our models. To begin our discussion, I will introduce some of the oft-used terminology and outline the fundamental concepts behind an ANN (Artificial Neural Network).

2.1 Artificial Neural Networks

Neural Computing or “neural networks” is a branch of computing that is designed to adopt the basic structural architecture of a human brain to perform a certain task or function. Typically, the framework for such networks are built on simple mathematical functions. The fundamental property or building block that an ANN shares with a human brain is called the neuron. In humans, any mechanical or non-mechanical task such as perception, recognition or memory requires the use of our brain. An ANN, of course, uses an artificial neuron to learn from data samples that require short- or long-term memory (e.g., classification, forecasting, pattern identification). In order to understand an ANN, we need to know what a neuron is.

2.1.1 Biological Neurons

A biological neuron fundamentally comprises a cell body, an axon, and a large number of dendrites as shown in figure 2.1. Neurons process information using electrochemical signals received through their dendrites which are further propagated or suppressed

depending upon the threshold level of the cell. A propagated signal, “fired” by a neuron, is carried to the neighboring neuron across a synapse; the synapse sits at a junction between the dendrite of a neuron and an axon of another neuron allowing the flow of information. The axons carry the processed information away from a neuron to another neuron. However, the neurons in the human brain have a far more complex structure than the simple picture painted here, for there are more than 100 billion neurons each with over 7000 synaptic weights. Yet, this simple analogy is an effective way to introduce an artificial neuron, given the scope of this dissertation.

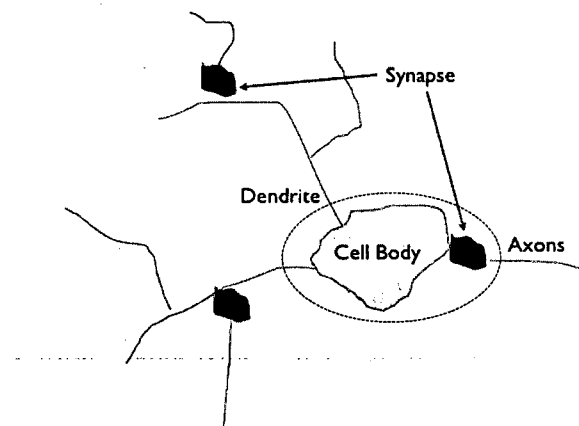


Figure 2.1 : A biological neuron (dotted oval) is composed of a cell *body*, a multitude of *dendrites*, *synapses* and an *axon*.

2.1.2 Artificial Neurons

The notion that ANNs are not meant to replicate the neurological details of a human brain but rather use it as a basic design platform to build mathematical models is generally underappreciated. An artificial neuron is a function-processing unit that

is key to the operation of a network of artificial neurons. Here, a set of inputs or a pattern (dendrites' synapses) is propagated through a corresponding set of "synaptic" weights. The incoming weighted information is then summed before being activated by a transfer function, which is usually non-linear, to produce an output signal of that node ("neuron"). Each node has a single output value which can either be fed back or passed on to another neuron. This process can be modeled using equation 2.1 for a single neuron.

A single neuron, however, may not be useful to solve a complex problem. In such instances, and in order to increase the available resources, the necessary computational power or "artificial intelligence" is derived by marshaling a network of processing elements or nodes, and such networks are designed based on the problem in hand. This will be illustrated by examples in section 2.3. Each neuron can be connected with many others through different topologies (e.g., layered networks), though not necessarily with its own neighbors [Mehrotra et al., 1997]. However, the laws of solid state physics imposes a limit on the computational power of the processors used, which in turn, limits the network size and the number of neurons that can be accommodated, significantly smaller compared to the tens of billions of biological neurons available to a human brain. An example of a processing element as a non-linear neuron model is shown in figure 2.2 (details in section 2.2) .

Biological neurons acquire knowledge through basic perception and experience, and when posed with a problem, are able to respond and react remembering from their acquired knowledge base. So let us raise the fundamental question of how an artificial neural network learns. In short, the answer is through adjustable "weights". More specifically, an ANN learns by adjusting its weights in response to the input-output patterns through a process called "learning". Learning is said to be complete when the output response reaches a desired level of accuracy specific to an input

pattern. The field of neurocomputing has achieved considerable progress over the last few decades to the extent that there exists a multitude of learning techniques.

The learning rules or the adaptation of the weights of an ANN can be classified into two large domains: supervised and unsupervised. In supervised learning, the network is initially provided with targets or “correct” answers which may be linear or non-linear functions of a specific set of inputs (e.g., forecasting to predict a future event based on past history). The weights are adjusted based on the error at the output after data are propagated from inputs to outputs. The process is iterated until the error between the actual and desired outputs is minimized. The weights are now said to be optimized. Unsupervised learning, unlike the supervised learning, does not have the luxury of knowing the target value or the “correct” answer beforehand e.g., vector quantization where the process is to divide up space into several connected regions and each point in space belongs to ones of these regions, clustering where it requires grouping together objects that are similar to each other [Mehrotra et al., 1997]. The system is expected to teach itself until it arrives at the “best” answer as determined by the optimized weights i.e., the error between measured and target values fall within an asking range. The technique that is germane to the examples below and eventually to my ANN forecast models, is a supervised learning rule.

2.2 A Non-linear Neuron Model

A simple *perceptron* is essentially a pattern-recognition machine whose learning rule was first discussed by Rosenblatt in 1962 and is illustrated through figure 2.2. For the p^{th} input pattern, say, $\mathbf{x}^p \equiv \{x_1^p, x_2^p, \dots, x_n^p\}$, the output o^p is written in this simple mathematical form by

$$o^p = \varphi(v^p) = \varphi\left(\sum_{i=0}^n w_{ji}x_i^p\right) \quad (2.1)$$

where $\varphi()$ is known as the “activation” or “transfer” function, w_{ji} represent the synaptic weights of their corresponding inputs x_i^p , ν^p denotes the net input, and o^p is the output of the neuron where there are n total inputs. x_0 , called as the bias neuron, here set at 1 (please see Haykin, 1999).

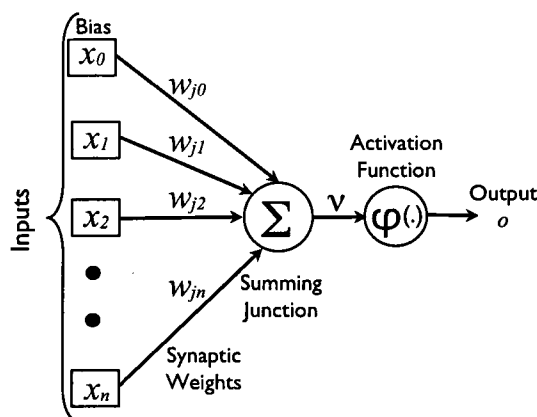


Figure 2.2 : A non-linear neuron Model. Σ and φ are defined in equation 2.1.

The activation function can be a linear function, non-linear functions such as the step function or the hyperbolic-tangent, and is chosen based on the specific problem at hand. The single most popular activation function used to build an ANN is the *sigmoid* function because it renders great advantages to any learning algorithm not only by being differentiable with non-zero values everywhere but also by the ease with which its derivatives are computed. It is also easy for algorithms to take advantage of the smoothness it offers. In fact, their input-output curves are similar to their biological counterparts. Experimental observations have shown that biological neurons and their neuronal firing rate is roughly sigmoidal, when plotted against the net input to a neuron [Mehrotra, 1997]. These “S-shaped” functions come in a variety of flavors

whose output is non-linear and asymptotically tends to lower and upper bounds at $-\infty$ and $+\infty$. This property can be appreciated later in section 2.3 when we discuss multilayer perceptrons as each neuron in such an architecture requires the knowledge of the derivative of the activation function for weight updates. Figure 2.3 illustrates the output of a commonly used sigmoid function called the *hyperbolic-tangent*,

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (2.2)$$

The resulting output of this configuration o^p is also known as the “firing” of a

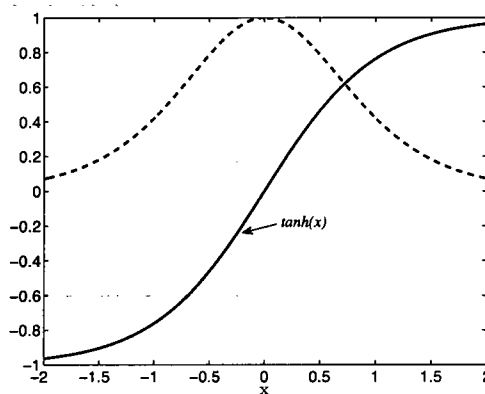


Figure 2.3 : A sigmoid function $\tanh(x)$ and its derivative (dashed curve).

neuron. The weight vector is adjusted each time a training pattern is presented, and in proportion to the error at the output until the algorithm converges. The weights for the next training cycle are adjusted according to

$$\Delta^p w_{ji} = \gamma x_i^p (D^p - o^p), \quad (2.3)$$

where $0 < \gamma < 1$ is the *learning rate* or step size and D^p is the desired target. The following sections will explain how these work through examples.

2.3 The Backpropagation Algorithm

A simple perceptron learning technique or perceptron-convergence procedure, unfortunately, is useful to solve only a limited set of problems i.e., problems that are linearly separable, because their inputs and outputs are linked directly and no other layers of processing elements lie between the input and output layer. The advantage is that the learning rule is relatively straightforward to develop. However, in dealing with a higher order problem or problems that are linearly inseparable (please see example in figure 2.4), one needs to effectively raise the learning potential by tapping a network of such perceptrons, the basic premise behind which a human brain functions. There are three different ways of how one can choose their network architecture to intimately connect a cluster of neurons in pursuit of better learning architectures: single-layer feedforward networks, multilayer networks, recurrent networks. For my work, I have

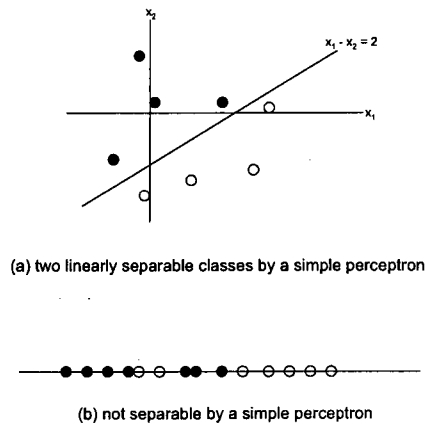


Figure 2.4 : An example of a two-class classification (black and open circles) : (a) is linearly separable by a simple perceptron, and (b) is not separable by a simple perceptron.

chosen the multilayer perceptron (MLP) architecture, a form of feedforward network, which can accommodate one or more layers besides the input and output layers. A layer that connects the input layer with the output layer is called *hidden layer*. A MLP network consists of an input, output and one or more hidden layers. An input layer is a non-computation layer consisting of input nodes, an output layer is a computation layer from which the final outputs of the network can be read, and hidden layers are also computation layers that are not part of either the input or output but intervene between the input and output layers. Having hidden layers helps to represent important features within a task [Rumelhart et al., 1986] i.e., to extract complex and useful features between input-output patterns. This was a motivating concept that first came into existence in the 1960's, but largely remained theoretical until 1974. The goal then, was to identify a way for synaptic weight modification in the context of a multilayer perceptron (MLP) as the simple perceptron learning rule cannot be implemented in an MLP with such extended network connectivity. An example of a two-layer MLP architecture is shown in figure 2.5. Here, the weights are indicated by arrows, for they represent the connection topology where the processing elements are the units that are connected.

The first original training algorithm for a MLP, which was considered one of the major breakthroughs in ANN history, was proposed by Paul Werbos in 1974, but did not catch on until 1986 when Rumelhart et al. popularized it in their book "Parallel Distribution Processing". They effectively demonstrated this training algorithm on a multilayer perceptron using a new learning rule called the backpropagation (BP) algorithm or simply called "backprop", a new formulation based on the old gradient descent technique. In a backprop algorithm, unlike a simple perceptron, there are two phases: a feedforward phase where the function signal propagates forward and a backprop phase where the error signal propagates backward. In the feedforward

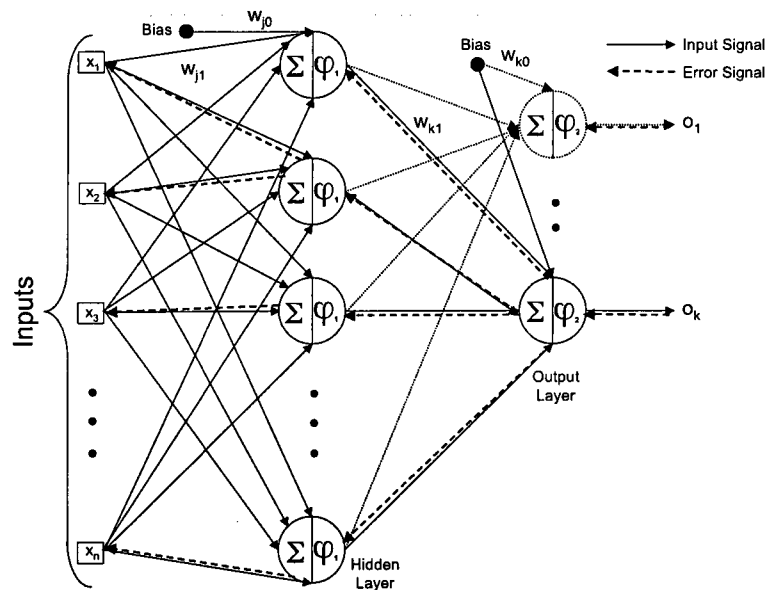


Figure 2.5 : Shown here is a generic architecture of a two-layer feedforward network which may be extended to any number of hidden layers.

phase, an input pattern presented to the input layer is processed by the neurons in the hidden layer where the weighted sum of its inputs is passed to the activation function before presenting their final outputs to the adjoining layer which may be either another hidden layer or the output layer. Function signals continue to flow across each layer until it reaches the output layer where the errors are computed based on the knowledge of the final output and the desired target. Once the errors are known, the weight corrections in each layer will be made during the “backpropagation phase” as the error signal propagates back and towards the input layer carrying the error. To understand the backprop algorithm, it is important to understand the weight-update rule based on the gradient descent procedure.

2.3.1 Learning through Gradient-descent Technique

A very easy-to-implement and a general purpose weight-minimization rule applicable to various flavors of the backprop algorithm is the gradient-descent. The main objective of this learning rule is to minimize the mean squared error or the cost function, a function of all the free parameters such as the weights and thresholds, at the output layer. Let us consider a two-layer feedforward network as the one in figure 2.5 with i, j and k denoting the nodes in the input, hidden and the output layer respectively, and where there are n, m and K total number of nodes in each layer.

Let's propagate a certain input pattern $\mathbf{x}^p \equiv \{x_1^p, x_2^p, x_3^p, \dots, x_n^p\}$ forward; x_i^p represents a certain input from the i^{th} node to the hidden layer. The net input to the j^{th} node in the hidden layer is then $\sum_{i=0}^n w_{ji}x_i^p$, where $x_0^p = 1$; the weights connecting the i^{th} node of the input layer to the j^{th} node of the output layer is denoted by w_{ji} . The output of the neuron j at the hidden layer is $x_{j(1)}^p = \varphi_1(\sum_{i=0}^n w_{ji}x_i^p)$ where $\varphi_1()$ is the *tanh* function. Next, the net input to the k^{th} node in the output layer is given by $\sum_{j=0}^m w_{kj}x_{j(1)}^p$, again $x_0^p = 1$; the weights connecting the j^{th} node of the hidden layer to the k^{th} node of the output layer is denoted by w_{kj} . The output of this node k is $o_k^p = \varphi_2(\sum_{j=0}^m w_{kj}x_{j(1)}^p)$ with $\varphi_2()$ the *tanh* function. The desired target output at the k^{th} node is D_k^p . For the given pattern, the total error can be written as a sum of the squared error (or the *cost function*) of all the output neurons using

$$E_k^p = \frac{1}{2}(D_k^p - o_k^p)^2 \quad (2.4)$$

where E_k^p defines the squared error. The total error for a pattern p is given by $E^p = \sum_{k=1}^K E_k^p$. Subsequent weight states are computed depending on the propagated error at the output layer in proportion to the local gradient of the cost function. The weight adjustment is made based on the knowledge of the direction of the local gradient of the error surface i.e., if the gradient is positive then the weights are

adjusted in the negative direction to decrease the error and vice versa. Thus, the weight-update rule requires computation of the partial derivatives of the local error with respect to the weights in the network.

Mathematically, the weight change to be applied to the output layer is

$$\Delta^p w_{kj} = -\gamma \frac{\partial E^p}{\partial w_{kj}} \quad (2.5)$$

Similarly, the weight change to be applied to the hidden layer is

$$\Delta^p w_{ji} = -\gamma \frac{\partial E^p}{\partial w_{ji}} \quad (2.6)$$

Further, expanding (2.5) using the chain rule, we get

$$\frac{\partial E^p}{\partial w_{kj}} = \frac{\partial E^p}{\partial \sigma_k^p} \frac{\partial \sigma_k^p}{\partial \varphi_2} \frac{\partial \varphi_2}{\partial w_{kj}} \quad (2.7)$$

after calculating the appropriate partials, we get

$$\frac{\partial E^p}{\partial w_{kj}} = - \sum_k (D_k^p - \varphi_2(\sum_j w_{kj} x_{j(1)}^p)) \cdot \varphi_2'(\sum_j w_{kj} x_{j(1)}^p) \cdot x_{j(1)}^p, \quad (2.8)$$

φ_2' denotes the partial derivative, which when inserted into (2.5) yields the weight update rule for the output layer

$$\Delta^p w_{kj} = -\gamma \delta_{k(2)}^p x_{j(1)}^p \quad (2.9)$$

where

$$\delta_{k(2)}^p = \sum_k (D_k^p - \varphi_2(\sum_j w_{kj} x_{j(1)}^p)) \cdot \varphi_2'(\sum_j w_{kj} x_{j(1)}^p) \quad (2.10)$$

represents the local error at the output node(s) also known as the “delta”. Similarly, we can write down the weight corrections to be applied to the hidden layer through

$$\Delta^p w_{ji} = -\gamma \delta_{j(1)}^p x_i^p \quad (2.11)$$

where

$$\delta_{j(1)}^p = \left(\sum_{k=0}^m \delta_{k(2)}^p w_{kj} \right) \varphi_1' \left(\sum_{i=0}^n w_{ji} x_i^p \right) \quad (2.12)$$

The presence of the φ' term facilitates the rate of change of weights to accelerate in those regions of the weight space where φ' is large. One beauty of this technique lies in the sleek way of computing the partials. This procedure is much more accurate than the simple perceptron weight-update rule presented earlier in equation 2.3 and is only slightly harder to adapt into a numerical problem.

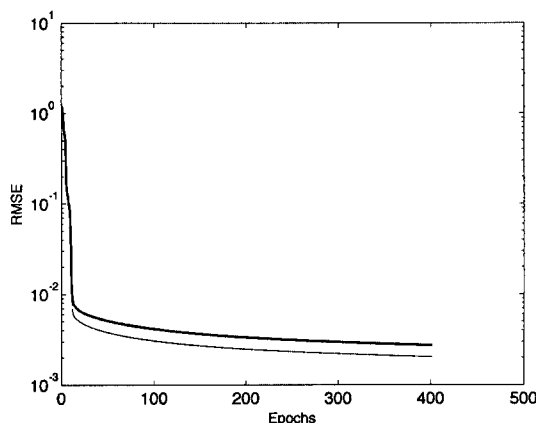


Figure 2.6 : An example of a learning curve for a function approximation problem ($y = \frac{1}{x}$). Shown here is the RMSE Error for the training (thin curve) and test data (thick curve), recalled at the end of each time step. 1 epoch = 100 training patterns.

Figure 2.6 is an illustration of the time evolution of the performance of the network as the error function is minimized. Ideally, one would want the algorithm to recall the error periodically on both the training and validation set to conduct an instant evaluation of the network as the training proceeds, i.e., as the cost function is minimized. This, to a large degree, helps the network from under- or over-fitting the data, and ensures the algorithm produces the best weight-minimization possible (figure 2.7).

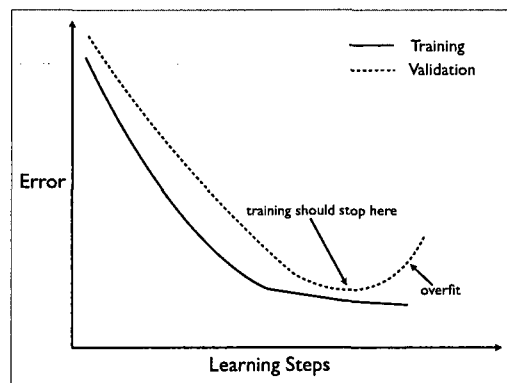


Figure 2.7 : Cross-validation: recalling the error on the training and validation data simultaneously as training proceeds. Learning is terminated as soon as the error on the validation set begins to rise.

A simplified version of the working architecture of my proposed models involving a two-layer feedforward network can be best represented using the figure 2.5 earlier. My actual models, however, involves a network with one output node and a set of nodes in the hidden layer; the number of neurons in the hidden layer for each model is determined based on trail and error. The crux of the proposed forecast algorithms, to be discussed later in the thesis, involves backpropagation as its principal learning technique, and the fundamental training steps applicable to our model specifications are outlined as follows:

1. Initialize weights randomly in both layers $-0.9 < w_{ji}, w_{kj} < 0.9$; set the learning rate γ .
2. Choose an input pattern and apply it to the input layer.
3. Propagate the pattern through the weights and the activation function forward at the hidden layer and then on to the output layer.

Note: The output layer activation function is typically a bounded function of

the form given by figure 2.3 and whose derivative is also bounded. Even though I did not notice a significant difference in the model performance when the two functions were used interchangeably (*sigmoid* versus *linear*) at the output layer, all my models employ a linear activation function at the output layer and a non-linear activation function (*sigmoid*) in the hidden layer. In nonlinear regression analysis, the output layer transfer function is preferred to be linear. In this case, the approximation error at the output can be modeled as *white Gaussian noise* [Haykin, 1999], and this property means that the amplitude falls in the range $(-\infty, \infty)$ requiring the output layer neuron to be linear.

4. Once at the output layer, compute the error δ_k^p using (2.10) and the target value D_k^p corresponding to the input pattern; p denotes each pattern; the function $\varphi'()$ reduces to a constant value.
5. Now compute the error at the hidden layer nodes using $\delta_{j(1)}^p$.
6. Update the weights using $\Delta^{p(2)}w_{kj} = -\gamma\delta_k^p x_{j(1)}^p$, $\Delta^{p(1)}w_{ji} = -\gamma\delta_{j(1)}^p x_i^p$; γ is the learning rate and takes values between 0 and 1.

Note: It is easy to see that the weight update equation for the hidden layers has the same form as the output layer. However, the way the local errors are computed at the hidden layer is unique to the backprop algorithm.

7. Go through steps 2 to 6 for all patterns until the error reaches a desired minimum value.

These step-by-step instructions are easy to implement and could be generalized to any MLP network with a BP rule. Larger backprop-based networks may take several iterations and sufficiently longer to train. A walk through of the detailed algorithm

is illustrated in the following subsection, and can be adapted to any problem with a two-layer backprop network.

2.3.2 Backpropagation algorithm for a two-layer network

```

1: set number of hidden layer neurons; output layer neuron
2: set learning rate
3: set batch /* number patterns for batch learning */
4: set goal /* set performance goal */
5: set activation function  $\leftarrow$  hyperbolic - tangent
6: initialize weights /* both layers */
7: old weights  $\leftarrow$  weights /* recalled during weight update */
8: set bias  $\leftarrow$  1
9: while error > goal do
10:  for i = 1 to batch do
11:    x  $\leftarrow$  (pattern, 1) /* the bias always equals 1 */
12:    output  $\leftarrow$  desired output
13:    output at hidden layer y  $\leftarrow$  activation function(hidden layer weights * x)
14:    output at output layer yo  $\leftarrow$  activation function(output layer weights * y)
15:    delta at output layer  $\leftarrow$  (1 - yo * yo) * (output - yo)
16:    error  $\leftarrow$  error + (output - yo)2
17:    k  $\leftarrow$  output layer weights * error at output layer
18:    compute delta at hidden layer
19:    weights at output layer  $\leftarrow$  learning rate * delta at output layer * y
20:    weights at hidden layer  $\leftarrow$  learning rate * delta at hidden layer * x
21:  end for
22: end while

```

23: *store weights*

2.3.3 Notes on Backprop

One of the problems I encountered during the course of this work is finding an algorithm that has the right mix of speed and stability. The standard backprop algorithm based on the gradient-descent procedure used to train a MLP NN, by the virtue of its simplicity, is one of the most successful and widely used algorithms [Ham and Kostanic, 2001], though its minimization procedure is often blamed for its poor speed of convergence and stability. Constant learning rates throughout the procedure compound the problem of slower learning rate. But, because of the greater flexibility they offer, one can restrict some user-specified parameters such as momentum (a positive constant to increase the rate of convergence by minimizing the oscillations in the weights space during training) or learning rate to guarantee convergence. Unfortunately, there are no easy nor automatic ways to specify the initial conditions, and even if there are, such moderate reforms do not provide the thrusts to accelerate the rate of convergence significantly.

Furthermore, initial conditions are critical to numerical problems based on stochastic processes; standard guidelines do exist in the literature that one can choose from to set some of the initial conditions (e.g., the Nguyen and Widrow's weight initialization algorithm [Ham and Kostanic, 2001]). From what was discussed earlier, the backprop procedure is based on one such format that relies on its initial conditions. It is the single most popular algorithm that provides a wide range of controls (e.g., initial weights, learning mode, learning rate, stopping criterion) to the user prior to a training process; the user also gets the choice over the weight-search algorithms. However, the gradient-search directions in the weight space are performed using defined learning rules.

There exists a range of techniques in the literature for supervised learning which were primarily motivated by the slow rate of convergence observed in the backprop. Second-order optimization techniques such as the conjugate gradient or Newtonian method offers significant improvement to the speed of convergence while being applicable to large-scale problems [Haykin, 1999]. Typically, the rate of convergence is quadratic and at least an order of magnitude faster than the gradient search. During the course of this dissertation, I had the opportunity to experiment with both these classes before eventually settling down with a flavor of the quasi-Newton method called the Levenberg-Marquardt (LM) technique upon which my entire work is based [Ham and Kostanic, 2001]. The fundamental idea behind the LM technique can be derived from the conjugate gradient algorithm, a combination of the method of steepest descent and the Newton's method, described below.

2.4 Conjugate Gradient Algorithm

A superior technique, compatible with the backprop procedure, useful to accelerate the learning rate and suitable for any large-scale problems is the conjugate gradient algorithm, hereafter called the CG algorithm. A CG method performs a unique gradient search in the weight space utilizing second-order derivatives for the cost function minimization (or the Hessian); the inclusion of the Hessian in the algorithm could be memory expensive, but its computation can be overcome through some implicit means which may be fast and inexpensive, through numerical approximations instead of first principles. As opposed to the gradient-descent method which uses an instantaneous estimate of the gradient of the error function, the CG algorithm uses the knowledge of the eigenvalues and *eigenvectors* of the Hessian at a given point of time in the weight space. The new search direction is a linear combination of all

the previous search directions while the learning rate is a function of the *eigenvalues*. Obviously, the new search direction is always non-interfering, and the weights are, therefore driven from an arbitrary state to the minimum of the weight space in a finite amount of steps and along those directions.

2.4.1 Newton's Method

As mentioned earlier, any explicit way to calculate the Hessian could be computationally taxing, and in some large-scale problems it could even prove practically impossible. Nonetheless, the prescription for the Newton's method aims to bring out the role of the Hessian in any Newton, quasi-Newton, or CG-based algorithms as they attempt to find the directions of fastest error descent by computing the Hessian through explicit or implicit means.

Consider a quadratic form of the error function as the one described in (2.3) where $E(w) = E(w_1, w_2, \dots, w_n)$ is a function of all the available weights in the network. For a small change going from w to w^* , we can write $E(w)$ through a Taylor series as

$$E(w) = E(w^*) + (w - w^*)^T \nabla E|_{w=w^*} + \frac{1}{2}(w - w^*)^T H|_{w=w^*} (w - w^*) + O(w^3) \quad (2.13)$$

where H is the Hessian given by

$$H_{ij} = \frac{\partial^2 E}{\partial w_i \partial w_j} \quad (2.14)$$

with $\nabla E=0$ at the minimum where $w = w^*$. Ignoring third and higher order terms, (2.13) becomes

$$\Delta E = E(w) - E(w^*) \approx \frac{1}{2}(w - w^*)^T H|_{w=w^*} (w - w^*) \quad (2.15)$$

If we describe a set of eigenvectors (c_i) for the Hessian, H obeying $H_i c_i = \lambda_i c_i$, then $\Delta w^* = w - w^* = \sum_i \alpha_i c_i$, where α_i are some coefficients determined by the Δw^* and

the c_i . Then (2.15) becomes

$$\Delta E = \frac{1}{2} \sum_i \lambda_i \alpha_i^2 \quad (2.16)$$

implying that the change of error is largest along the directions where the eigenvalues are greatest. Further, from (2.13),

$$\nabla E = \sum_i \alpha_i \lambda_i c_i \quad (2.17)$$

Recall the weight update rule from (2.5) and (2.6),

$$\Delta w = -\gamma \nabla E = -\gamma \sum_i \alpha_i \lambda_i c_i \quad (2.18)$$

The weights are adjusted in accordance with the above equation until α_i approaches close to 0 as w proceeds toward w^* . In other words, the rate of convergence will be decided by the eigenvalues computed through the explicit knowledge of E and H , with a constant step size γ . This procedure lays the foundation for the CG algorithm which in and of itself is the Newton's method for quadratic $E(w)$.

2.4.2 Conjugate Directions

This method proposes to find a set of conjugate vectors $\{d_1, d_2, d_3, \dots, d_n\}$ in non-interfering directions, i.e., directions that do not “zig-zag” or point to the old search direction (section 2.3.1 earlier), corresponding to each weight vector in the network such that

$$\Delta w = w_1 - w^* = \sum_j^n \alpha_j d_j \quad (2.19)$$

where w_1 is some arbitrary starting point; the weight update rule is analogous to that of Newton's method. α_i is any scalar that minimizes the cost function, typically evaluated from a line search and can be computed without the explicit knowledge of the Hessian, provided the cost function is quadratic. The weight vector at each

iteration is a linear combination of all the previous search directions and the weight updates are parallel to the conjugate directions d_j . We can write every successive conjugate direction as a linear combination of the previous conjugate direction as

$$d_{j+1} = -g_{j+1} + \beta_j d_j \quad (2.20)$$

where β_j denotes a scaling factor and g_{j+1} denotes the old search direction. Again, the scaling factor can be computed without calculating the Hessian explicitly using the Polak-Ribière or the Fletcher-Reeves formula [Ham and Kostanic, 2001].

Thus, a slight modification of the backprop procedure to include the CG procedure would, in principle, minimize the cost function, assuming it is quadratic, guaranteed to converge in a small number of steps (\leq number of weights in the network), thereby greatly accelerating the rate of convergence.

2.4.3 The CG Algorithm

The following CG algorithm offers the proposed modification to the backprop procedure discussed earlier.

1. Initialize the weight vectors to some random values, w_1
2. Select a pattern and apply it to the network and compute the errors at the output and hidden layers using the backprop procedure, and compute the cost function.
3. Set the initial search direction $d_1 = -g_1$; $g_1 = \frac{\partial E}{\partial w}|_{w=w(1)}$
4. Update the weight vector through $w_{(j+1)} = w_{(j)} + \eta d_j$. Typically, η is estimated through a line search that minimizes the cost function. Terminate the algorithm when $\|d_j\|$ is sufficiently small.

5. Compute the new gradient from $g_{j+1} = \frac{\partial E}{\partial w}|_{w=w(j+1)}$
6. Compute new search direction from $d_{j+1} = -g_{j+1} + \beta_j d_j$
7. Calculate β_j , which has different forms. The form I chose is the Polak-Ribière formula given by $\beta_j = \frac{g_{j+1}^T(g_{j+1}-g_j)}{g_j^T g_j}$. This has the ability to “self-reset” when the search hits a flat surface.
8. Go to step 4 and repeat until it converges.

2.4.4 Levenberg-Marquardt Method

The Levenberg-Marquardt method (LM) is a highly reputable member of the family of Newtonian minimization techniques, especially known for its fast rate of convergence. It is an optimization technique that is very similar to the CG method in that both offer a quadratic rate of convergence without computing the Hessian explicitly. Here, like in CG, we approximate the Hessian using the Jacobian, J , which is a matrix comprising the first order derivatives of the network error with respect to the weights. The difficulty in implementing this technique on a NN MLP is the computation complexity it presents. However, with the help of sufficient processing resources, this can be implemented with ease. All my forecast models were trained using this quasi-Newton algorithm available from the *MATLAB Neural Network Toolbox of The MathWorks, Inc.*. The weights are updated according to:

$$w_{j+1} = w_j - H_j^{-1} g_j \quad (2.21)$$

with the Hessian approximated by $H \approx J^T J$, we can then rewrite (2.21) as

$$w_{j+1} = w_j - \alpha_j g_j, \quad (2.22)$$

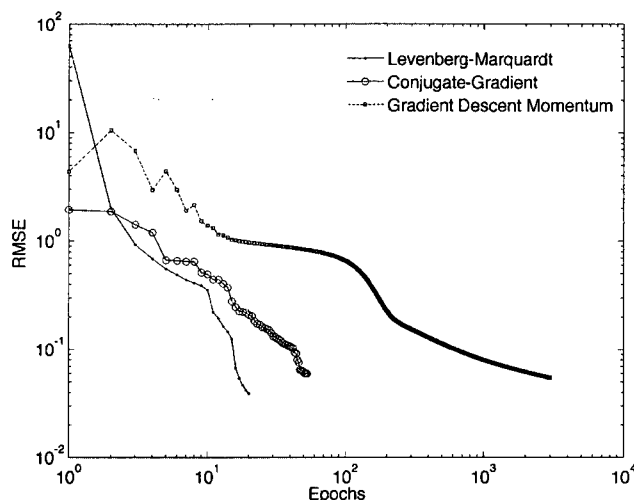


Figure 2.8 : A comparison of the speed of convergence using different methods.

reminiscent of the gradient descent method. Here, $\alpha_j = 1/\mu_j$, where μ_j is learning rate parameter. Computing the Hessian directly from first principles can be computationally challenging. Therefore, the simplest approach is to calculate the partial derivatives, the components of the Jacobian matrix and the resulting Hessian, using numerical approximation.

As an example, I use the function approximation problem ($y = \frac{1}{x}$; $x = [0.51, 3.50]$, $\Delta x = 0.01$) as a benchmark to compare the typical performance curves (figure 2.8) of all the techniques addressed in this chapter, and clearly, LM is the fastest by far. This is a single input-single output problem solved by 4 neurons in the hidden layer of a *tanh* function and 1 neuron in the output layer of a linear function. We refer again to equation 2.4 for the cost function. The data is subdivided randomly into three equal parts for training, testing and model validation. Figure 2.9 shows the learning histories, performed using the LM method, on both the training (smooth curve) and test (dashed curve) set as the cost function is minimized. The training continues

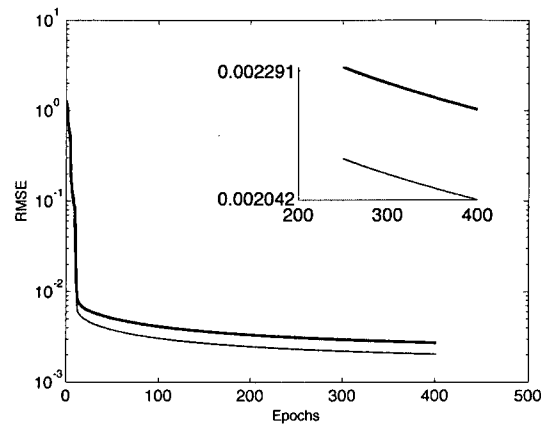


Figure 2.9 : Learning history for the function approximation problem performed using the LM routine. Training RMSE (thin curve) is shown along with the test RMSE (thick curve).

until it reaches its maximum allowed time of 400 epochs (1 epoch = 100 training patterns before each weight update), at which point the RMSE on the training data approaches the desired precision of ≈ 0.002 ; the results are reproducible with different sets of random samples. Note that here, in order to look closer at the time evolution of the training histories, the RMSE was allowed to reach a much higher precision from that shown in figure 2.8.

We now turn our attention to building a research methodology aimed at exploiting the properties of the feedforward neural architecture discussed here.

Chapter 3

Research Methodology

This chapter broadly describes the problem tackled, approach taken and the data needed to develop the proposed prediction models. It is divided into three sections. The first section will be a description of the data gathering efforts, specifically outlining the satellites and instruments providing the data. The second section introduces the research methodology and raises hypothetical questions in the hope of finding an optimal solution. The last section describes the key terms and statistics used to validate and verify the models.

3.1 Data and Instruments

In this study, I used two kinds of data: (1) solar wind velocity and interplanetary magnetic field (IMF), and (2) geomagnetic indices: Kp, Dst and AE, which were gathered from different observatories around the world. The Boyle Indices (BI's) used in this study were derived using archived data from ACE. To stretch the timeline, I also used solar wind and IMF data from WIND and IMP-8 (discussed in succeeding subsections) respectively to derive the BIs [Boyle et al., 1997]. Thus, the total available time line of observations spanning 13 years enabled me to train the network over a complete solar cycle. Though it was easy and straightforward to derive the BIs, data gaps and other undesirable features caused hiccups to an otherwise continuous data stream. Therefore, a thorough inspection was needed before the data could be used effectively.

3.1.1 WIND

WIND was launched in 1994, and one of its mission goals is to provide complete plasma, energetic particle, and magnetic field input for magnetospheric and ionospheric studies. Its data is maintained by the MIT Space Plasma Group and is publicly available from http://ftpbrowser.gsfc.nasa.gov/wind_swe_2m.html. I used three years of solar wind data from WIND for this work (1995-1997).

3.1.2 IMP-8

IMP-8 or IMP-J was one of the oldest-serving satellites for the space physics community. Launched in 1973, it was designed to measure magnetic fields, plasmas, and energetic charged particles of the Earth's magnetotail and magnetosheath and of the near-Earth solar wind. The IMP-8 data is also publicly available from the Space Plasma Group at the MIT (<ftp://space.mit.edu/pub/plasma/imp/www/imp.html>). The data I used from IMP-8 (1995-1997) complemented the data obtained from WIND.

3.1.3 ACE

ACE stands for Advanced Composition Explorer. It was launched in 1997 with nine science instruments including six high resolution spectrometers to measure the elemental, isotopic, and ionic charge state composition of nuclei from H to Ni originating from solar photospheric and coronal material transported to 1 AU and other sources elsewhere in the galaxy. One of the primary mission objectives of ACE is to provide continuous measurements of the solar wind, low energy solar and interplanetary particles, and cosmic rays, requiring an orbit outside the Earth's magnetosphere. The modified halo orbit about the Sun-Earth system's libration point, L1 meets this requirement [Stone et al., 1998].

ACE level 2 data products (<http://www.srl.caltech.edu/ACE/ASC/level2/index.html>) of SWEFAM (solar wind plasma) and MAG (magnetic field) gave an uninterrupted data set to derive the BI. Unlike WIND and IMP-8, ACE is strategically placed at the L1 point to constantly monitor the upstream solar wind conditions. Assuming an average solar wind velocity of 400 km/sec, this position allows a 45 minutes to an hour lead time before the solar wind hits the Earth. I used the 1-minute data from ACE which was in turn used to derive 1 and 3 hour integrated-time averages; the data covers 1998 to 2007.

3.1.4 Kp

The official values of 3 hour averaged Kp were obtained from GeoForschungsZentrum (GFZ), Potsdam, Germany. They have been maintaining records that date back to 1932 and are publicly available. Unlike the solar wind and IMF data, the official Kp record is fairly uninterrupted. For my analysis, I used the Kp data from 1995 to 2007.

Since Kp is a three-hour index, we construct a new 1-hour cadenced time series using quadratic interpolation through splines, a technique preferred over oversampling. These newly constructed averages will henceforth be denoted as the measured “1-hour” Kp index. For example, corresponding to a certain three hour period say between 0600 and 0900 UT, centered at 0730, the new 1-hour Kp index will have points centered at 0630, 0730 and 0830. In the following discussions, Kp* is the ANN predicted Kp.

3.1.5 Dst and AE

The Data Analysis Center for Geomagnetism and Space Magnetism, World Data Center for Geomagnetism, Kyoto provides AE and Dst in real time or “quick-look” (<http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html>); it also maintains the official

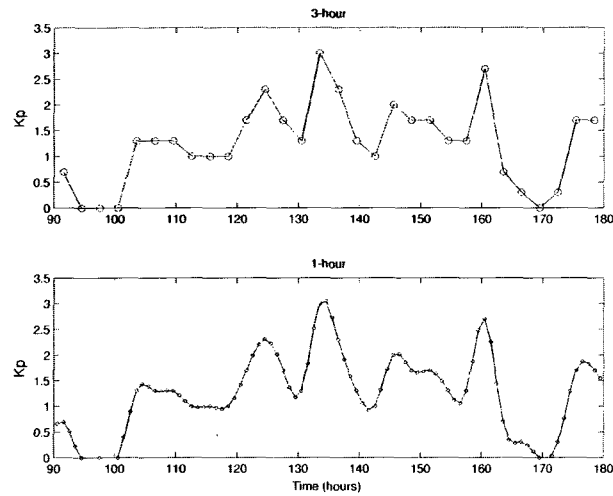


Figure 3.1 : Shown here are the 3-hr Kp and 1-hour Kp smoothed quadratically.

archived Dst and AE data.

3.1.6 Real-time data

Finally, for our operational needs, the Space Weather Prediction Center (SWPC) of the National Weather Service (<http://swpc.noaa.gov/>) provides regular updates of planetary Kp (3 hour average) or Kp proxies (see chapter 1). As far the solar wind and IMF is concerned, the online data is available from <http://www.swpc.noaa.gov/ace/>. Data in Level 2 format is downloaded at regular intervals to the Rice Space Institute's (RSI) website to feed the real time models. The data summary is shown in table 3.1

3.2 Problem Definition

The goal of this dissertation is to design a set of computer algorithms with good prediction capabilities for moderate to severe storms and to forecast Kp, Dst and AE

Table 3.1 : Summary of Data

Source	Type	Span	Average resolution
ACE	Solar wind + IMF (BI)	1998-2007	1 hour
WIND	Solar wind (BI)	1995-1997	1 hour
IMP-8	IMF (BI)	1995-1997	1 hour
GFZ	Kp	1995-2007	3 hour
OMNIWeb	Dst	1995-2007	1 hour
OMNIWeb	AE	1995-2007	1 hour

up to 3 hours ahead in near-real time, and further to be significantly better than the existing models. To achieve this goal, I primarily focus on training and validating an ANN, using the time histories of the BI and the corresponding target indices or the BI on its own. Here, I have chosen the standard multilayered backpropagation network for the purposes of training, validation and testing, particularly adapting Levenberg-Marquardt, a flavor of conjugate gradient algorithm, for training [Haykin, 1999].

Given the ANN's success in short term space weather forecasting, as mentioned in chapter 1, I propose to investigate the following as a feasible research methodology:

- (1) to predict Kp with a lead time of 1 hour from solar wind only,
- (2) to predict Kp with a lead time of 3 hours from solar wind only,
- (3) to predict Kp with a lead time of 1 hour from solar wind and previous time history of Kp,
- (4) to predict Kp with a lead time of 3 hours from solar wind and previous time

history of Kp,

- (5) to predict Dst with a lead time of 1 hour from solar wind only,
 - (6) to predict Dst with a lead time of 3 hours from solar wind only,
 - (7) to predict AE with a lead time of 1 hour from solar wind only,
- and (8) to predict AE with a lead time of 3 hours from solar wind only.

3.3 Hypotheses to Test

1. I will investigate the effectiveness of the BI in raising the baseline further in short-term (3 hours or less) geomagnetic activity index forecasting. I will further investigate whether legitimate forecasting is plausible beyond lead times of over 3 hours, perhaps up to 6 hours.
2. Given that the new algorithms developed will be quantified and tested against historic data for new performance standards, I will investigate the feasibility of issuing probabilistic forecasts so that key discriminator levels can be stated with well defined confidence limits.

3.4 Proposed Models

Part of the job of a space weather forecaster is to aid the process of decision making during critical times and to make the available space weather information more useful. In order to achieve this goal, one would want to forecast at different cadences. Here, I choose to forecast at Kp, Dst and AE at 1-hour and 3-hour cadences. Recall, Kp is a 3-hour index while AE and Dst are hourly indices. The purpose of a 1-hour cadenced prediction is to provide a short-term warning to certain end-users for whom such a warning can be of benefit; for example, satellite operators, electrical transmission line companies, and airlines with polar routes. A 3-hour predicted AE and Dst will

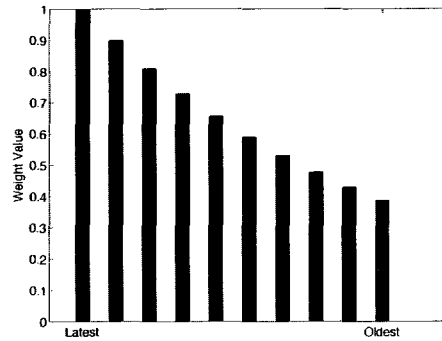


Figure 3.2 : Each input to the ANN is weighted according to equation 3.1. The most recent value of the solar wind/target index is given the maximum weight of unity.

be regarded as a new baseline for further studies since most of the best estimates available in the literature are delivered for the next two hours from the predicted time.

3.4.1 Training Parameters

In addition to the basic ANN parameters we saw earlier, some of the key terms and operational definitions that will also help interpret the recipe behind the models are defined here.

Input Weights

Inputs to the ANN architecture will follow a simple rule that the most recent value of the BI and Kp receives the maximum weight (w) of unity, the second most recent value receiving 0.9 ($0 < w < 1$) of the previous, and so on (w^n , $n = 0, 1, 2, \dots$) [equation 3.1 and figure 3.2]. This procedure is an arbitrary choice. The weighting scheme explicitly:

$$\text{Input weights} = [1.0 \ 0.90 \ 0.81 \ 0.73 \ 0.63 \ 0.59 \ 0.53 \ 0.48 \ 0.43 \ 0.39] \quad (3.1)$$

The rationale behind this scheme of weighting the inputs is based on constructing models aimed at having the most recent solar wind values contribute the most to the predicted results. We will revisit (chapter 5) this weighting scheme after further analysis.

Error Measure

As a reminder, a couple of key terms related to network optimization, and which are useful to track the time evolution of the network training error, are the RMSE (Root Mean Square Error) and the ARE (Average Relative Error), computed at the end of each epoch before the weights are updated. They are given by:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (target_t - output_t)^2} \quad (3.2)$$

$$ARE = \frac{1}{N} \sum_{t=1}^N \frac{|target_t - output_t|}{|target_t|} \quad (3.3)$$

where $target_t$ is the desired value and $output_t$ is the measured value at the output layer. N is the total number of training samples. Testing the quality of the final result will not be based just on the RMSE error. However, initial network diagnostics and pruning can be done through observing the learning curves based on monitoring the RMSE.

Data for Training

Copious data are available for this study, covering more than a full solar cycle. A 3-hour average means having over 30000 points covering almost 100000 hours worth of data. Therefore, a good approach to building a steady network is to judiciously stratify the data into three smaller but distinct samples, one for training, one for

validating the error and the other for testing the model. The inputs and the outputs to the ANN are both normalized to fall in the range $[-1, 1]$. An ideal network is one which can generalize “unseen” data that is not represented in the training. The models will be trained, tested and validated from a random selection of samples. All models using hourly cadenced dataset will have approximately 47% of the total data participating in training and approximately 36% and 17% will be reserved for testing and validation purposes respectively. Similarly, models running on 3-hour cadence will have approximately 59%, 22% and 19% of the total data reserved for training, testing and validation respectively. Furthermore, while using the Kp index as inputs, the most recent Kp value is not used deliberately keeping in mind that there will be delays in obtaining the “nowcasted” Kps during real time operations, and we do not want the algorithm to depend on a value which is not available in real time.

All the models use a “hyperbolic-tangent” transfer function in their hidden layer and a “linear” transfer function in their output layer, and employs the LM minimization technique for training. The number of neurons in the output layer is held constant at 1, because this is a time-prediction problem. Furthermore, one “epoch” is defined as the time between any two successive weight updates corresponding to a batch size or sample size of the training set. However, the number of neurons in the hidden layer and learning rate will be decided based on the training performance using trial and error.

3.4.2 Model 1: Kp prediction with 1-hour lead time using only the BI

We seek to find the optimum look-back time for the best predictive power i.e., how much of the solar wind history is needed in order to get the best prediction. We apply the standard method of a “sliding window” procedure to the feedforward network to get the best predictive estimate. This procedure is represented by equation 3.4, i.e.,

given the values of t , $t-1$, $t-2$, ..., $t-n$, we want to forecast $t+1$. In other words, we want to determine the value of the optimum n in equation 3.4. The following function is the generic form of the exploratory tests performed, whose parameters are summarized in table 3.2.

$$Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 0, 1, \dots, 9. \quad (3.4)$$

' t ' represents the epoch in question while ' $t-1$ ', ' $t+1$ ' means 1 hour behind and 1 hour ahead of t respectively. Kp_{t+1}^* is the forecasted value, and each BI_t , BI_{t-1} etc are hourly averages of the BI. Figure 3.3 is a schematic representation of the model architecture, shown here for the case $n = 6$. The training set (input-output pairs) were chosen randomly. The LM routine adopts batch processing, wherein the weights are updated only after all the elements in the batch, defined by the number of training samples, have been propagated to the output layer. Figures 3.4 and 3.5 show their learning histories based on the RMSE error for both training (black) and test (blue) data set. Based on the training histories, we can see that the best function is possibly $n = 8, 9$ or 10 or equivalently 8, 9 or 10 hours of the BI history (lower RMSE and ARE for $Kp > 4$); any input time history beyond 10 hours fails to improve the performance. The optimum value of n can be given only following further analysis.

3.4.3 Model 2: Kp prediction with 3-hour lead time using only the BI

This model follows model 1 except that it predicts Kp 3 hours ahead and uses 3-hour averages of the BI, giving us the advantage of using longer time history in inputs. The ANN parameters are shown in table 3.3 and the learning histories in figures 3.7 and 3.8. The generic form is given by the following equation:

$$Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), n = 0, 1, \dots, 9. \quad (3.5)$$

Table 3.2 : Predict Kp_{t+1}^* from BI.

#	Model 1 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test	ARE [†] Test
1	BI_t	1	58	0.01	0.904	0.907	0.247
2	BI_t, BI_{t-1}	2	58	0.01	0.803	0.803	0.205
3	BI_t, BI_{t-1}, BI_{t-2}	3	58	0.01	0.742	0.753	0.188
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}$	4	58	0.01	0.719	0.732	0.176
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}$	5	58	0.01	0.716	0.722	0.179
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}$	6	58	0.01	0.712	0.723	0.181
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}$	7	60	0.01	0.700	0.710	0.173
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}$	8	60	0.01	0.702	0.712	0.174
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}$	9	62	0.01	0.692	0.710	0.172
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}$	10	62	0.01	0.687	0.709	0.171

[†] $Kp > 4$

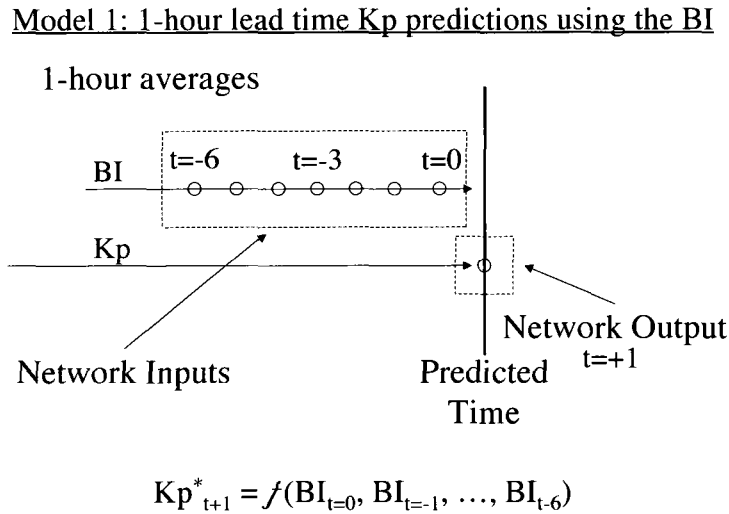


Figure 3.3 : Diagram showing an input vector and its corresponding output, constituting a pattern. Here $n=6$ is chosen for example.

Since this procedure adopts 3-hour binning it gives tremendous advantage in terms of the available time history of data before each prediction. The number of hidden nodes were adjusted based on trial and error and the learning rate has been lowered (50% of model 1), given the low volume of training samples here, to provide the same opportunity as model 1 received. Figure 3.6 shows the network inputs (7) and its corresponding desired output at the predicted time of $t = +3$, for case $n = 6$. The data has been organized by constructing 3-hour averages of the BI and Kp over the entire 11-year time length and the input-output pairs are constructed by applying the standard sliding window technique. The model has been designed by training ≈ 19000 , testing ≈ 7000 and validating ≈ 5000 input-output patterns chosen randomly from the 11-year period. It appears that the best function here has either $n = 7$ or 8 (21 or 24 hours of input history).

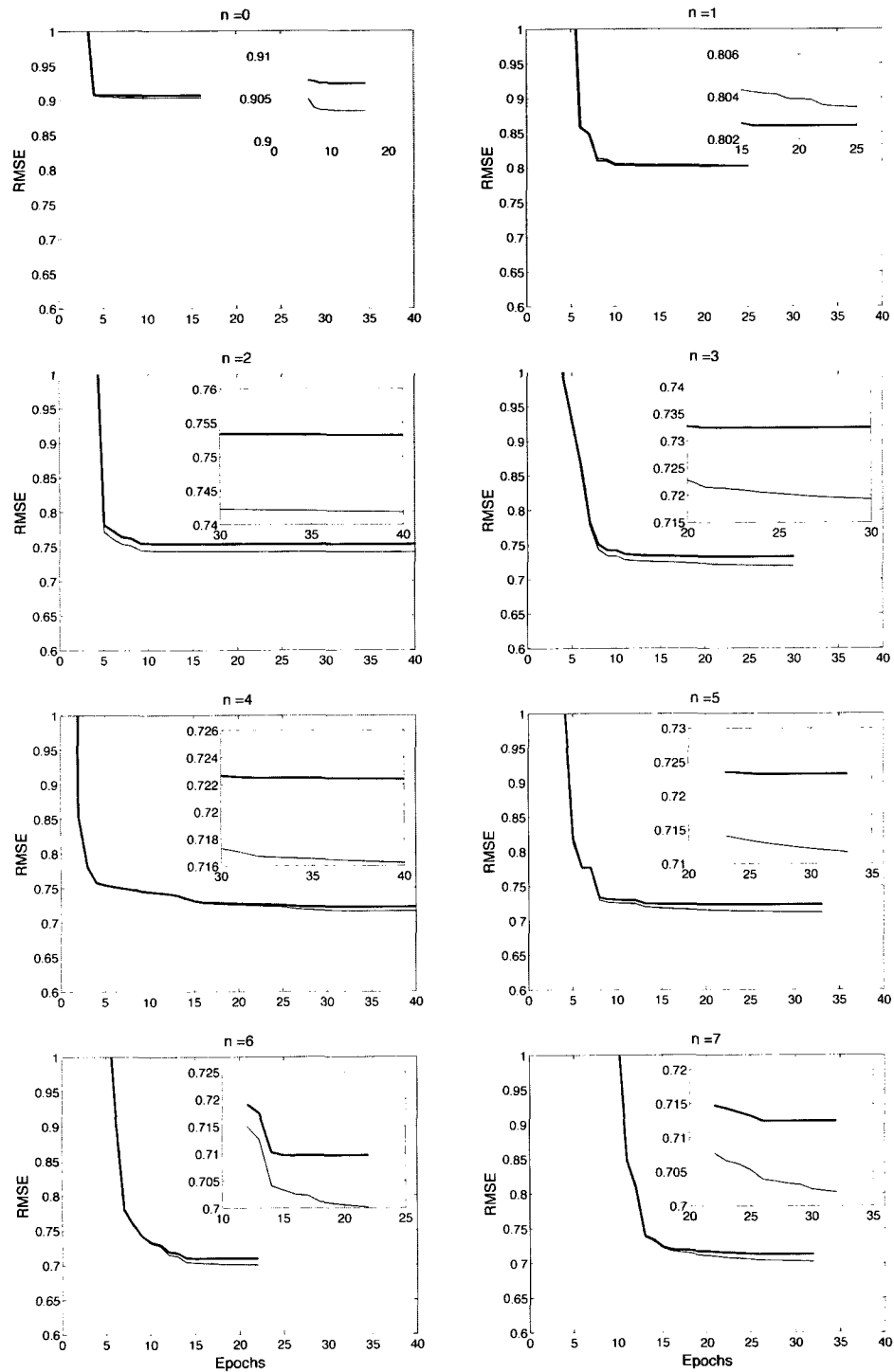


Figure 3.4 : Model 1 learning histories to predict Kp_{t+1}^* using hourly averages of BI, corresponding to table 3.2 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 0, 1, \dots, 7$.

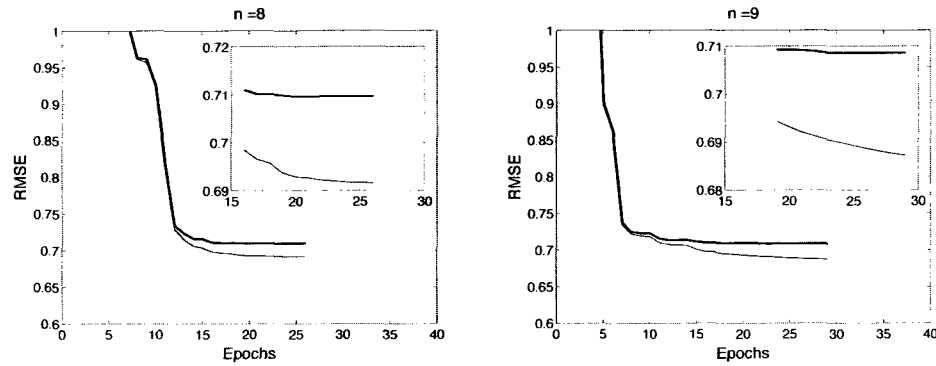


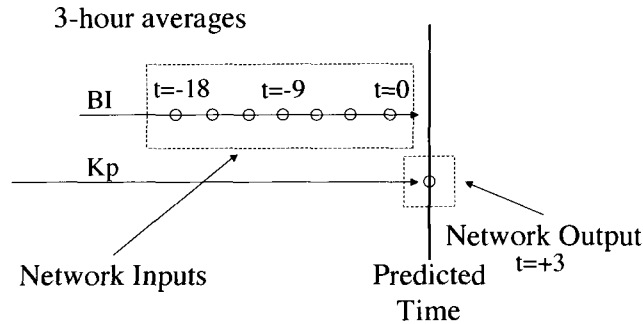
Figure 3.5 : Model 1 learning histories to predict Kp_{t+1}^* using hourly averages of BI, corresponding to table 3.2 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), n = 8, 9$.

Table 3.3 : Predict Kp_{t+3}^* from BI

#	Model 2 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test	ARE [†] Test
1	BI_t	1	48	0.005	0.846	0.854	0.219
2	BI_t, BI_{t-3}	2	48	0.005	0.834	0.848	0.222
3	BI_t, BI_{t-3}, BI_{t-6}	3	48	0.005	0.822	0.851	0.228
4	$BI_t, BI_{t-3}, \dots, BI_{t-9}$	4	48	0.005	0.818	0.823	0.213
5	$BI_t, BI_{t-3}, \dots, BI_{t-12}$	5	50	0.005	0.816	0.824	0.220
6	$BI_t, BI_{t-3}, \dots, BI_{t-15}$	6	52	0.005	0.806	0.826	0.218
7	$BI_t, BI_{t-3}, \dots, BI_{t-18}$	7	52	0.005	0.796	0.811	0.199
8	$BI_t, BI_{t-3}, \dots, BI_{t-21}$	8	52	0.005	0.796	0.825	0.213
9	$BI_t, BI_{t-3}, \dots, BI_{t-24}$	9	54	0.005	0.806	0.819	0.208
10	$BI_t, BI_{t-3}, \dots, BI_{t-27}$	10	54	0.005	0.789	0.829	0.209

[†] $Kp > 4$

Model 2: 3-hour lead time Kp predictions using the BI



$$Kp_{t+3}^* = f(BI_{t=0}, BI_{t=-3}, \dots, BI_{t=-18})$$

Figure 3.6 : Diagram showing an input vector and its corresponding output, constituting a labeled pattern. Here $n = 6$.

3.4.4 Model 3: Kp Prediction with 1-hour Lead-Time using the BI and Kp history

Here, I propose a slight variation to the models described above in that we introduce the time history of the target index to the network along with the solar wind data, i.e., for a set of BI in the input there is a corresponding set of Kp values lagged relative to the BIs. Predictions improve with the usage of the past history of Kp in inputs [e.g., Wing et al., 2005]. The general form of the tests performed is given in equation 3.6, the testing parameters in table 3.4 and the learning histories in figures 3.10 and 3.11. Since this is a 1-hour model, we use Kp values as described in section 3.1.4. Note that the history of Kp values are weighted, using the scheme applied for the BI, and lagged 3 hours relative to the BI because of non-availability of Kp at $t=0$ (and at $t = -1$ and $t = -2$ in some cases).

$$Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-n-3}), \quad n = 0, 1, \dots, 9 \quad (3.6)$$

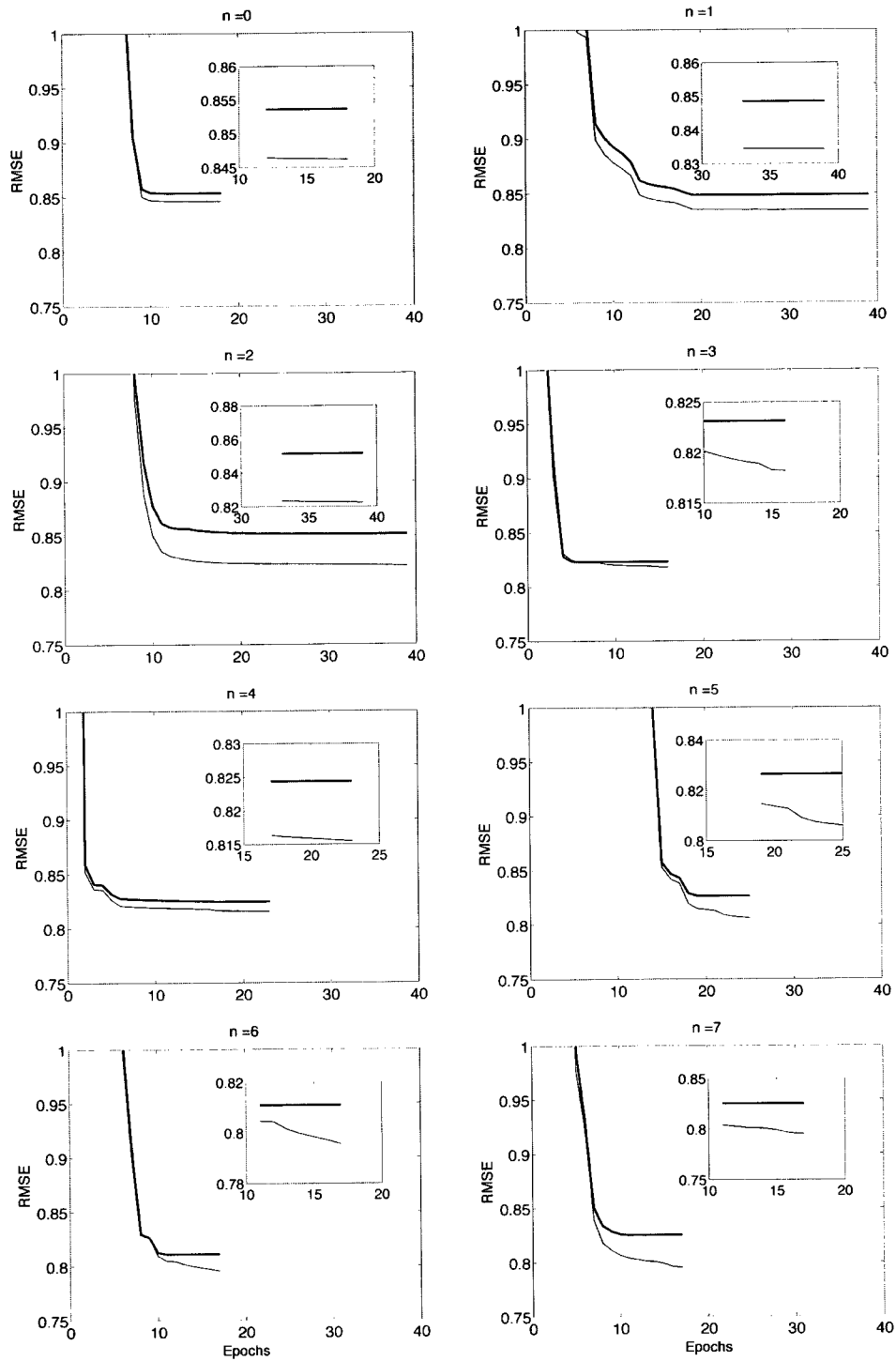


Figure 3.7 : Model 2 learning histories using 3-hour averages of the BI, corresponding to table 3.3 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3})$, $n = 0, 1, \dots, 7$.

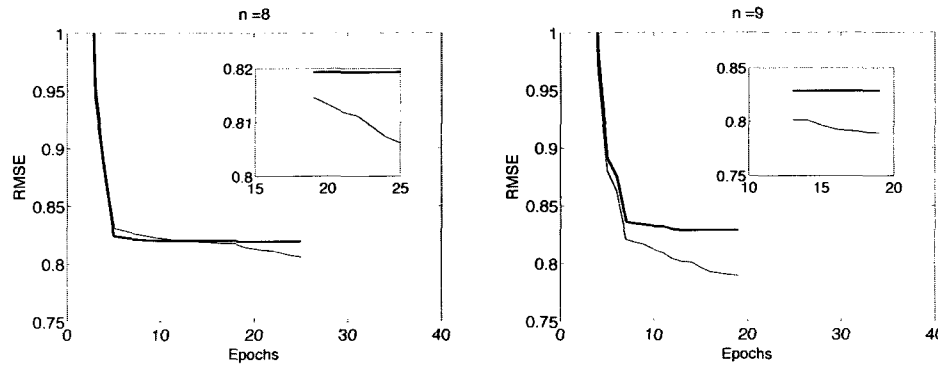


Figure 3.8 : Model 2 learning histories using 3-hour averages of the BI, corresponding to table 3.3 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3})$, $n = 8, 9$.

Training is conducted in the same manner as we would perform real-time operations and hence, Kp_t has been left out. In the training (and of course in the real-time predictions), I never use a known Kp index value to predict itself; if the previous hour's Kp index is not known, the prior measured Kp index is just duplicated. It is also worth noting that in order to provide greater credence to the models, network validation and testing were performed on two distinct subsets of the data. This, to a large degree, reflects the network's competency in generalizing new data and, to some degree, helps set a new performance standard against which any future models may be tested.

We use the official Kp values for model development and validation. However, for real time operational purposes, the time history of Kp will be 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, which can be downloaded from http://www.swpc.noaa.gov/rt_plots/kp_3d.html. There is, however, a 30-40 minute lag before the data is publicly available, mainly owing to processing delays but still usable within an hour. Therefore, we create the model

Model 3: 1 hour lead time Kp predictions using BI and Kp

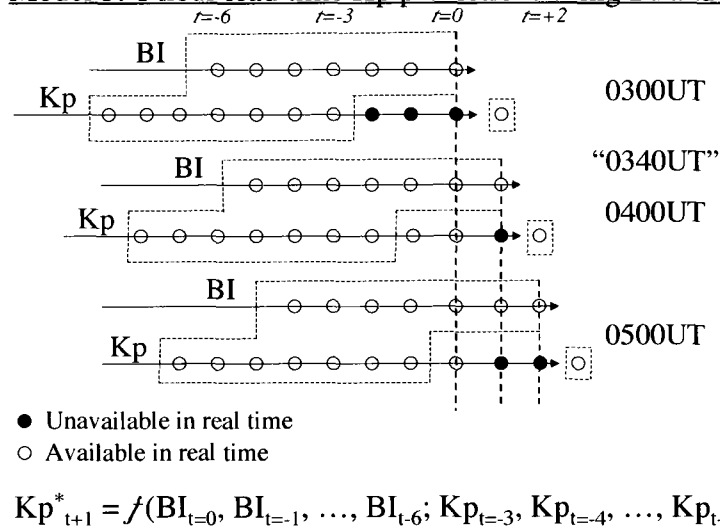


Figure 3.9 : Diagram showing an input vector and its corresponding output, constituting a pattern. Here $n = 6$.

keeping the real time operations in mind and exclude the last known Kp value both in training and retrospective analysis. Figure 3.9 illustrates this. For example, at 0300 UT the last value of Kp during the interval 0000-0300 UT is unknown in real time. At around 0340 UT, the last value of Kp becomes available. Yet, only two out of the three points are available at 0400 UT. Moving the window further by another hour to 0500 UT, only one out of three points is available. In retrospect, we therefore force the network to learn from what is available at that time and start at $t = -3$ instead of $t = 0$.

We pair the input vectors and its corresponding target as shown as “dotted window” in figure 3.9. For example, taking $n = 6$, we have a set of 7 BIs and 7 Kps as inputs to the network, and for which, the desired target is Kp_{t+3} . This constitutes a training pattern. We then slide the window, incrementing by 1 hour, to generate

the next set of training vectors, and the process is repeated sequentially over the entire volume of data. Training patterns are shuffled randomly and then chosen and designated for training ($\approx 47\%$), testing ($\approx 36\%$) and validation ($\approx 17\%$). This model consumes more training time because of the addition of Kp in the inputs. However, the results are much improved compared to model 1, as can be inferred from table 3.4 with remarkably low RMSE and ARE for $n = 6$.

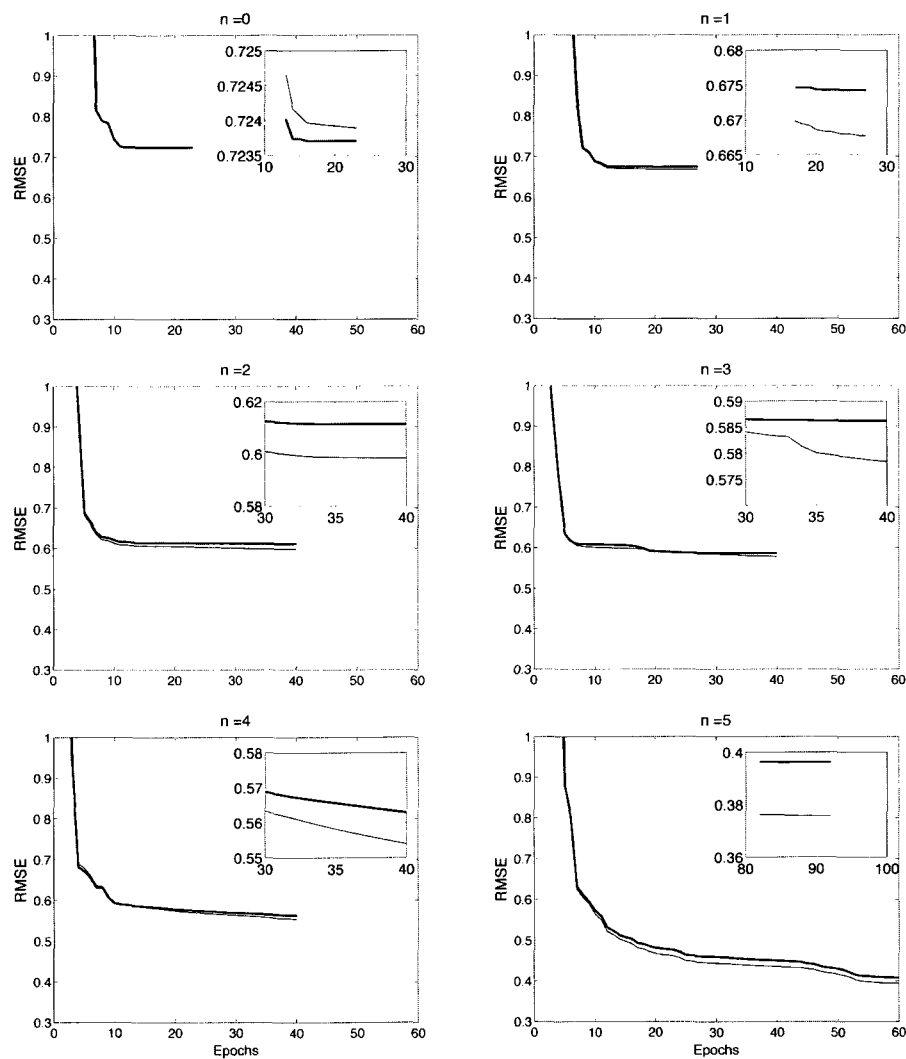


Figure 3.10 : Model 3 learning histories using hourly averages of BI and Kp, corresponding to table 3.4 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-n-3}), n = 1, 2, \dots, 5$.

Table 3.4 : Predict Kp_{t+1}^* from BI and Kp history

#	Model 3 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test	ARE [†] Test
1	$BI_t; Kp_{t-3}$	2	62	0.01	0.724	0.724	0.176
2	$BI_t, BI_{t-1}; Kp_{t-3}, Kp_{t-4}$	4	62	0.01	0.668	0.674	0.163
3	$BI_t, BI_{t-1}, BI_{t-2}; Kp_{t-3}, Kp_{t-4}, Kp_{t-5}$	6	62	0.01	0.598	0.611	0.145
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-6}$	8	60	0.01	0.578	0.586	0.135
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-7}$	10	60	0.01	0.554	0.563	0.124
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-8}$	12	62	0.01	0.376	0.396	0.077
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-9}$	14	64	0.01	0.517	0.545	0.119
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-10}$	16	66	0.01	0.537	0.555	0.124
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-11}$	18	66	0.01	0.445	0.493	0.098
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-12}$	20	68	0.01	0.520	0.565	0.124

[†] $Kp > 4$

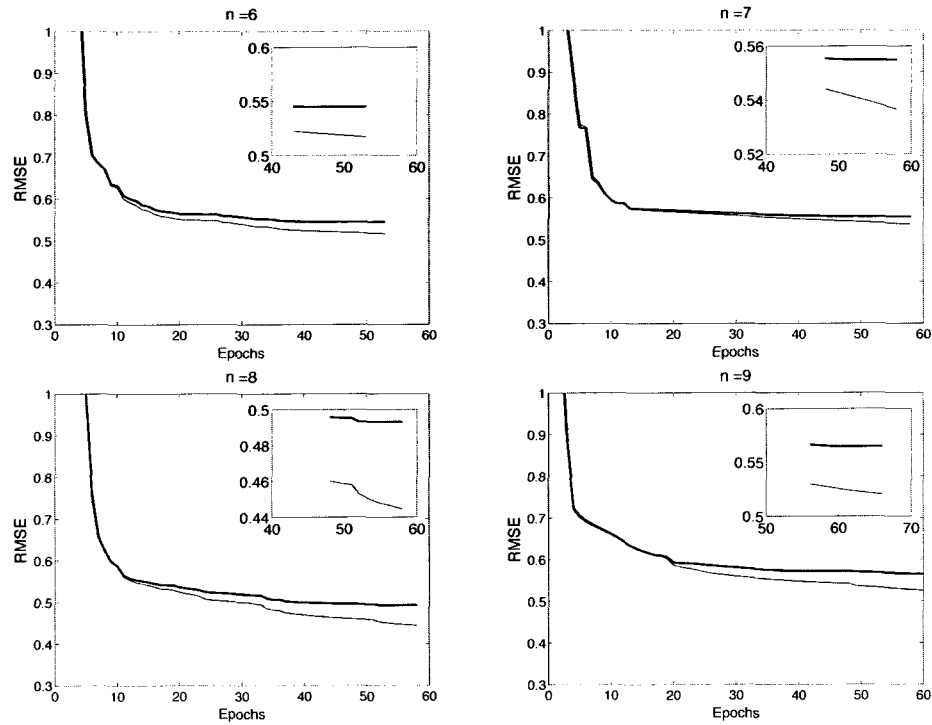


Figure 3.11 : Model 3 learning histories using hourly averages of BI and Kp, corresponding to table 3.4 (thick curve is the test data and thin curve is the training data). $Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-n-3})$, $n = 6, 7, 8, 9$.

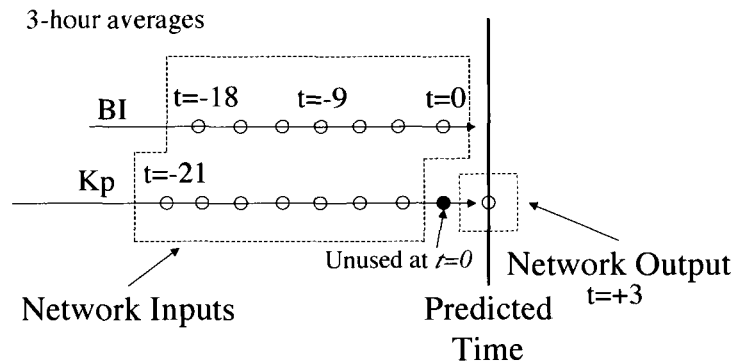
3.4.5 Model 4: Kp Prediction with 2-hour Lead-Time using the BI and Kp history

The proposed model 4 is similar to model 3 in the input sequence but for the 3-hour averages used here. Equation 3.7 describes the generic form of the experiments performed. Table 3.5 lists the model parameters and the learning histories are illustrated in figure 3.13 and 3.14. Just like in model 3, the history of Kp values are lagged 3 hours relative to the BI because Kp is not available at $t=0$ in real time. We apply the similar sliding window technique explained in model 3 for sampling the training vectors (figure 3.12).

$$Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-(n+1)*3}), \quad n = 0, 1, \dots, 9 \quad (3.7)$$

We do see a slight improvement in the results by bringing Kp time history in inputs. However, interestingly, longer time history of inputs does not guarantee a major improvement (RMSE and ARE for $n = 0$ and 9 are not significantly different). A clear distinction can be made after further statistical analysis through skill scores and hit/miss rate etc.

Model 4: 3-hour lead time Kp predictions using BI and Kp



$$Kp^*_{t+3} = f(BI_{t=0}, BI_{t=-3}, \dots, BI_{t=-18}; Kp_{t=-3}, Kp_{t=-6}, \dots, Kp_{t=-21})$$

Figure 3.12 : Diagram showing an input vector and its corresponding output, constituting a pattern. Here $n = 6$.

Different learning parameters were experimented in order to get the lowest RMS error possible on the test set. It appears that the uncertainties in the predictions increase with the lead time. The pick of the models and their validations will be discussed in detail in chapter 5. Since the Kp index is quantized (28 total bins), we will show the exact hit versus misses statistics through histograms. We will test the network performances using a couple of case studies involving geomagnetic storms. We will also examine the effect of persistence in models 3 and 4.

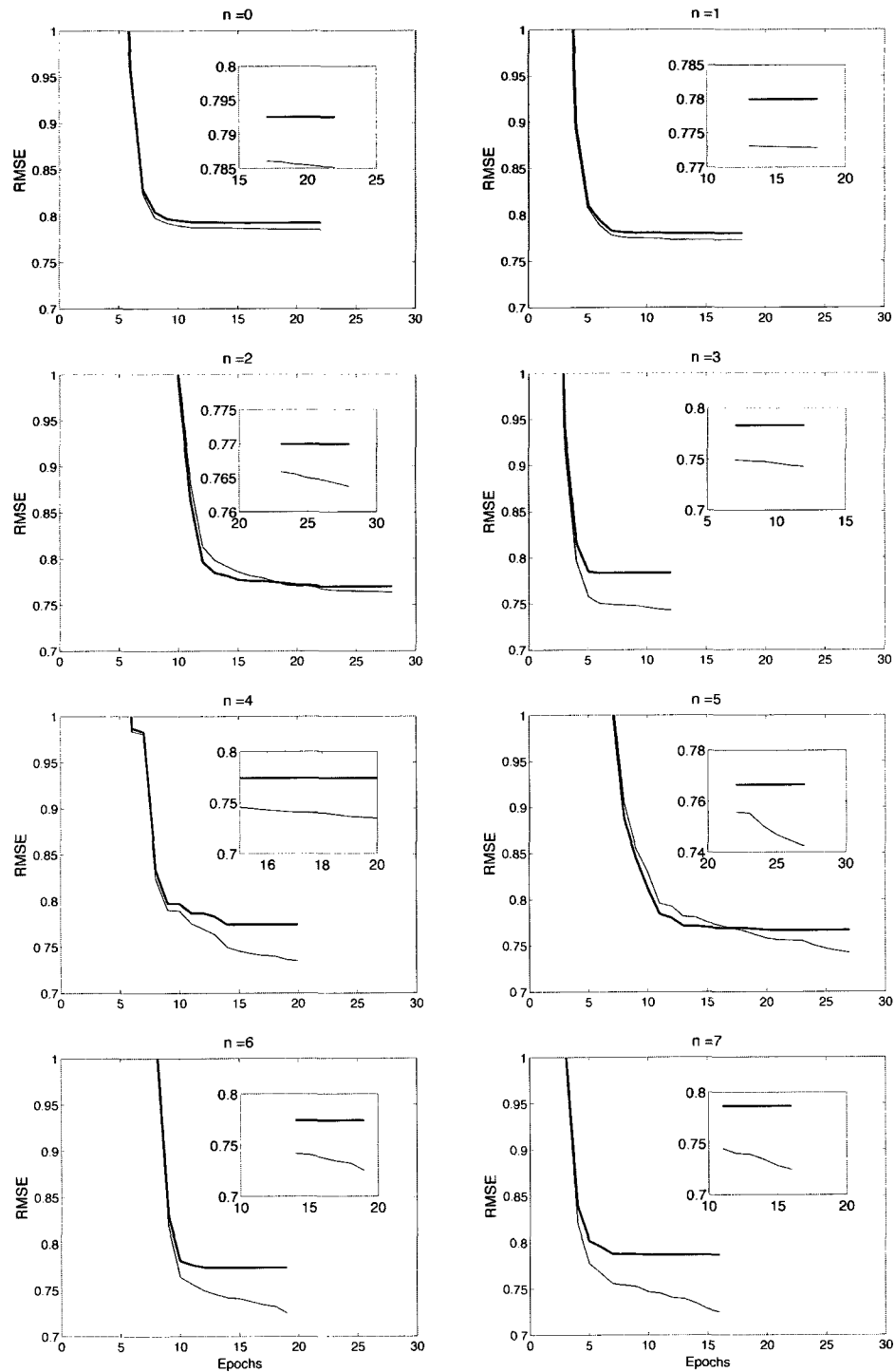


Figure 3.13 : Model 4 learning histories using 3-hour averages of BI and Kp, corresponding to table 3.5 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-(n+1)*3})$, $n = 0, 1, \dots, 7$.

Table 3.5 : Predict Kp_{t+3}^* from BI and Kp

#	Model 4 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test	ARE [†] Test
1	$BI_t; Kp_{t-3}$	2	61	0.005	0.785	0.793	0.204
2	$BI_t, BI_{t-3}; Kp_{t-3}, Kp_{t-6}$	4	61	0.005	0.773	0.778	0.195
3	$BI_t, BI_{t-3}, BI_{t-6}; Kp_{t-3}, Kp_{t-6}, Kp_{t-9}$	6	61	0.005	0.764	0.770	0.195
4	$BI_t, BI_{t-3}, \dots, BI_{t-9}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-12}$	8	61	0.005	0.743	0.783	0.205
5	$BI_t, BI_{t-3}, \dots, BI_{t-12}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-15}$	10	63	0.005	0.735	0.774	0.197
6	$BI_t, BI_{t-3}, \dots, BI_{t-15}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-18}$	12	63	0.005	0.742	0.767	0.199
7	$BI_t, BI_{t-3}, \dots, BI_{t-18}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-21}$	14	63	0.005	0.725	0.774	0.188
8	$BI_t, BI_{t-3}, \dots, BI_{t-21}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-24}$	16	65	0.005	0.725	0.787	0.202
9	$BI_t, BI_{t-3}, \dots, BI_{t-24}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-27}$	18	65	0.005	0.718	0.770	0.186
10	$BI_t, BI_{t-3}, \dots, BI_{t-27}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-30}$	20	65	0.005	0.701	0.778	0.190

[†] $Kp > 4$

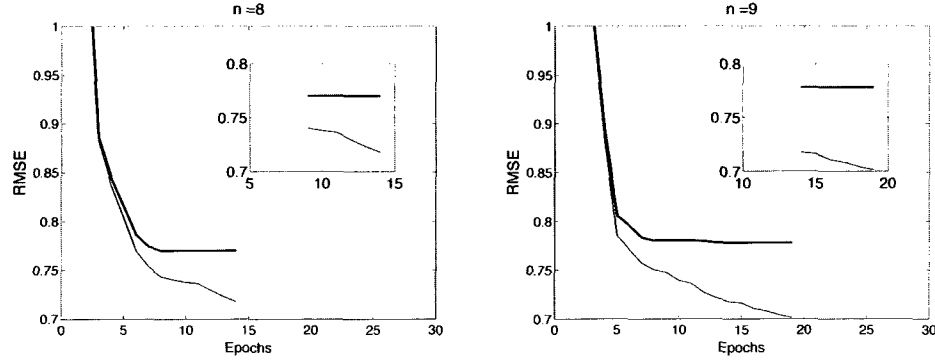


Figure 3.14 : Model 4 learning histories using 3-hour averages of BI and Kp, corresponding to table 3.5 (thick curve is the test data and thin curve is the training data). $Kp_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-(n+1)*3})$, $n = 8, 9$.

Furthermore, we will see how the new models can be implemented into the existing “spacalrt” system. The following subsections will discuss the proposed research methodologies to get the best functions for predicting the Dst and the AE index.

3.4.6 Model 5: 1-hour lead time Dst predictions from BI

This model is similar to the Kp model 1 in that it uses 1-hour averages of the BI but predicts Dst one hour ahead. We want to determine the optimum n in equation 3.8. The model inputs and the corresponding ANN training parameters are shown in table 3.6. Figures 3.15 and 3.16 show the ANN learning histories and are similar in trend to the Kp models. Here, the BIs are scaled linearly as opposed to the logarithmic scale used in the Kp models. However, the inputs are still weighted using equation (3.1).

$$Dst_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), \quad n = 0, 1, \dots, 9 \quad (3.8)$$

From table 3.6, we can see a clear trend emerging: as the number of inputs increase, the performance increases. The ARE threshold is set at $Dst < -40$ nT. It is also worth noting that, however, beyond 10 inputs the network performance fails to improve.

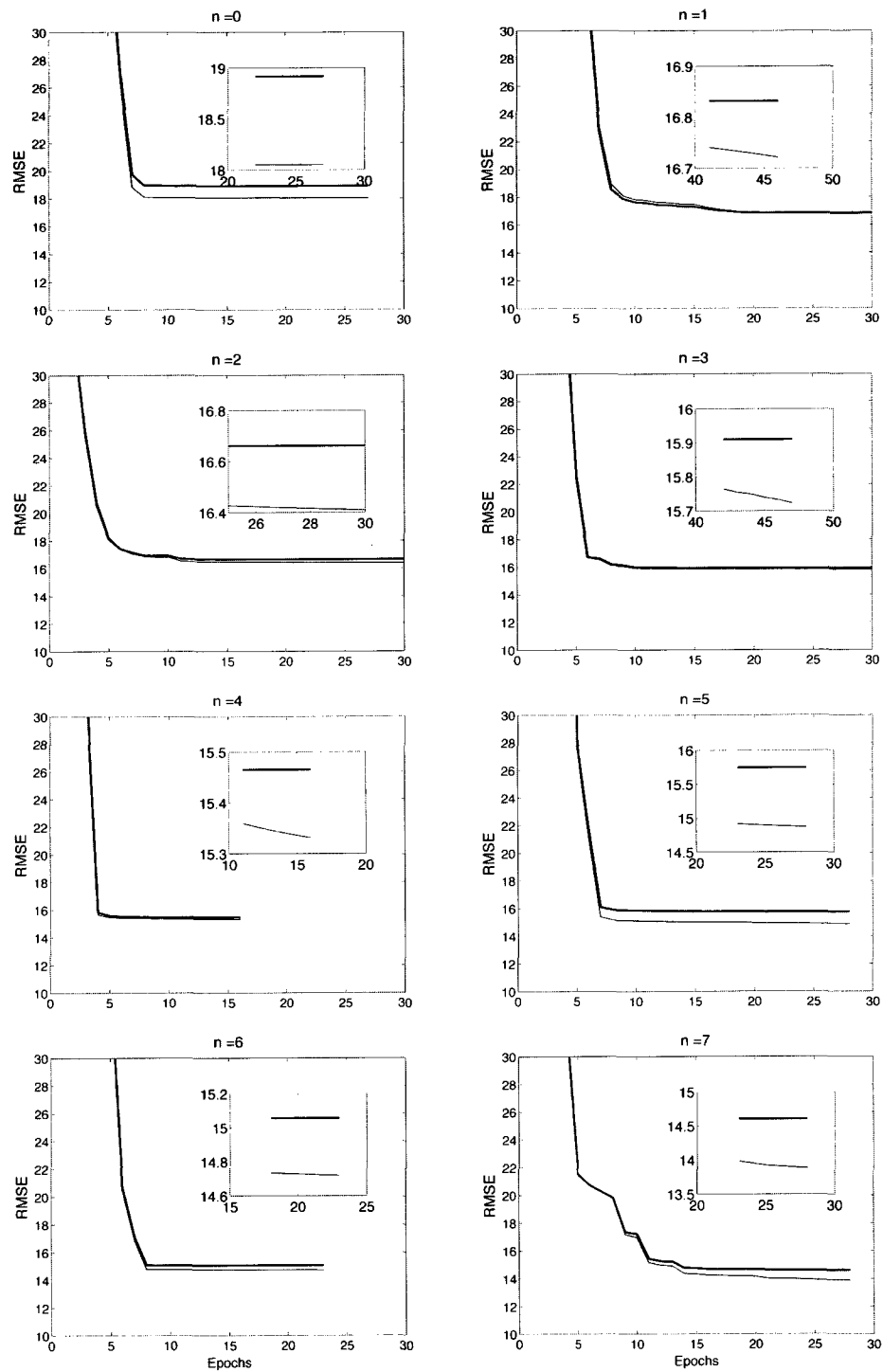


Figure 3.15 : Model 5 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $Dst_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n})$, $n = 0, 1, \dots, 7$.

Table 3.6 : Predict Dst_{t+1}^* from BI

#	Model 5 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test (nT)	ARE [†] Test
1	BI_t	1	59	0.05	18.05	18.91	0.491
2	BI_t, BI_{t-1}	2	59	0.05	16.72	16.83	0.454
3	BI_t, BI_{t-1}, BI_{t-2}	3	59	0.05	16.41	16.66	0.439
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}$	4	60	0.05	15.72	15.91	0.420
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}$	5	64	0.05	15.33	15.46	0.397
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}$	6	64	0.05	14.87	15.75	0.383
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}$	7	64	0.05	14.72	15.06	0.371
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}$	8	66	0.05	13.89	14.61	0.371
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}$	9	66	0.05	14.27	14.34	0.354
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}$	10	66	0.05	14.05	14.68	0.349

[†]Dst < -40 nT

3.4.7 Model 6: 3-hour lead time Dst predictions from BI

This model is similar to the Kp model 2, but predicts Dst 3 hours ahead. Since the final measured Dst are 1-hour averages, for this model, we construct 3-hour averages using the conventional 1-hour average. The model 6 inputs and the corresponding ANN training parameters are shown in table 3.7. Figures 3.17 and 3.18 shows the ANN learning histories. We want to determine the optimum n in equation 3.9 shown

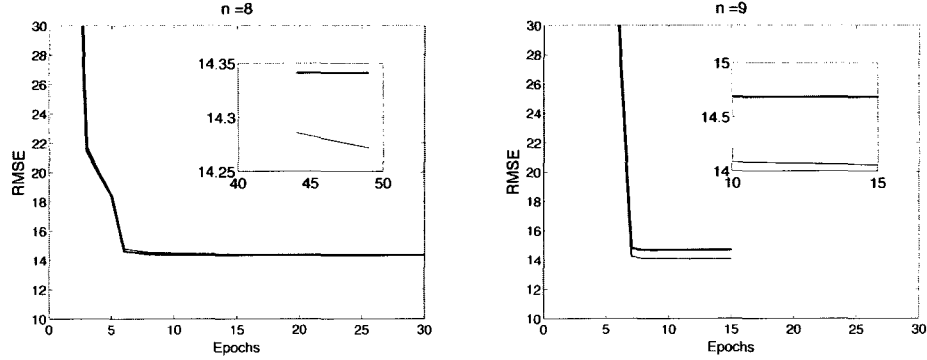


Figure 3.16 : Model 5 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $Dst_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n})$, $n = 8, 9$.

below

$$Dst_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), \quad n = 0, 1, \dots, 9 \quad (3.9)$$

Given the preliminary results, it is rather difficult to come to a conclusion about the best performing function. Interestingly, the overall RMSE (measured in nT) and ARE ($Dst < -40$ nT) are better here compared to the 1-hour Dst model.

3.4.8 Model 7: 1-hour lead time AE predictions from BI

This model is also similar to the Kp model 1 in that it uses 1-hour averages of the BI, but predicts AE one hour ahead. We want to determine n in equation 3.10.

$$AE_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n}), \quad n = 0, 1, \dots, 9 \quad (3.10)$$

The model inputs and the corresponding ANN training parameters are shown in table 3.8 with figures 3.19 and 3.20 showing the ANN learning histories. The RMSE values are large here. One possibility is that only the solar wind component of the AE index has been modeled well here.

Table 3.7 : Predict Dst_{t+3}^* from BI

#	Model 6 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test	ARE [†] Test
1	BI_t	1	48	0.001	15.47	15.75	0.463
2	$\text{BI}_t, \text{BI}_{t-3}$	2	48	0.001	14.31	14.63	0.379
3	$\text{BI}_t, \text{BI}_{t-3}, \text{BI}_{t-6}$	3	50	0.001	13.84	10.85	0.371
4	$\text{BI}_t, \text{BI}_{t-3}, \dots, \text{BI}_{t-9}$	4	50	0.001	13.14	13.92	0.383
5	$\text{BI}_t, \text{BI}_{t-3}, \dots, \text{BI}_{t-12}$	5	50	0.001	12.97	13.65	0.361
6	$\text{BI}_t, \text{BI}_{t-3}, \dots, \text{BI}_{t-15}$	6	52	0.001	12.84	12.87	0.380
7	$\text{BI}_t, \text{BI}_{t-3}, \dots, \text{BI}_{t-18}$	7	52	0.001	12.51	13.29	0.338
8	$\text{BI}_t, \text{BI}_{t-3}, \dots, \text{BI}_{t-21}$	8	52	0.001	12.29	13.02	0.321
9	$\text{BI}_t, \text{BI}_{t-3}, \dots, \text{BI}_{t-24}$	9	54	0.001	12.19	12.77	0.342
10	$\text{BI}_t, \text{BI}_{t-3}, \dots, \text{BI}_{t-27}$	10	54	0.001	11.73	12.83	0.338

[†]Dst < -40 nT

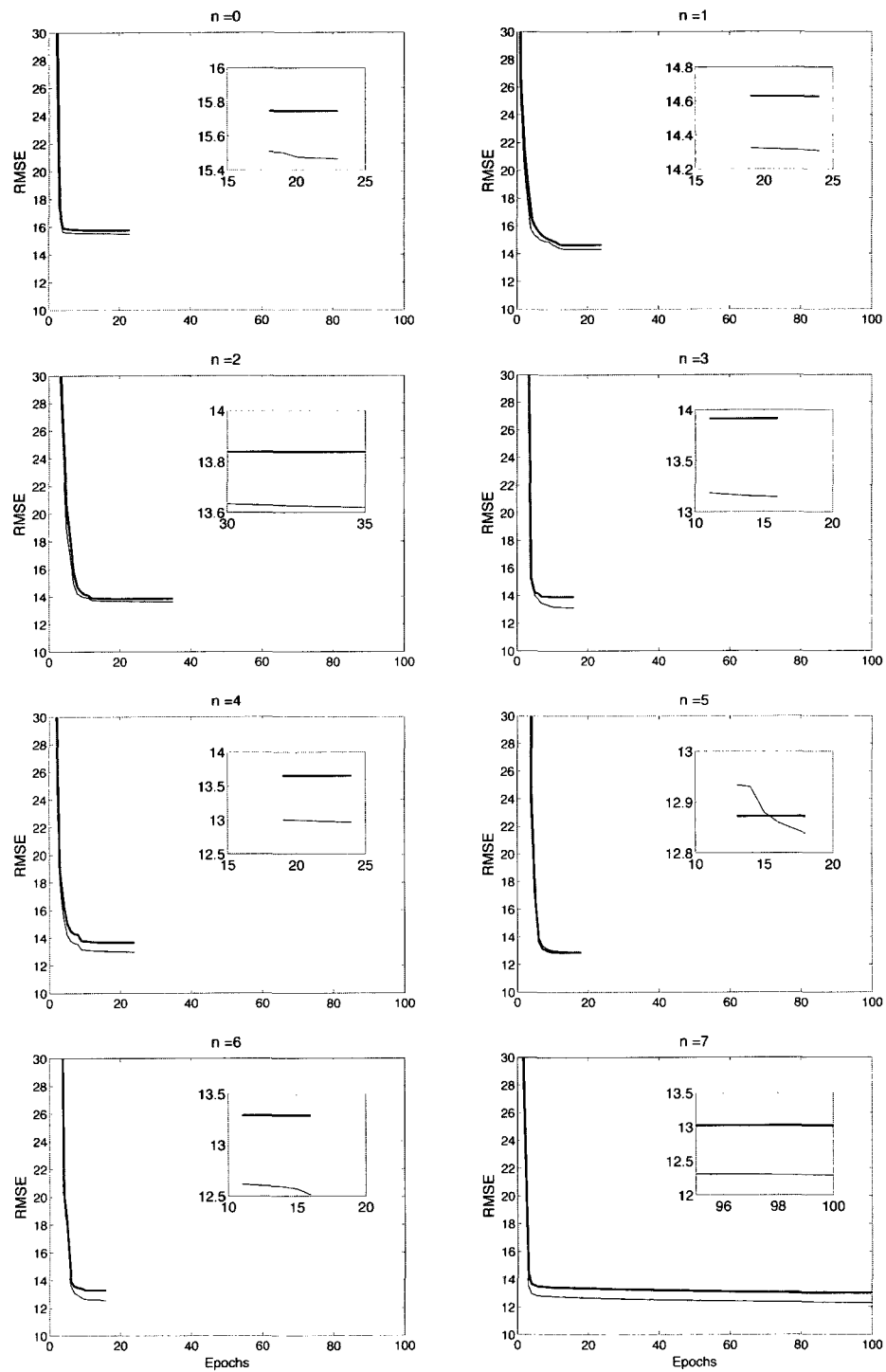


Figure 3.17 : Model 6 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $Dst_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3})$, $n = 0, 1, \dots, 7$.

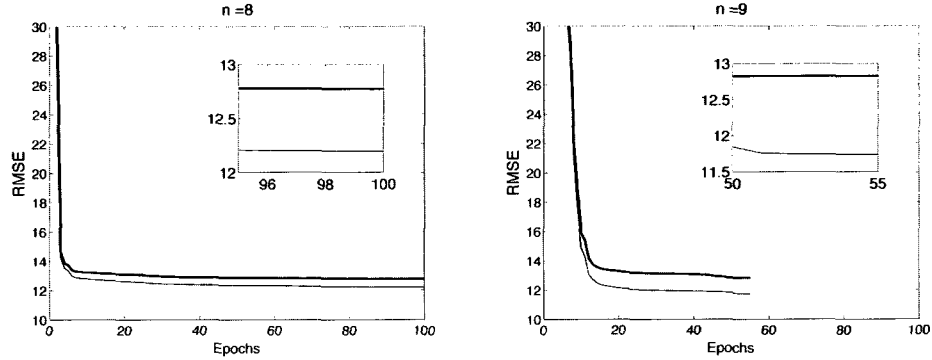


Figure 3.18 : Model 6 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $Dst_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3})$, $n = 8, 9$.

3.4.9 Model 8: 3-hour lead time AE predictions from BI

The last model, model 8, resembles the Kp model 2, but predicts AE 3 hours ahead. Since the final measured AE are 1-hour averages, for this model, we construct 3-hour averages using the conventional 1-hour average. The model 8 inputs and the corresponding ANN training parameters are shown in table 3.9. Figures 3.21 and 3.22 shows the model 8 learning histories. We want to determine n in equation 3.11.

$$AE_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3}), \quad n = 0, 1, \dots, 9 \quad (3.11)$$

3.5 Model Validation

At this juncture, it is important to point out the different criteria used by various authors for model validation and testing; the results of the models discussed in section 1.3 are mostly stated in terms of a linear correlation of the predicted versus the actual values. However, inferences as to the accuracy of the models cannot be drawn from linear correlation coefficients alone. Also, performance standards differ based on whether or not their measurements were propagated to the Earth during their

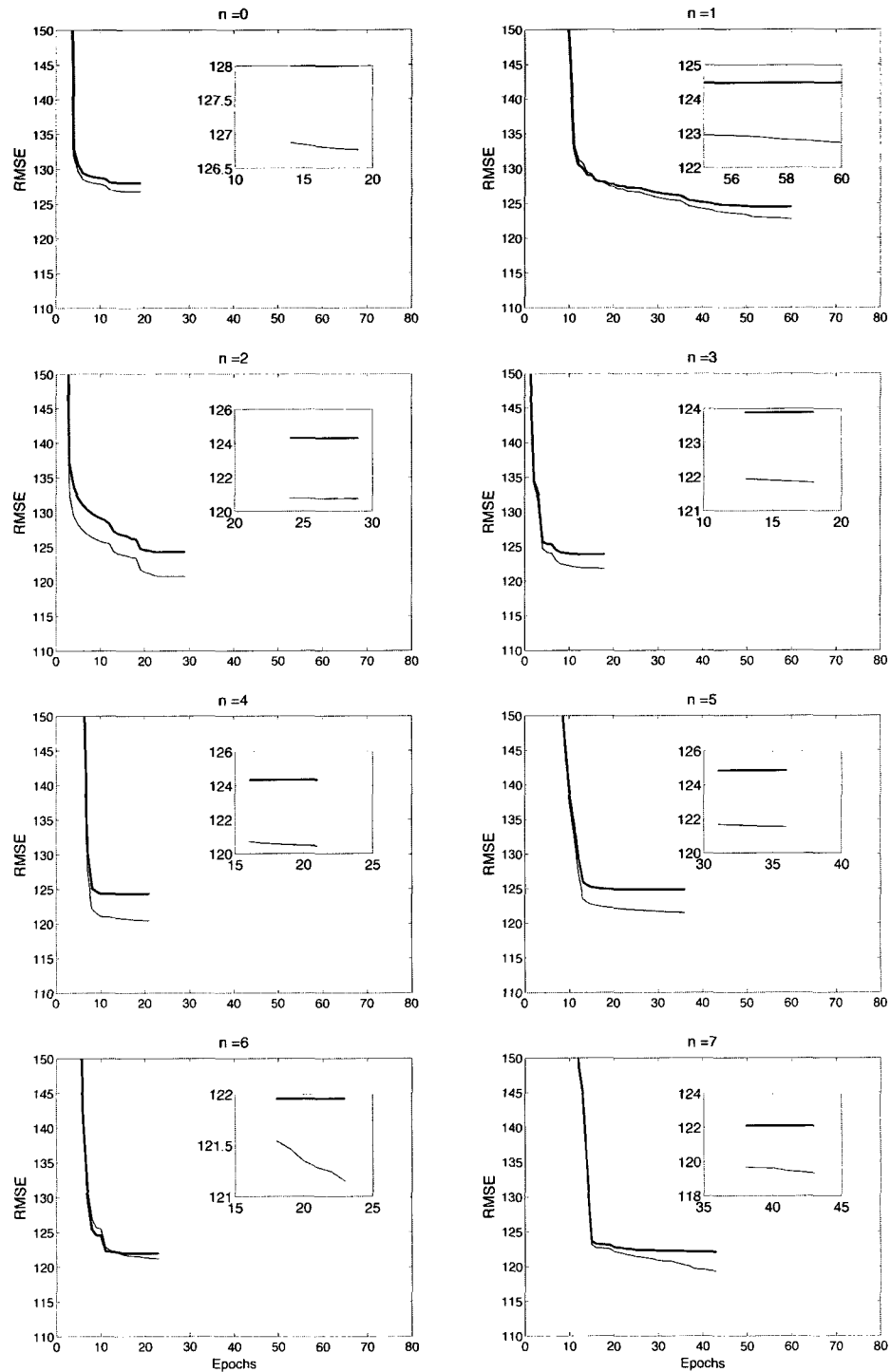


Figure 3.19 : Model 7 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $AE_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n})$, $n = 0, 1, \dots, 7$.

Table 3.8 : Predict AE_{t+1}^* from BI

#	Model 7 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test	ARE [†] Test
1	BI_t	1	58	0.005	126.77	127.99	0.289
2	BI_t, BI_{t-1}	2	58	0.005	122.72	124.47	0.277
3	BI_t, BI_{t-1}, BI_{t-2}	3	58	0.005	120.74	124.29	0.292
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}$	4	60	0.005	121.83	123.88	0.294
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}$	5	62	0.005	120.45	124.32	0.284
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}$	6	62	0.005	121.53	124.85	0.287
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}$	7	64	0.005	121.15	121.95	0.282
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}$	8	64	0.005	119.33	122.10	0.288
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}$	9	66	0.005	121.99	125.70	0.303
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}$	10	66	0.005	117.03	123.67	0.290

[†]AE > 500 nT

analysis. Time-propagated models allow usage of solar wind data from satellites located potentially anywhere and not just at L1 alone (for example APL Kp models offer that advantage). Therefore, unless models are tested on a level ground, it is often difficult to bring out their accuracies and inaccuracies specific to a storm and the discrepancies they might exhibit. One of the ways to eliminate model discrepancies, which I will show later in the discussion, is through a cross-correlation analysis.

I did not branch out too much while looking for different metrics to validate

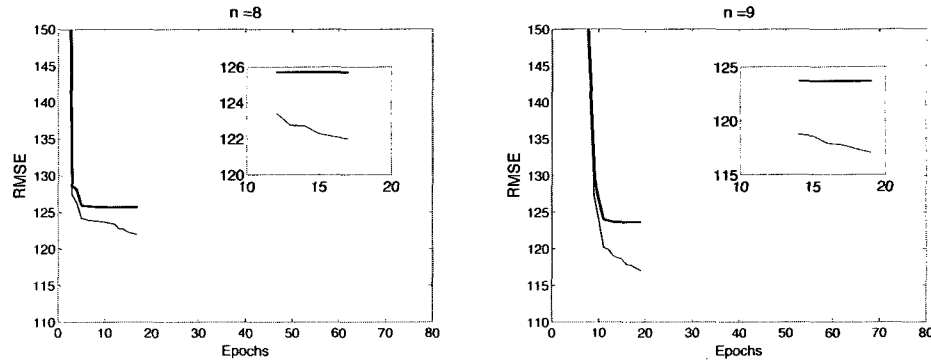


Figure 3.20 : Model 7 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $AE_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-n})$, $n = 8, 9$.

Table 3.9 : Predict AE_{t+3}^* from BI

#	Model 8 Inputs	ANN Parameters					
		Total Inputs	Hidden Nodes	LR γ	RMSE Training	RMSE Test	ARE [†] Test
1	BI_t	1	50	0.001	134.05	131.16	0.356
2	BI_t, BI_{t-3}	2	50	0.001	134.32	133.22	0.364
3	BI_t, BI_{t-3}, BI_{t-6}	3	52	0.001	131.03	135.82	0.369
4	$BI_t, BI_{t-3}, \dots, BI_{t-9}$	4	52	0.001	131.21	133.06	0.355
5	$BI_t, BI_{t-3}, \dots, BI_{t-12}$	5	52	0.001	130.90	137.71	0.370
6	$BI_t, BI_{t-3}, \dots, BI_{t-15}$	6	52	0.001	129.73	130.72	0.362
7	$BI_t, BI_{t-3}, \dots, BI_{t-18}$	7	52	0.001	131.67	138.81	0.361
8	$BI_t, BI_{t-3}, \dots, BI_{t-21}$	8	52	0.001	127.92	132.89	0.334
9	$BI_t, BI_{t-3}, \dots, BI_{t-24}$	9	52	0.001	127.77	132.71	0.367
10	$BI_t, BI_{t-3}, \dots, BI_{t-27}$	10	54	0.001	128.44	133.41	0.347

[†]AE > 500 nT

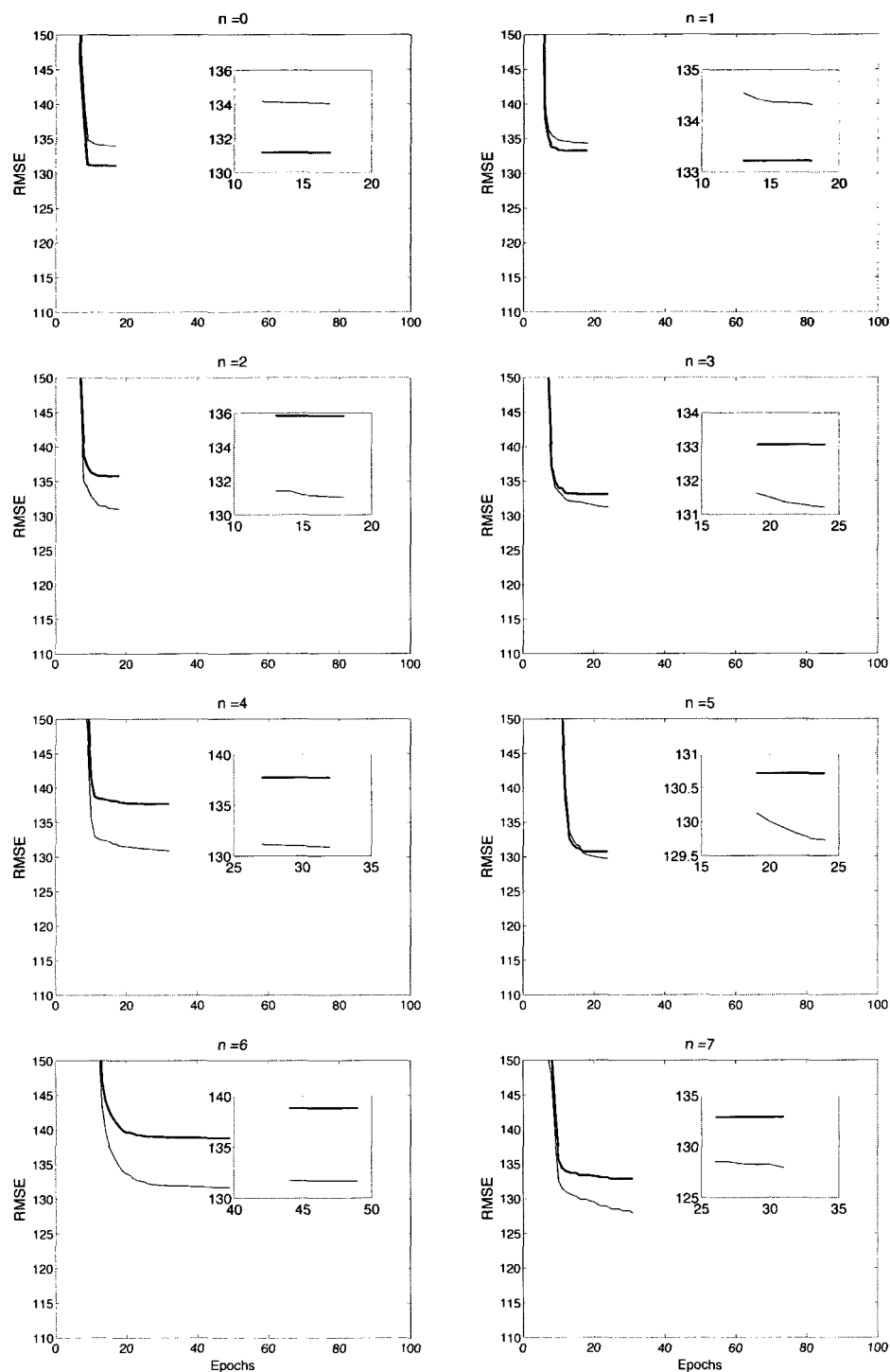


Figure 3.21 : Model 8 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $AE_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3})$, $n = 0, 1, \dots, 7$.

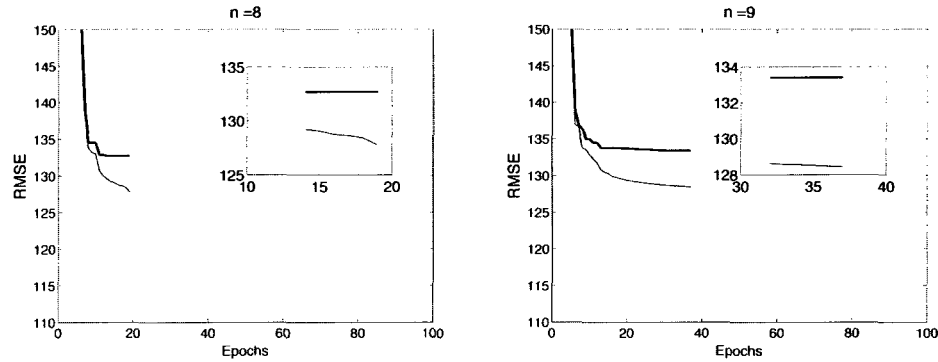


Figure 3.22 : Model 8 learning histories corresponding to table 3.6 (thick curve is the test data and thin curve is the training data). $AE_{t+3}^* \equiv f(BI_t, BI_{t-3}, \dots, BI_{t-n*3})$, $n = 8, 9$.

these models, but remained mostly within the confines of the basic statistics that are commonly noted in the literature.

3.5.1 Skill Scores

Forecasts that are based on a well-defined set of threshold limits are known as categorical forecasts; “categorical” means that the forecast consists of a flat statement that one and only one of a set of possible events will occur [Wilks, 1995]. 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 outcome of the forecast events

Table 3.10 : Contingency Table

		Observations	
		Yes	No
Forecasts	Yes	a	b
	No	c	d

are conveniently displayed using an $i \times j$ contingency table with $i \times j$ representing the possible combinations of all possible events. Table 3.10 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 2×2 contingency table. In this study, I choose the Heidke Skill Score (HSS) to represent my forecasting scheme. HSS is defined by [Wilks, 1995]:

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)} \quad (3.12)$$

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.

We can also express the probability of detection (POD) using the entries in table 3.10 as

$$POD = \frac{a}{a + c} \quad (3.13)$$

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

$$FAR = \frac{b}{a + b} \quad (3.14)$$

Perfect forecast get a FAR of 0, and the worst FAR is 1. In concurrence with the Heidke skill score (given by equation 3.12) and the linear correlation coefficient r , the forecast accuracy of the models is also characterized by RMS error as defined by

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (X_t - Y_t)^2} \quad (3.15)$$

and, prediction efficiency (PE) defined as

$$PE = \sqrt{r}, \quad (3.16)$$

where

$$r = \frac{\sum_{t=1}^N [X'_t Y'_t]}{[\sum_{t=1}^N X_t'^2]^{1/2} [\sum_{t=1}^N Y_t'^2]^{1/2}} \quad (3.17)$$

where r , X_t , Y_t represent the linear correlation coefficient, predictions and the actual values respectively. $X' = X_i - \bar{X}$ and $Y' = Y_i - \bar{Y}$ represents the deviation from the mean.

3.5.2 Tests of Significance

The linear correlation coefficient r can be useful in assessing the significance of a linear fit intrinsically; for “good” correlations (r values close to +1), the points have the tendency to be on a straight line. However, inherent data uncertainties or a few outlying data points could worsen the fit. Moreover, one cannot make an objective judgement on the fairness of a fit merely based on correlation coefficient alone. Non-linear relationships between the two variable X and Y mentioned above may not be recognized using linear correlations, and the correlations may be too sensitive to one or a few outlying pair of points [Wilks, 1995]. Therefore, model comparisons and analysis of verification statistics performed through linear correlations using a

joint distribution of the forecasted and the observed values should go beyond simple correlation coefficients to substantiate their significance.

A fair number of test choices, formally known as “tests of significance”, are available to unravel the quantitative significance of a statistical sample. To start off a hypothesis test, we define a “null distribution” along with the test level, a 5% rejection level is commonly chosen. Typically, a “null hypothesis” will be a trivial argument that one wants to defeat while defining it will be most crucial step in any tests of significance.

For example, a t-test of a correlation coefficient can investigate whether the difference between the sample correlation coefficient and zero is statistically significant. Naturally, the “null hypothesis” is defined for the linear correlation of the sample as being zero. One can then calculate the one-sample t-statistic using the student’s t-distribution through [Kanji, 1999]

$$t = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n - 2}, \quad (3.18)$$

where $n - 2$ represents the total degrees of freedom, and n being the sample size. The t probability distribution is similar to, but has heavier tails than the Gaussian distribution [Wilks, 1995]. If the t -statistic value ends up in the region that is sufficiently far from the rejection level, then the null hypothesis is rejected as too trivial to have been observed.

Z-tests are more appropriate to investigate the level of significance of r with respect to a specified value say ρ_0 where, r and ρ_0 are both derived from the same distribution. One can write the Z-test statistic using

$$Z = \frac{Z_1 - \mu_{Z_1}}{\sigma_{Z_1}}, \quad (3.19)$$

where

$$Z_1 = \frac{1}{2} \log_e \frac{1+r}{1-r}; \quad \mu_{Z_1} = \frac{1}{2} \log_e \frac{1+\rho_0}{1-\rho_0}; \quad \sigma_{Z_1} = \frac{1}{\sqrt{n-3}} \quad (3.20)$$

Further, we also assume that the sample distribution of the variables, used to arrive at r , are both normal distributions, have independent variances and have a linear relationship between them.

3.5.3 Autocorrelation Function

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

$$r_k = \frac{\sum_{t=1}^{N-k} (X_t X_{t+k}) - (N-k)\bar{X}^2}{\sum_{t=1}^N X_t^2 - N\bar{X}^2} \quad (3.21)$$

where r_k denotes the autocorrelation coefficient and k denotes the lagged time step. It is done by comparing two time series i.e., a given series is compared with itself except they are time shifted by a unit in time relative to each other; the time shifts are called “lags” and the unit of time is defined by the sampling interval. The equation (3.21) is the most commonly used function to compute lagged correlations and resembles the form of (3.17). An autocorrelation of a time series with itself is 1 i.e., $r_0 = 1$ meaning unshifted time series produces perfect correlation with itself. Thus, the autocorrelation function is simply the collection of correlations at various time lags.

In order to study the effects of “persistence”, one would want to know the lagged correlation at a time step of +1 or -1. However, it is also very common to compute the correlations at other time steps to understand other effects. For example, it is used to study the magnetospheric response time to the solar wind, which is roughly

non-linear and varying over several hours. In forecasting terminology, if the autocorrelation plot doesn't decay towards zero after a few hours/days then making a reasonably accurate forecasts in that time range should be fairly easy [Wilks, 1995]. A cross-correlation on the other hand compares two distinct time series at various lags. Conceptually, it is similar in idea to using linear correlations but with various lags as in autocorrelations. We simply extend the idea of autocorrelation to a bivariate time series in order to understand the inherent associations, between solar wind data and geomagnetic activity, for example.

Chapter 4

Solar Wind-Magnetosphere Coupling

This chapter presents a brief overview of the fundamentals of the Earth's magnetosphere, physics of the solar wind-magnetospheric coupling, and the physical processes governing magnetospheric convection. It will also discuss the BI's applicability to space weather.

4.1 Solar wind-Magnetospheric Interactions

The Earth's magnetosphere is a giant cavity in the interplanetary space above the Earth's atmosphere where the energy density is dominated by the geomagnetic field. One of the key constituents of the causally coupled solar-terrestrial system is the solar wind, a supersonically flowing magnetized plasma originating at the solar corona. It is a continuous stream of ionized gas composed primarily of hydrogen, secondarily of helium and traces of other heavier elements. In the vicinity of the Earth, the solar wind speed is highly variable, typically falling in the range of 200-800 km s⁻¹. When the supersonic solar wind flow interacts with the magnetosphere, it forms a standing bow shock that slows and deflects the solar wind. Furthermore, it alters the size and shape of the magnetosphere, and the energy input into the magnetosphere, making the system quite complex. The boundary between the solar wind and the magnetosphere is called the magnetopause.

The three principal pressure components of the solar wind which determines the size and shape of the magnetosphere, in addition to being responsible for the momen-

tum transfer into the magnetosphere, are the dynamic pressure (ρv_{sw}^2), the thermal pressure (nkT), and the IMF pressure ($B^2/2\mu_0$). The magnitude of dynamic pressure typically far exceeds the thermal and magnetic pressures by at least an order of 100 [Russell, 2007], and as a result, it plays a significant role in controlling the general morphology of the magnetosphere. Under steady-state conditions, the resulting configuration of the magnetosphere is such that the distance to the dayside magnetopause (the standoff distance) is compressed at $\approx 10R_E$, where the geomagnetic field pressure ($B_{MS}^2/2\mu_0$) balances the solar wind pressure (figure 4.1). On the night side of the earth, the solar wind stretches the magnetosphere into an elongated tail. The solar wind transfers momentum and energy by applying forces normal to the

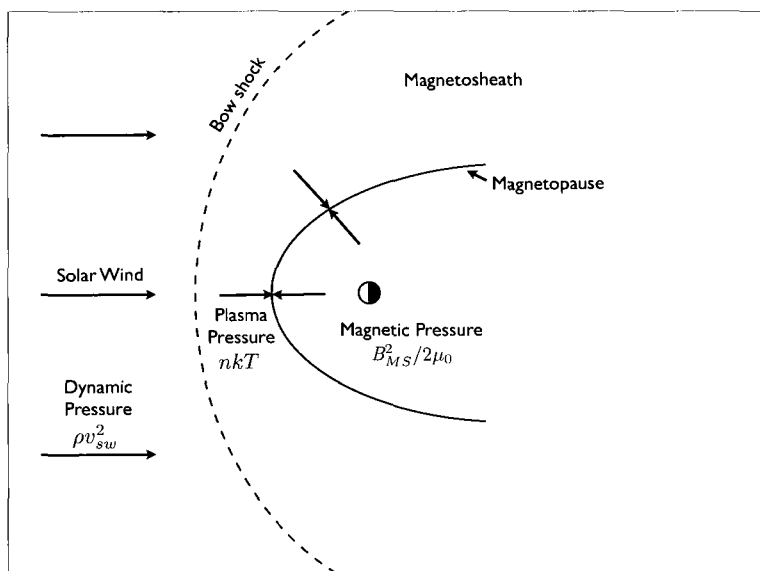


Figure 4.1 : Pressure balance between solar wind dynamic pressure and magnetic pressure of the magnetosphere. Plasma pressure is normal to the magnetopause.

magnetopause or tangentially stressing across the magnetopause to produce various large-scale processes of the Earth's magnetosphere ranging from auroras to particle

injection in the magnetotail causing ring currents. Although most of its energy flux is diverted away by the magnetosphere, a sufficient fraction of it is trapped by the magnetosphere where it is temporarily stored, and eventually dissipated through a range of different mechanisms (e.g., Joule heating in the high-latitude ionosphere, particle precipitation). The solar wind-magnetosphere coupling transfers the available energy into the magnetosphere via both “magnetic” or “non-magnetic” processes. Magnetic processes are those which imply interaction between the solar wind magnetic field and the geomagnetic field, whereas non-magnetic processes are those which involve viscous-like interactions generated via some micro- or macro-instabilities [Baumjohann and Paschmann, 1987].

We can apply the laws of ideal MHD fluid dynamics to learn about the solar wind-magnetosphere interactions because the scale lengths existing within the plasma (e.g., ion gyroradius, collisional length) are small compared to the size of the space structures such as the magnetopause or the bow shock. Since the solar wind constituents are ionized, and therefore highly conductive, it is generally safe to assume that the flux within the plasma is “frozen into” the plasma, a result first obtained by Hannes Alfvén in 1942 using an ideal MHD fluid approximation. Moreover, in an infinitely-conducting magnetized plasma, a steady bulk flow with velocity v requires the existence of an electric field in its rest frame, satisfying the ideal MHD condition

$$\mathbf{E} = -v \times \mathbf{B} \tag{4.1}$$

where v is the velocity in km/s, \mathbf{E} is the electric field in $\mu\text{V}/\text{m}$, and \mathbf{B} is the magnetic field in nT. A property of equation 4.1 is that, particles in a plasma attached to a certain field line will continue to remain on that field line, with their particle motions perpendicular to the field. Thus one can imagine a “flux tube” moving as a single entity. The following section describes various processes driving the solar

wind-magnetosphere coupling.

4.2 Magnetospheric Convection

One of the main processes that results from the solar wind-magnetospheric interactions is magnetospheric convection. In MHD, while the term “convection” refers to the bulk motion of the plasma, “magnetospheric convection” describes the circulation of plasma in the magnetosphere and in the magnetically-connected ionosphere. Plasma flow patterns over the polar regions have been recorded by ground-based instruments to show that it flows anti-sunward (noon to midnight) in the outermost layer of the magnetosphere and in the high-latitude regions of the ionosphere and then back to the dayside in the inner magnetosphere and in the low-latitude ionosphere, thereby completing the circulation [Hill, 1983]. It has been realized since the 1950s, that these are roughly stationary magnetospheric flow patterns occurring on both the dawn and dusk side which appear to mimic thermally driven convection cells, and hence the term convection. The magnetospheric convection system, hereafter called simply, convection, can be described either in terms of the local electric field \mathbf{E} or the plasma bulk velocity v using equation 4.1 which may be measurable using any low orbiting spacecraft that can detect ionospheric flow patterns.

This bulk motion has been ascribed to two leading mechanisms: magnetic reconnection of terrestrial field lines and the IMF (open-model); and viscous interaction-driven closed-model magnetosphere.

4.2.1 Open and Closed Models

Axford and Hines (1961) proposed a “closed-model” where they assumed a “viscous-like” component being responsible for the solar wind and magnetosphere interaction

to generate streamlined convective flows as shown in figure 4.2. They pointed out that viscous-like “non-magnetic” interactions arise from viscosity generated through macro- or micro-instabilities occurring near the magnetopause (e.g., Kelvin-Helmholtz instability, gradient drift, diffusion) as the solar wind flows past the Earth and carried to depth by a form of hydromagnetic eddy viscosity [Axford and Hines, 1961]. Any friction along the boundary would provide a mechanism for solar wind to transfer momentum into the magnetosphere. This fluid-like behavior is reminiscent of a falling raindrop experiencing viscous drag near the droplet/air interface which led them to suggest the resulting closed circulatory patterns inside the magnetosphere. It is a classical view that isolates the planetary magnetosphere from interplanetary space and IMF through a closed magnetic boundary, and is therefore called a “closed-model”. The geomagnetic field lines at low-latitudes having both ends on the Earth with a roughly dipole shape are referred to as “closed” field lines; the model also includes longer high-latitude lines diverging from the Earth to the magnetospheric tail.

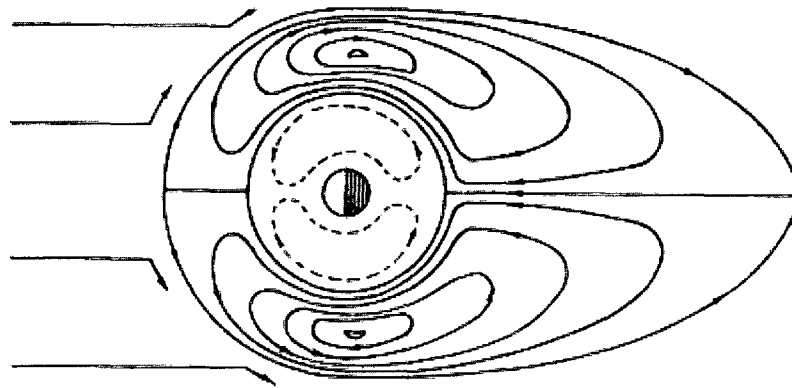


Figure 4.2 : Equatorial convection pattern in a closed magnetosphere due to viscous interaction between the solar wind (Sun is to the left) and the magnetosphere, as envisioned by Axford and Hines (1961).

Here, the magnetopause is assumed to be a closed boundary and the magnetic field lines are not allowed to cross the solar wind-magnetosphere interface. Their model postulates that a part of the solar wind momentum is transferred across the boundary of the magnetosphere to the plasma within. A convective flow carries the plasma anti-sunward and towards the tail, and is constrained to remain within the magnetosphere as the material is tied to the geomagnetic field. But under steady state conditions, an excess build-up of the material near the tail cannot continue forever. Consequently, a tail pressure build up causes a return sunward flow then completes the convection around the region closest to the Earth, resulting in a two-cell circulatory flow pattern with the field lines mapping to the polar cap as shown in figure 4.3; the anti-sunward flow occurs on the geomagnetic field lines just within the magnetopause boundary. Taking $\mathbf{E} = -\nabla\Phi$ in (2.1) implies $v \cdot \nabla\Phi = 0$, and therefore, the observed plasma flow lines are also the electric equipotentials.

Several in-situ measurements (e.g., AE-C, AE-D and S3-3 satellite data) have confirmed two-cell convection patterns and obtained voltage drops of 40-70 kV [Stern, 1996]. The observed efficiency from this mechanism, however, can only explain 10% of mass and momentum transfer overall [Baumjohann and Paschmann, 1987]. Therefore, there must be some other mechanisms besides the viscous processes contributing to the convection, especially during strong IMF conditions.

In an alternate model, Dungey (1961) applied magnetic reconnection, rather than viscous-like interaction, to magnetospheric physics to propose a new theory of convection, called the “open-model”; it “opens” the magnetosphere for magnetic flux and solar wind to cross the magnetopause at magnetic neutral points and lines [Parks, 1992]. It introduced the idea of polar cap convection produced by magnetic reconnection between the Earth’s magnetic field and southward IMF. Here, the magnetic tension due to newly-reconnected field lines near the subsolar X-line (denoted by 1

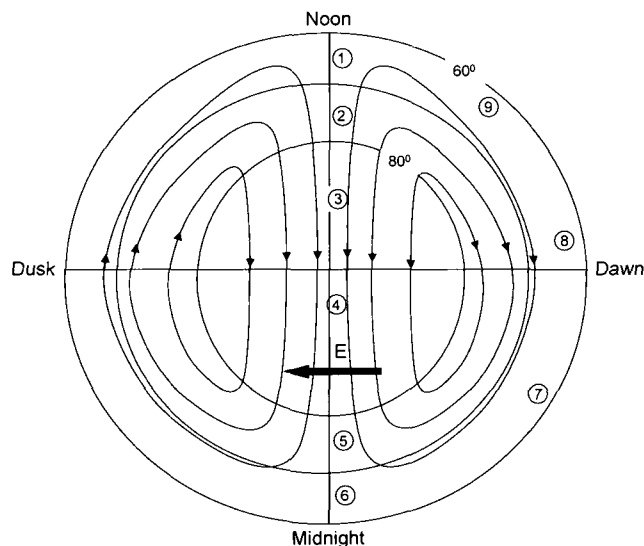


Figure 4.3 : Schematic view of the two-cell convection pattern over the polar cap shown in the geomagnetic latitude coordinate system [Kivelson and Russell, 1995].

and 1' in figure 4.4) transfers the solar wind momentum to the magnetosphere and ionosphere by dragging the field lines and plasma with it. A second X-line is formed in the distant tail where an open field line from each pole reconnects and allows the sunward flow of the flux. The resulting solar wind electric field \mathbf{E} (directed from dawn to dusk for southward IMF) maps down along the open equipotential field lines over the polar cap ionosphere. A schematic sketch representing a reconnection process for a purely southward IMF is shown in figure 4.4. As opposed to the closed-model, the anti-sunward flow here occurs on magnetically “open” field lines (one foot on each pole and one in the solar wind). The rate of energy flow in to the magnetosphere is then directly proportional to the rate at which southward IMF flux is convected to the magnetosphere and to the dawn-dusk asymmetry of the electric field resulting in strong convection. Conversely, for northward IMF, the model predicts a weak

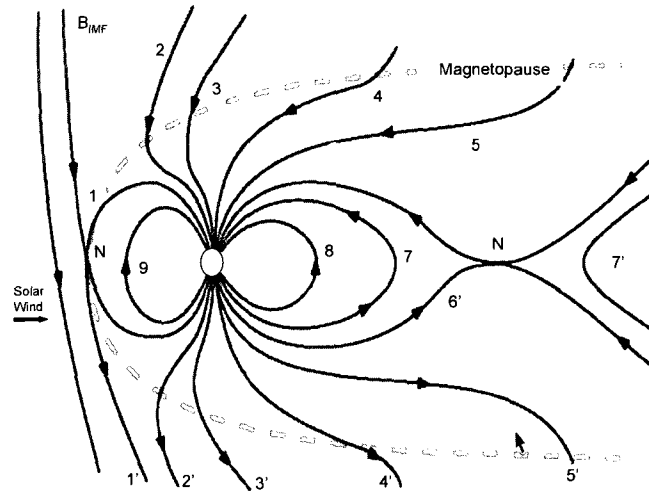


Figure 4.4 : Dungey’s model of the reconnecting magnetosphere for a purely southward IMF [Adapted from Kivelson and Russell, 2007]. N denotes the neutral point ($\mathbf{B} = \mathbf{0}$), formed near the equatorial plane.

polar cap convection and solar wind electric field \mathbf{E} , and therefore, does not provide a strong coupling between the solar wind and the magnetosphere. This model also proposed two-cell convection patterns which qualitatively resemble those of the closed-model. Its major success over the closed-model is that it predicts both an increase of strength with the negative z component of the IMF and a skew in direction with the y -component. Spacecraft observations have successfully associated various IMF orientations with the magnetic merging mechanism and resulting convection patterns [e.g., Crooker 1979; Reiff and Burch, 1985]. Some of the effects associated with dayside merging that have been observed are dayside magnetopause moving inward, auroral zone displaced equatorward, substorms etc.

While both these models can coexist continuously and, by no means, are mutually exclusive [Cowley, 1982], their relative importance is very different. Tests of these predictions show that open-model or the IMF-dependent dayside merging mechanism

is dominant with theoretical and observational estimates suggesting that it accounts for roughly 80% of the total potential required to drive convection. Nevertheless, for over four decades, both the models have stood out for their contrasting, but deep insights to magnetospheric convection, and for providing the fundamental basis that the electric field across the antisunward flow region is a good measure of the solar wind-magnetosphere coupling efficiency.

4.3 Role of the Ionosphere

The ionosphere, a partially ionized region in the Earth's atmosphere above 60 km, is another important component of the flux transfer process whose role in convection cannot be overstated. Field aligned currents (called "Birkeland" currents) are an important component of the electrodynamic structure of the magnetosphere that is involved in the energy extraction process [Vasyliunas, 1982]. The two major components to the field-aligned currents are known as Region 1 and Region 2 currents: region 1 currents flow at high-latitudes poleward, flowing down into the ionosphere on the dawn side and up on the dusk side, and are produced as a result of the solar wind and IMF driving the magnetosphere; region 2 currents, caused by pressure gradients in the inner magnetosphere, are the equatorward component of the field-aligned currents flowing in the low-latitude ionosphere. The Region 1 current system appears to persist even during periods of low geomagnetic activity and has a strong correlation with the Kp index, while region 2 currents show variable intensities besides being an important element of auroral electrojets [Iijima and Potemra, 1976].

While the region of the ionosphere over the poles (auroral zone) contains the low-altitude portion of the open field lines that connect to the solar wind and tail lobes, the region around the poles at lower geomagnetic latitudes contain closed field lines

threaded to the Earth. In the ideal case, the boundary between open and closed field lines is defined by the foot points of the “X lines” (N’s in figure 4.4). These field lines allow the magnetosphere currents to flow into the ionosphere through the field aligned currents to close the circuit, thereby completing the convection process [Iijima and Potemra, 1976]. The field-aligned currents related to the sunward return flow in the

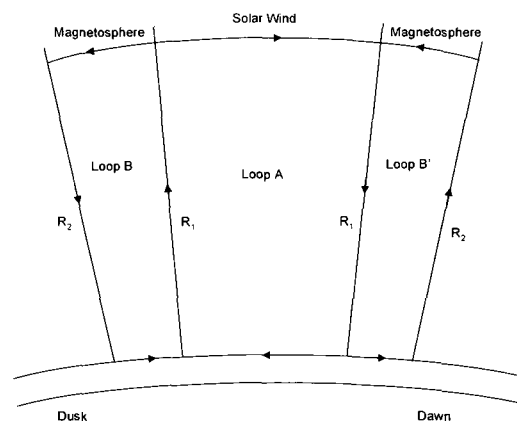


Figure 4.5 : A schematic view toward the sun and over the north polar cap showing the region of field-aligned currents. Region 1 and region 2 currents are labeled as R_1 and R_2 respectively [Adapted from Russell, 2007].

magnetosphere is shown in figure 4.5.

The other two important current systems inside the ionosphere are the Pedersen current and the Hall currents, both flowing across the field lines. The Pedersen current flowing along \mathbf{E} , driven by the dawn-dusk electric field, produces enough force to maintain the anti-sunward convection. The Hall current is the component of \mathbf{E} but flows in the direction of $-\mathbf{E} \times \mathbf{B}$. The amount of current flowing (e.g., magnitude of the Birkeland currents) in the ionosphere is largely determined by the ionospheric conductivity which varies with solar flux, seen as particle flows; observations have

pointed out the existence of conductivity gradients in the ionosphere [e.g., Moses et al., 1987; Ridley et al., 2004]. The effects of conductance and conductivity gradients are such that it creates a dawn-dusk asymmetry in the polar cap potential which, in principle, could eventually decide the maximum potential build up over the polar caps [Hill et al., 1976].

An active area of research that will enhance our understanding of the solar wind-magnetosphere coupling besides shedding light on processes such as auroras and substorms is the magnetosphere-ionosphere coupling (“MI”) coupling.

4.4 Polar Cap Potential

The Earth’s polar cap is a region of anti-sunward convection, produced primarily as a result of the two crucial processes discussed in section 4.2. The measure of the polar cap potential (PCP) is a direct measure of the rate of plasma flow through the magnetospheric convection system [Reiff and Luhmann, 1986]. The convection, if steady, can be expressed as an electric field with its integral across the antisunward flowing portion called the PCP. This convection persists for nominal solar wind conditions ($v \sim 400 \text{ km s}^{-1}$, $B \sim 10 \text{ nT}$). PCP, driven by the solar wind, is the fundamental measure of the coupling rate between the solar wind and the Earth’s magnetosphere; it also corresponds to the ionospheric plasma flow that is a low-altitude magnetospheric phenomenon. The existence of convective cell patterns over the polar cap have been duly confirmed and measured through high latitude observations from low-orbiting spacecraft (e.g., OGO-6 [Heppner, 1972]; AE-C and AE-D [Reiff et al., 1981]; S3-2 data [Doyle and Burke, 1983]; S3-3 data [Wygant et al., 1983]; DMSF data [Boyle et al., 1997]). Richmond and Kamide, [1988] developed a technique called Assimilative Mapping of Ionospheric Electrodynamics (AMIE) combining both ground- and space-

based measurements for mapping high-latitude electric fields and currents; measurements ranging from electric fields from radars and satellites to magnetic perturbations obtained from ground-based magnetometers. Ruohoniemi and Baker [1998] applied high-frequency radars of the the Super Dual Auroral Radar Network (SuperDARN) to get measurements of the $\mathbf{E} \times \mathbf{B}$ drift of ionospheric plasma over extended regions of the high-latitude ionosphere.

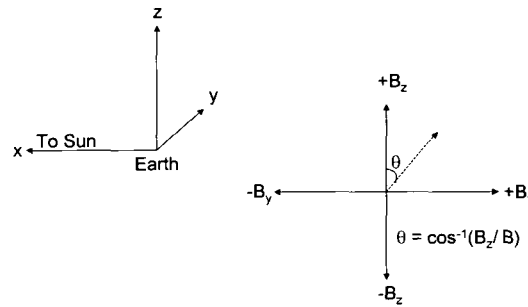
Each of these observational techniques has its own limitations, either spatially or temporally. For example, flow or electric field measurements in the upper ionosphere of low-orbiting satellites yields a value every 90 minutes from each polar cap, and though radars and magnetometers operate on a 24x7, 365-days-per-year basis, their spatial resolutions are limited. Geomagnetic perturbations could not be predicted using such low-cadence, low-altitude field measurements, however. Instead, space weather predictions focus on the search for coupling functions using the solar wind parameters such as its mass density, bulk velocity, and its magnetic field strength which can be measured upstream from Earth. Estimates of the PCP from the solar wind through empirical [e.g., Boyle et al., 1997; Newell et al., 2007; Reiff and Luhmann, 1986] or theoretical approximations [e.g., Hill et al., 1976; Siscoe et al., 2002a], available in the literature, provide an easy way to characterize the strength and magnitude of the convection system. In a closed scalar form (Φ), its magnitude ranges from as low as 10 kV to as high as 240 kV [Hairston et al., 2003].

In this dissertation, we focus on functions derived from the solar wind and IMF as best estimates of the PCP. Vasyliunas et al [1982] gave a quantitative estimate of the PCP, through dimensional analysis, of the form

$$\Phi_{PCP} = v_{sw} B_T L_{CF} Q(\theta, M_A^2) \quad (4.2)$$

where B_T is the projection of the IMF on the solar magnetospheric y-z plane, L_{CF} is

the Chapman-Ferraro scale length, and $Q(\theta, M_A^2)$ is a dimensionless function of two variables, M_A the Alfvén-Mach number, and θ is the clock angle as shown in figure 4.6. They also argue that the two other dimensionless quantities that can occur in this expression are the Pederson conductivity (Σ_p) and the Reynolds number related to measuring the relative importance of the inertial to viscous effects. Several merging models of the form of equation (4.2) [e.g. Sonnerup, 1974; Gonzales and Mozer, 1974] have been developed and tested observationally [e.g., Doyle and Burke, 1983; Reiff et al., 1981; Wygant et al., 1984]. The conclusions from these studies are that measured potentials during storms are consistent with the theoretical merging rates, while during periods of strong northward IMF processes other than reconnection mainly contributes to the PCP.



Geocentric Solar Magnetospheric (GSM) coordinate system

Figure 4.6 : The Geocentric Solar Magnetospheric (GSM) coordinate system. X-axis points to the Sun and the dipole is in XZ plane. Y-axis is perpendicular to the Sun-Earth line and point towards the dusk side.

Fundamentally, an effective increase or decrease in solar wind speed, strength of field, or a change of IMF orientation causes the cross-magnetospheric electric field to change, and hence the PCP. Convection patterns are known to vary with the

orientation of the IMF, particularly its B_z , as expected from the merging mechanism, and to the magnitude of the solar wind speed. It is now accepted, based on the observations discussed above, that as the IMF B_z decreases, the cross polar electric field increases and consequently, the PCP increases linearly up to ~ 160 kV [Reiff and Luhmann, 1986] corresponding to an upstream solar wind electric field of ~ 6 mV m $^{-1}$. Convection also arises due to magnetic merging between the IMF and magnetotail field lines, during periods of northward IMF ($B_z > 0$) [e.g., Russell, 1972; Reiff, 1982; Zanetti et al., 1984], resulting in convection cells known as “lobe cells” [Reiff and Burch, 1985]. However, during such times the measured PCP was strongly influenced by the number of hours IMF remained northward [Wygant et al., 1983]. All these studies led to two major quantitative conclusions about the extrema of the PCP estimates, the existence of “residual” and “saturation” potentials.

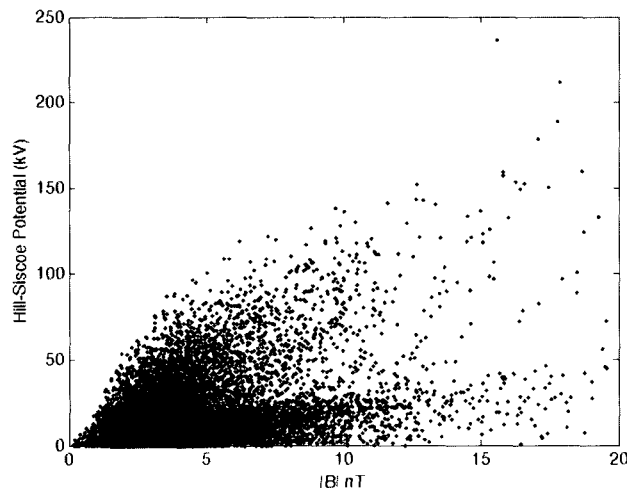


Figure 4.7 : Hourly averages of PCP are plotted against $|B|$ for 2006-2007 using the Hill-Siscoe formulation [Siscoe et al., 2002a]. PCP asymptotes around 160 kV.

Empirical studies focussing on correlation between the IMF and PCP substantiated the presence of strong residual effects resulting in “residual” potentials, which

were later ascribed to “viscous” processes. A observational study by Wygant et al. [1983] measured these residual potentials to point out that only 10-20 kV can be from viscous sources, consistent with the theoretical expectations [e.g., Hill, 1979; Pu and Kivelson, 1983] and experimental data [e.g., Sanders et al., 1980; Mozer, 1984]. On the other extreme, when the IMF amplitudes are large ($|B| > 10$ nT), PCP tends to saturate around 150-200 kV (figure 4.7).

The effects of polar cap “saturation” is now quite well understood through several observations and techniques, and will be discussed in detail in section 4.6. Thus, a simple relationship between the IMF and the PCP cannot be found over their whole dynamic range. We now turn our attention to the Boyle potential, another form of solar wind-magnetosphere coupling function.

4.5 The Boyle Index: A Solar wind-Magnetosphere Coupling Function

As large volumes of the IMF data became more and more available, a large number of quantitative relationships have emerged to parameterize the coupling between the solar wind and the magnetosphere [e.g., Reiff and Luhmann, 1986; Newell et al., 2007; Siscoe et al., 2002a; Weimer, 2001; Wygant et al., 1983]. While some of them had theoretical motivation, the rest were purely empirical. The details involving the polar cap convection processes are becoming clear with a wide array of studies to date, and such studies have made the distinction between the “viscous” and “merging” terms contributed by solar wind plasma and IMF data respectively; some of the commonly noted viscous terms in the literature are n , v , nv , p , $p^{1/3}$, $p^{1/2}$, v^2 etc, and some commonly used merging terms are B_S , vB_T , $vB^2 \sin^4(\theta/2)$ etc. One of the benefits of applying this concept of coupling functions to space weather is underscored by a recent

finding by Newell et al. [2007] that the presence of a solar wind-dependent viscous interaction term and the IMF-dependent merging term in a solar wind coupling function makes it a good candidate to describe the state of the magnetosphere system over a wide variety of magnetospheric activity [e.g, $vB_T \sin^4(\theta_c/2)$, $v^{4/3}B_T \sin^4(\theta_c/2)p^{1/6}$, $vB_T^{2/3} \sin^{8/3}(\theta_c/2)$ for merging terms; n , v , nv for viscous terms] .

One such relationship is the Boyle potential or the Boyle Index (BI), an empirical formula derived by Boyle et al. [1997] through an analysis comparing 3.5 years of Defense Meteorological Satellite Program (DMSP) flow data to solar wind parameters. It approximates the steady state PCP through solar wind and IMF, with its best fit formulation given by:

$$\Phi = 10^{-4}v_{sw}^2 + 11.7B \sin^3\left(\frac{\theta}{2}\right) \text{ kV}, \quad (4.3)$$

where v_{sw} 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]. This is a steady-state model in that they imposed a 4-hour steadiness criterion to the IMF in their study. The viscous or IMF-independent term here ($10^{-4}v_{sw}^2$) is proportional to the solar wind flow energy, and it contributes 16 kV (assuming $v_{sw} = 400 \text{ km sec}^{-1}$, $\theta = 0$ for B_Z northward), consistent with observations ranging between 10-20 kV from viscous contributions [e.g., Wygant et al., 1983] while during extreme solar wind conditions caused by high-speed streams, the viscous contribution can reach ~ 30 kV. Furthermore, their IMF-dependent “merging” term does not depend on the solar wind pressure significantly. In general, an hourly-averaged BI is a good predictor of the polar cap potential drop for sufficiently long and steady periods of solar wind and IMF, and for BI less than 160 kV. However, beyond 160 kV, the BI overestimates the PCP.

4.6 Polar Cap Saturation

One of the earliest studies on polar cap voltage “saturation” was done by Hill et al. in 1976. They theorized that during periods of sufficiently large IMF, region 1 currents driven by the solar wind produce magnetic field perturbations that are comparable and opposite to the local field, eventually limiting the rate of dayside reconnection and placing an upper bound on the PCP. As a result, ionospheric line-tying restricts the convection at larger potentials while the convection due to dayside reconnection dominates at lower limits. This result is elegantly put in a simplified form, called the Hill model, as follows:

$$\Phi_{pcp} = \frac{\Phi_m \Phi_S}{\Phi_m + \Phi_S}, \quad (4.4)$$

where Φ_{pcp} approaches Φ_S , the saturation potential for $\Phi_m \gg \Phi_S$, and Φ_{pcp} approaches Φ_m , the magnetospheric convection potential for $\Phi_m \ll \Phi_S$. In other words, Φ_{pcp} is limited by the smaller of Φ_m and Φ_S . Recently, Siscoe et al., [2002a], starting from the Hill limiting expression of (4.4), provided a quantitative expression for the PCP as a function of the solar wind parameters and the ionospheric (Pedersen) conductivity. They also pointed out that an external mechanism in the form of solar wind Alfvén-Mach number decrease might influence the potential, possibly causing the saturation. The quantitative form of the Hill-Siscoe potential is given by:

$$\Phi_{pcp} = \frac{57.6 E_{sw} P_{sw}^{1/3} D^{4/3} F(\theta)}{P_{sw}^{1/2} D + 0.0125 \xi \Sigma_p E_{sw} F(\theta)} \text{ kV}, \quad (4.5)$$

where $E_{sw} = |V_{sw} \times B_{sw}|$ is the upstream solar wind electric field, P_{sw} is the ram pressure exerted by the solar wind, D is the Earth’s dipole field, $F(\theta) = \sin^2(\theta/2)$, ξ is a dimensionless coefficient based on the currents in the ionosphere, Σ_p is the Pedersen conductivity obeying the relation $\xi = 4.45 - 1.08 \log(\Sigma_p/1S)$.

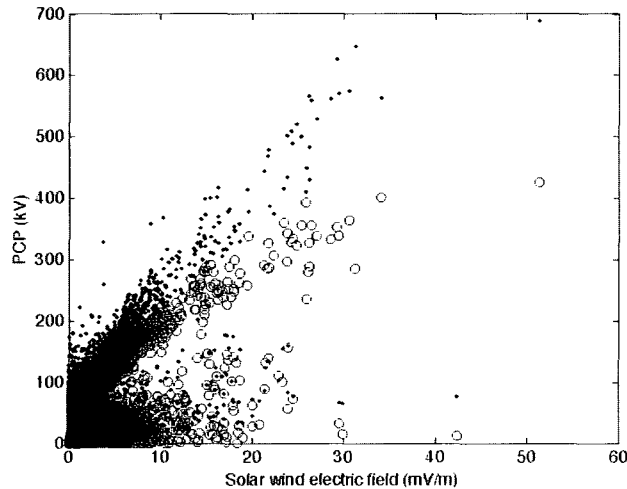


Figure 4.8 : A 11-year data (using 1-hour averages) comparing the Boyle index (black diamonds) with the Hill-Siscoe model (open circles) as a function of solar wind electric field. While the former shows a linear rise, the latter asymptotes to a constant value for large electric fields.

Polar cap voltage saturation has been studied by several authors in the past using a variety of observational techniques (e.g., satellite probes, radar measurements, magnetic measurements). We now have enough evidence to substantiate the theory of saturation. 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. Recently, Ridley [2005], examined 13 different events and clearly showed evidence of saturation, and further showed that saturation tends to occur when the solar wind Mach number decreases (externally). However, a couple of issues still remain to be addressed: the ways of determining if saturation is caused by an internal or external mechanism, and to what extent; and determining the true value of the saturation potential. Further studies are required to answer these questions. I have compared the BI with Hill-Siscoe formulation using 11 years'

worth of data (1997-2007) using hourly cadence, illustrated in figure 4.8. Here, the Pedersen conductivity (Σ_p) for the Hill-Siscoe model is 10, a model parameter.

4.7 Newell functions: another coupling formula

Newell et al. [2008], from a rigorous analysis, have shown that the single coupling function from the solar wind is enough to predict a wide variety of magnetospheric phenomenon without relying on the time history of the target index. Just like the BI, their coupling function contains a magnetic-merging term and a viscous term. Using their merging term ($d\Phi_{MP}/dt = v_{sw}^{4/3} B_T^{2/3} \sin^{8/3}(\theta/2)$) alone, they were able to correlate 9 out of 10 indices of magnetospheric activity including the Kp and the AE index. However, combining their top-performing viscous function ($n^{1/2} v_{sw}^2$) with the merging term provided the best combination overall to predict up to 61% of variance across all indices. Their merging term is proportional to the rate at which the field lines are convected towards the magnetopause (v), the strength of the IMF (B_T), the length of the merging line and the probability of field lines impacting the magnetosphere (obtained from empirical fit). The one index that is of specific interest to the work of this thesis is the Kp index. They approximated the Kp index through:

$$Kp = 0.05 + 2.244 \times 10^{-4} \frac{d\Phi_{MP}}{dt} + 2.844 \times 10^{-6} n^{1/2} v^2, \quad (4.6)$$

$$\text{with, } \frac{d\Phi_{MP}}{dt} = v^{4/3} B_T^{2/3} \sin^{8/3}(\theta/2), \quad \theta = \cos^{-1}(B_z/|B|) \quad (4.7)$$

After studying two 11-year periods for Kp they conclude that their function can predict the variance in Kp up to 75%, $r = 0.866$ (see figure 4.9). Furthermore, and interestingly, they claim that it is quite simple to use just two physics based terms to predict Kp better than an optimally trained neural network.

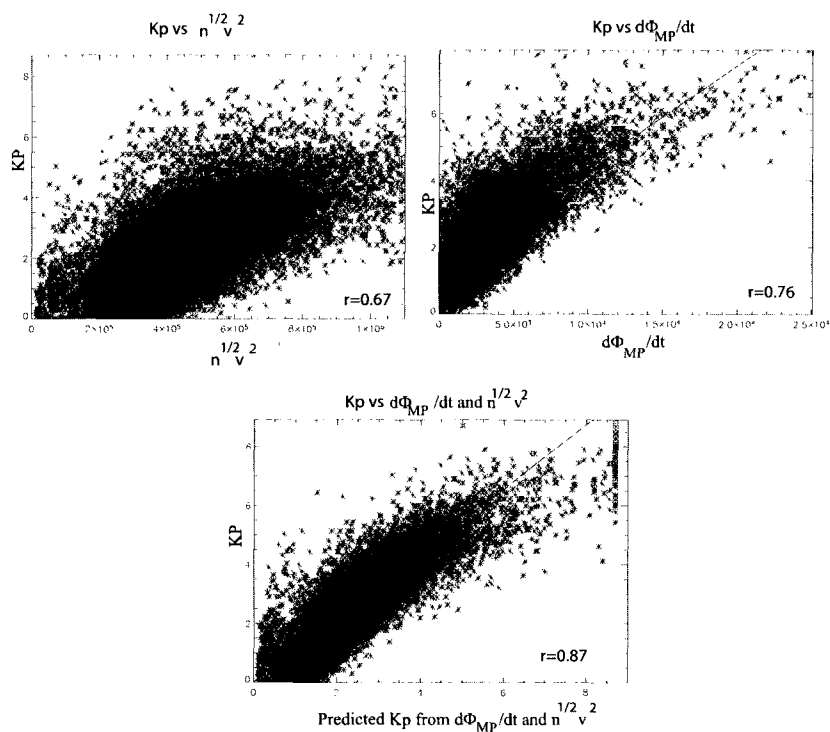


Figure 4.9 : K_p predicted using equation 4.6, Newell et al., [2008].

4.8 Boyle Index: Effect of Preconditioning Events

Coronal mass ejections (CME) from the sun, large solar flares, magnetic 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. Preconditioning is also caused by a colder and denser plasma that is convected inward leading to increased ring current development as well as contributing to the plasma sheet during the main phase of a storm. An increased ring current development is typically associated with magnetospheric and ionospheric disturbances.

Figure 4.10 presents an example of a preconditioning event. It shows a time series plot of the solar wind, magnetospheric and Kp index values of a storm ($BI > 200$ kV; $Kp > 6$) that occurred on 14 April 2006. The Dst index is commonly used to measure the strength of the ring current which in this case is, $Dst < -110$ nT. This event has been chosen for its steady high Boyle index lasting over 5 hours (> 194 kV). 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. Success of a forecast algorithm, therefore, depends on training the network with preconditioning events and the magnetospheric response to such events. Due to the limitations imposed by the learning algorithms, it may not be feasible to look back into a preconditioning event and the whole storm interval simultaneously beyond 9 hours—for an average storm, it would mean that one would have to weight the inputs for as long as 20-25 hours in order to achieve that. However, the models 2, 4, 6 and 8 (see chapter 5) have a time resolution of 3 hours and in principle, should be able to capture both the preconditioning and the storm more thoroughly.

4.9 The BI and space weather

The BI plot (figure 1.11) was launched into real-time mode in Oct. 2003, purely for a scientific and educational motive. It is available from <http://space.rice.edu/>

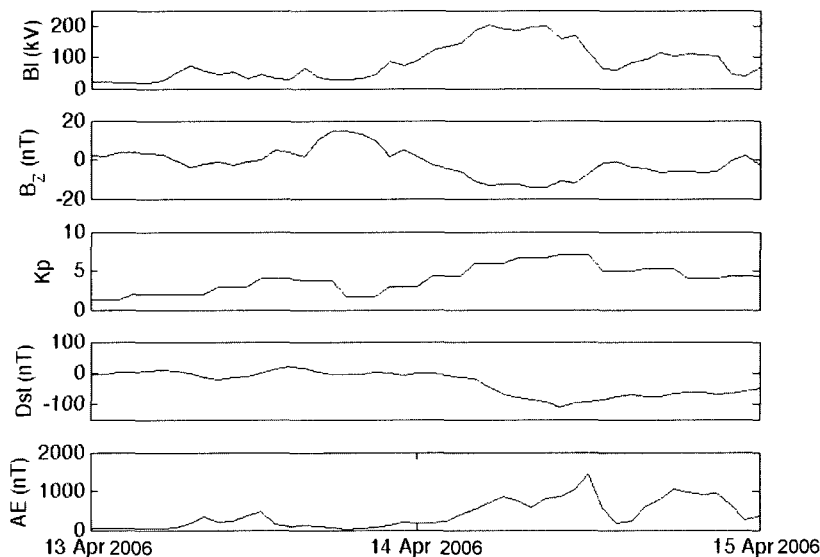


Figure 4.10 : Time series plots of the ACE and a few geomagnetic indices following a CME in April 2006 (3-hour averages). From top to bottom: (a) derived Boyle index, (b) B_z (IMF), (c) official Kp values, (d) the Dst, and (e) the AE index.

ISTP/wind.html as a courtesy of the *Rice Space Institute*. Since its inception, the subscribers to the “spacalrt” mailing list receive email notices whenever the 10-minute BI average exceeds 200 kV, called “red alerts”. In over 6 years of real-time operations, it has enjoyed a lot of success and no major storm ($Kp > 6$) has been missed, which demonstrates its value as a forecasting tool.

The BI calls for steady-state conditions to prevail in the IMF for at least four hours, whereas this whole study is based on 1 and 3-hour averages of the BI. However, by setting up a non-linear neuron model through ANNs, we can capture the non-steady states in the solar wind by weighting time-integrated BI over time i.e., looking back several hours into the past. This is particularly useful to study the processes driving

the convection, which tends to be non-linear. Furthermore, the BI does not exhibit an apparent saturation effect, but tends to be linear over its whole dynamic range. Interestingly, the general linear response of the BI to the solar wind, the observed linear correlation between the $\log(\text{BI})$ with K_p (figures 1.4) and the BI with Dst and AE (figures 1.5) could be high under extreme conditions and in the domain where these indices are large. This could be vital for an operational setup, setting limits and thresholds for alerts, etc. This study is unique in that it is the first ANN K_p

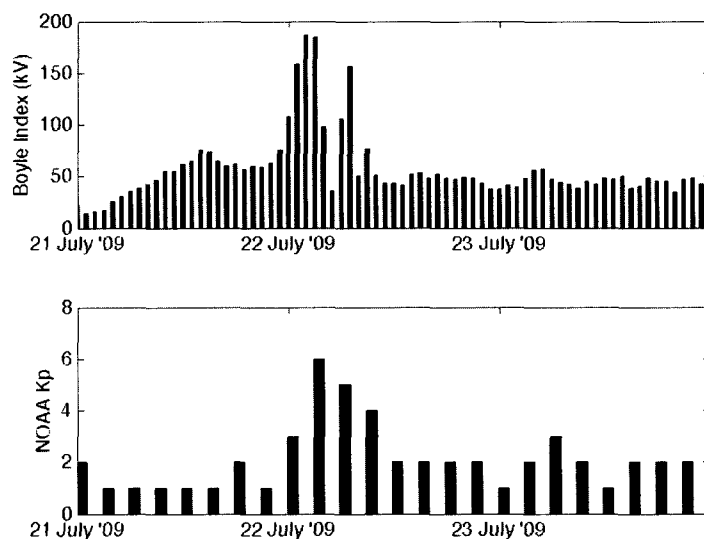


Figure 4.11 : Real-time “snapshot” of the BI (1-hour average) and the NOAA K_p during a recent magnetospheric activity on 22 July 2009.

prediction model to use only a single input (a coupling function) or, at the most, two inputs (a coupling function plus the NOAA nowcast- K_p). These claims will be tested in the succeeding chapters. Figure 4.11 is an example of a real time plot, and in this case, during a magnetospheric activity (22 July 2009) as the new solar cycle 24, which started in December 2008, begins to ramp up. The success of our ANN predictions during this event will be discussed later in chapter 5.

Chapter 5

Scientific Results

This chapter deals with an in-depth analysis of the Boyle Index (BI) as a potential forecasting parameter in order to find statistical correlations between the BI versus Kp, Dst and AE. A true correlation and a good performance skill score will enable us to characterize the global geomagnetic activity indices in terms of the BI which is derived from the solar wind. I will apply the Heidke Skill Score (HSS) for a set of defined discriminator levels. Simple linear predictors gave us an idea of how well the BI can predict the changes in these indices to their next time step. Further analysis through cross-correlations will help us ascertain the average magnetospheric response time to changes in the solar wind and interplanetary magnetic field (IMF) conditions. These insights are particularly useful as we weight and time-integrate both the BI and the geomagnetic indices using an artificial neural network (ANN). Furthermore, in order to eliminate the statistical uncertainties in the correlation coefficients and to rule out the possibility that the results obtained here are not occurrences of mere chance, we rely on a large data set to perform the skill score statistics.

In order to present an unbiased view, part of the data was withheld by random selection for the purposes of testing alone. In other words, the data were not part of the ANN training. Otherwise, we run the risk of contaminating our scores by bringing an “artificial” skill effect into them. This procedure, where part of the data is withheld for the purposes of testing, is called “cross-validation”. Earlier in this dissertation, I introduced this idea, but in a different context i.e., in training an ANN, wherein

their primary purpose will be to ensure the minimization algorithm does not under- or over-fit the data to avoid memorization rather than generalization.

5.1 Linear 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 10 kV, and can reach well over 500 kV before or during severe storms. My analysis is facilitated by choosing a logarithmic transformation to scale the BI and to be in tune with the Kp index which is quasi-logarithmic. A scatter plot of a 3-hour average of the logarithm of BI and the following 3-hour Kp is shown in figure 5.1 with each quadrant representing either a hit, miss, false positive or a correct rejection. In this figure, as an example, the vertical and horizontal lines within the plot represents one arbitrary BI cutoff (discriminator level, here 110 kV) and corresponding Kp index (5) cutoff. A more rigorous approach is to compute the HSS distribution for a specific Kp threshold, say 4, by sliding the vertical BI

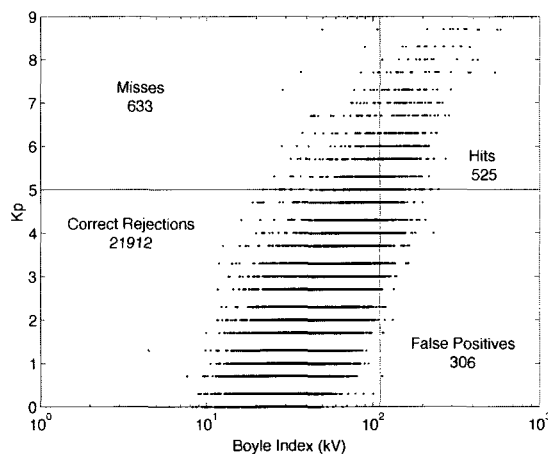


Figure 5.1 : 3-hour averaged $\log(\text{BI})$ versus the following 3-hour Kp of a complete data set between 1998 and 2005 ($r=0.74$) [Bala et al., 2009].

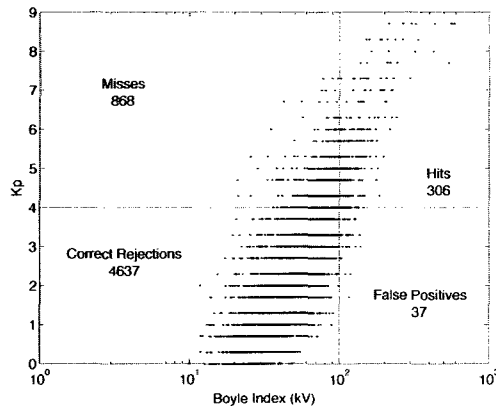


Figure 5.2 : Plot shows the 3-hour averaged $\log(\text{BI})$ versus the following K_p for 2003 and 2004 with a linear correlation coefficient of 0.785. Note that the chosen BI cut-off in this case is 100kV while the trigger level is reduced to $K_p = 4$ [Bala et al., 2009].

discriminator through the entire range of available BI values. The logarithm of the BI correlates well with the following 3-hour K_p ($r = 0.74$). We have already established that the BI is an overestimation of the PCP above 160kV, and that it is linear with increasing solar wind electric field. Unsurprisingly, the correlation is still valid and the trend line continues to be linear at higher K_p s, possibly because K_p is logarithmic as well but also because perhaps, even though the polar cap potential may saturate, the overall magnetospheric response may not. Since the BI does not include a saturation term, it will overestimate the true potential for major storms. However, since certain measures of geomagnetic activity do not saturate, a BI of 300 does imply a stronger storm than a BI of 200, even though the actual polar cap potential may turn out to be about the same because of saturation. The cut-off shown here has been deliberately chosen to emphasize a fact that in the 8 year period we studied, the likelihood of a storm, having a K_p index of 5 or higher, exceeds 95% when the average BI over the previous three hours is over 110 kV. However, a good number of “misses” occur with that discriminator level. Let us look another scenario (figure 5.2) where we reduce

the discriminator level to 100 kV, so the number of “misses” decreases, and the “false positives” increase. Depending on the kind of forecast and the level of sensitivity desired, one can set these cut-offs arbitrarily by trading a few hits for misses, right rejections for false alarms, and vice versa. It illustrates that during 2003 and 2004, when the BI fell below 100 kV, the number of “right rejections” dominated, suggesting that the magnetosphere is typically quiet (K_p index < 4) for $BI < 100$ kV .

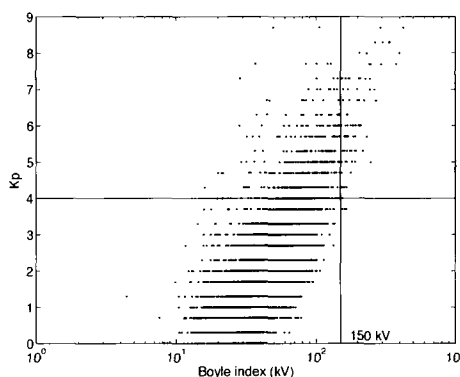


Figure 5.3 : 1-hour averaged $\log(BI)$ versus the following K_p ($r=0.71$) for an active period during 2000 and 2001 is shown here. BI cut-off in this case is 150kV [Bala et al., 2009].

As another example, during active periods of solar maxima, as shown in figure 5.3, hourly averages of the BI and K_p are correlated, where the K_p has been oversampled to one hour resolution i.e., values of the 3-hour average were smoothed to each of the preceding two hours. We identified 2000 and 2001 to have several severe events corresponding to their proximity to the maximum phases of solar cycle 23 [see Cane and Richardson, 2003]. During some very active periods, an hourly averaged BI can exceed 250 kV, in which case the geomagnetic K_p index could be over 7, causing major geomagnetic storms and low-latitude auroras to form within the succeeding few hours. This general trend of a linear rise in K_p with the preceding BI, regardless

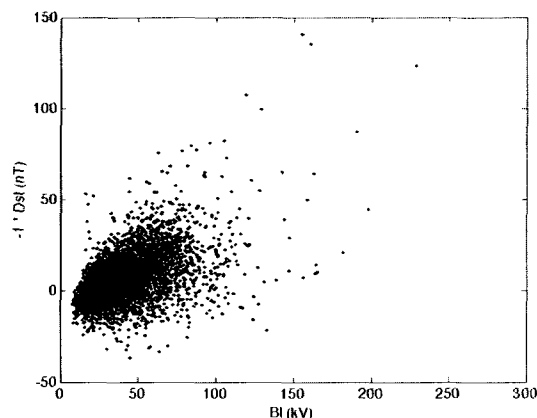


Figure 5.4 : 3-hour averages of the BI is compared against the Dst index for 2006-07.

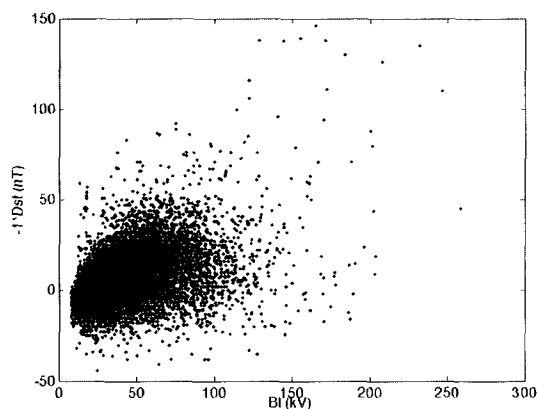


Figure 5.5 : 1-hour averages of the BI is compared against the Dst index for 2006-07.

of the cadence chosen, is critical to the training of an ANN as well as to space weather forecasters searching for solar wind and IMF-based parameters to predict magnetospheric phenomena.

Unlike Kp, correlations of BI with Dst (figures 5.4 and 5.5) and BI with AE (figure 5.6) are derived by comparing them on linear scales, because both AE and Dst are non-logarithmic and vary over a free range. Due to the large scatter in the plots discriminator levels are hard to set. The

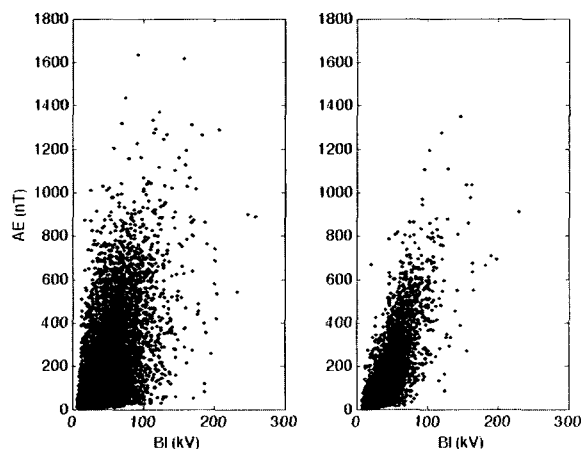


Figure 5.6 : 1-hour (left) and 3-hour (right) averages of the BI is compared against the AE index for 2006-07.

5.2 Cross-correlation Analysis

The Earth's magnetosphere is a highly non-linear and stochastic system. Cross-correlation analysis is one of the fundamental forms of time series analysis performed in the time domain that can offer clues about characteristic time scales of small- or large-scale processes. They offer the best choice in dynamical weather forecasts to explore the temporal correlations involved in a linear time series, be it uni- or bivariate. We therefore 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.

In figure 5.7, I plot the estimated crosscorrelation 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, I 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. We might wonder if the

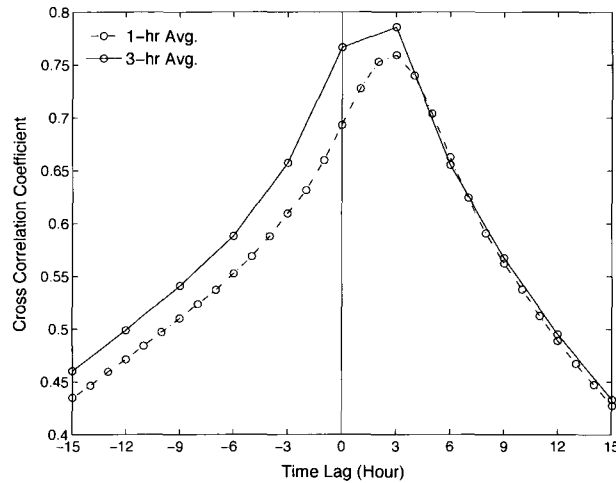


Figure 5.7 : Cross-correlation function of $\log(\text{BI})$ and Kp (oversampled) versus the time lag. Both 3-hour averages (solid curve) and 1-hour averages (dash-dot) display a strong correlation at 3 hour with the BI leading Kp [Bala et al., 2009].

1-hour cadence would produce a different result than the 3-hour cadence. Actually, an hourly averaged (dashed-dot curve) $\log(\text{BI})$ and Kp also has a good correlation between them and the strongest correlation (0.76) still occurs at a positive time lag of 3 hours. Figures 5.8 and 5.9 shows the crosscorrelation function for the BI vs Dst and AE respectively. For the Dst, the 3-hr cadence shows good correlations in the 3 to 6 hour time lags with highest correlation occurring at a time lag of 6 hours. The hourly cadenced Dst shows good correlations in the 3 to 6 hour time lags, just like the 3-hour averages. AE index shows good correlations in the 2-5 hour time range. There are no significant difference in the plots between 1 and 3 hour averages, and a clear trend line is visible. The existence of a clear trend line means that successive values are highly correlated with each other [Makridakis and Wheelwright, 1978], implying that the time series is predictable. This behavior can be seen for both positive and negative lags. For positive lags, there is a strong tendency for persistence. For negative time lags, however, the opposite is true and the past values negatively influence the future

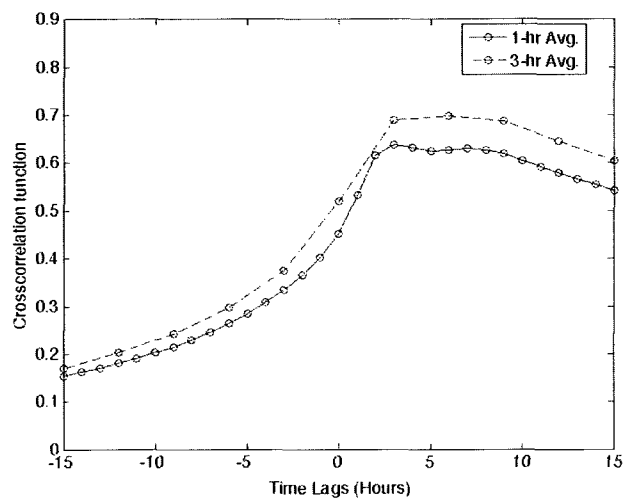


Figure 5.8 : Cross-correlation function of BI and Dst: 1-hour averages (solid curve) and 3-hour averages (dashed).

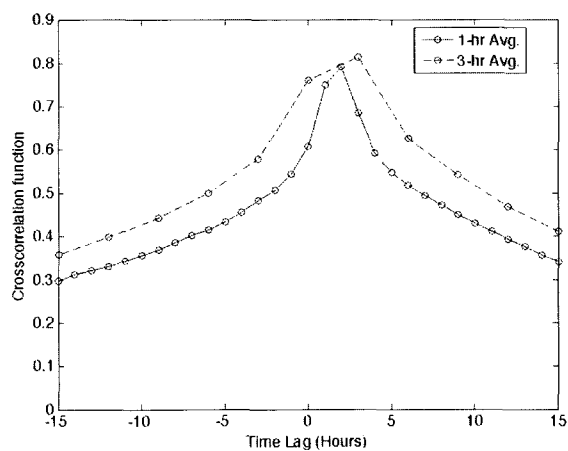


Figure 5.9 : Cross-correlation function of BI and AE: 1-hour averages (solid curve) and 3-hour averages (dashed).

values. We can also note that the lagged correlations are extremely small for large lags and decay rapidly after several hours, though it does not decay to zero even after several hours. We can therefore infer that with Kp, Dst and AE 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 1-6 hours where the correlations are significant at 95% confidence level ($r \geq 0.6$ from t-tests), in concurrence with the discussions we saw earlier about autocorrelations. Therefore, by training the preconditioning events using the BI on time scales imposed by cross-correlation analysis, reliable proxies can be estimated in advance. From an operational standpoint, these results strongly suggest the feasibility of using the BI to make short-term predictions of the magnetospheric activity.

5.3 Prediction Algorithms

In this section, I introduce the four Kp, the two Dst and AE algorithms and describe their architecture in detail. Also, I will validate and test these models using the skill scores described earlier. Furthermore, we will see whether or not “persistence contamination” influences the predictions and, if pure solar wind-driven models are better for forecasts.

The question I would like to find the answer to is, is it possible to create stand alone algorithms using “only” the BI and have them predict Kp as “efficiently” as algorithms using inputs that also use the time history of Kp? Given the amount of data and network resources needed, both Kp-dependent and Kp-independent models were large and complex to build, requiring a lot of training time. All our models are unique in that, unlike their predecessors which were trained using combinations of solar wind parameters, this is the first time ever that ANN-based Kp models were trained solely using a single input parameter, the BI. Two of my models, however, use both the prior Kp and the BI as input parameters. Organizing the data to the ANN means using just two input streams: the BI and either of Kp, Dst or the AE. The “final” archived geomagnetic data is fairly uninterrupted, but the solar wind data

have data gaps. The gaps were discarded before the training i.e., if the time history of the data stream specific to a input pattern had unusable points then that training pattern was discarded.

ANN's are good forecasting machines, provided they are well trained. Recall from chapter 2, the LM routine, a superior form of the gradient descent algorithm, offered the best choice to train my models. Weight optimization or training was achieved by adjusting various input parameters via trial and error, primarily through the research methodology prescribed in chapter 3. I was able to achieve my desired target error by having the network "cross-validate" as it trained, where the total sample was segregated equally for training and validation; cross-validation is the best solution to prevent "over-training". The training was halted as soon as the error reached the desired minimum or when the error on the validation set was acceptable; any overtraining will affect network generalization. Monitoring the instantaneous error as the training proceeds, manifested in the form of learning curves, is therefore important to the success of any learning procedures. The weights are now said to be at their closest to the global minimum.

As discussed earlier in chapter 3, all my models were developed by probing the network with different input time resolutions, using 1-hour and 3-hour averages, to know if the prediction accuracy varied with both the integration time (or the cadence) and the number of inputs i.e., inputs with different time histories. We shall go through the results and performance of each algorithm individually in the following subsections.

5.3.1 Model 1: 1-hour lead time Kp predictions from BI

This model took a straightforward approach to build. Our recent success, with only a few false alarms, in providing 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 (e.g. APL Kp models), from a forecast standpoint, if a model using solar wind data alone is essentially as effective as models which also require nowcast Kp, then simpler functions should be used which avoid the concerns about the availability or quantization of the nowcast Kp. Interestingly, in a very recent work, Newell et al. [2008] have shown that the use of a viscous term in addition to a merging term dramatically improves the predictability of geomagnetic indices such as Kp up to 75% (prediction efficiency with $r = 0.866$) without prior knowledge of the target index.

We use equation 3.4 to predict Kp approximately one hour ahead. Given that official Kps are 3-hour averages, data granularity of an hour or less is obtained by splines. The network training and validation was done in the manner described in section 3.4.2 and the outputs have a lead-time of 1 hour. Setting $n = 9$, the best function is written as

$$\mathbf{Kp}_{t+1}^* \equiv \mathbf{f}(\mathbf{BI}_t, \mathbf{BI}_{t-1}, \dots, \mathbf{BI}_{t-9}) \quad (5.1)$$

The results from the exploratory tests (based on randomly chosen test set containing 20457 total points) performed is listed in table 5.1 (and figures 5.12 and 5.13) giving the linear correlation coefficient, the RMSE of the test data, the HSS (with Kp discriminator set at 4), and the ARE. Based on the results, the best outcome was achieved using 10 hours of BI history. The prediction efficiency of the network using the BI as a stand alone to estimate the next upcoming Kp is 93% with a linear correlation coefficient of 0.865 between the real and predicted Kp (figure 5.10). The figure shows the 95% confidence limit as “prediction bounds” within which the pre-

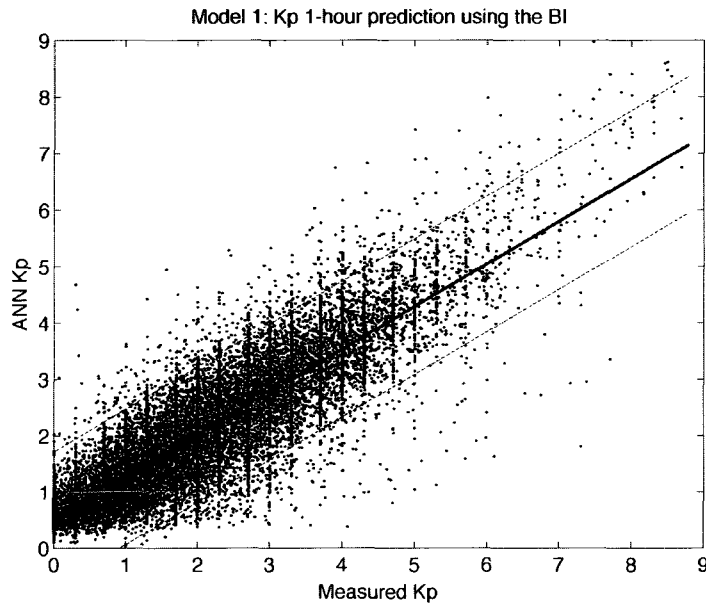


Figure 5.10 : Model 1: ANN Kp vs measured Kp, $r = 0.865$. The dashed lines are the prediction bounds with 95% confidence limit and the thick line is the linear fit.

dictions are likely to fall. It can be seen that the predictions are good for high Kps with a large scatter in the mid-range. Setting the Kp discriminator at 4, there are 1169 hits to 310 misses for a HSS of 0.609. Furthermore, when the Kp discriminator is set at 6, the HSS is found to be 0.564. The histogram (figure 5.11) uses the best function ($n=9$) to show the predicted values against the measured in 28 different bins spanning over the entire range of Kp.

My finding is that, for 1-hour averages, when the time history reached 6 hours, the prediction efficiency barely changed. This finding is consistent with another analysis by Johnson and Wing [2005] and Wing et al. [2005]. Therefore, introducing a sufficient time history guarantees a drastic change in the network dynamics in terms of capturing the nonlinearity in solar wind stream but does little to the prediction efficiency when it gets sufficiently large. This implies that the algorithm is approaching an optimal solution i.e., the larger the inputs, the slower and harder it is to

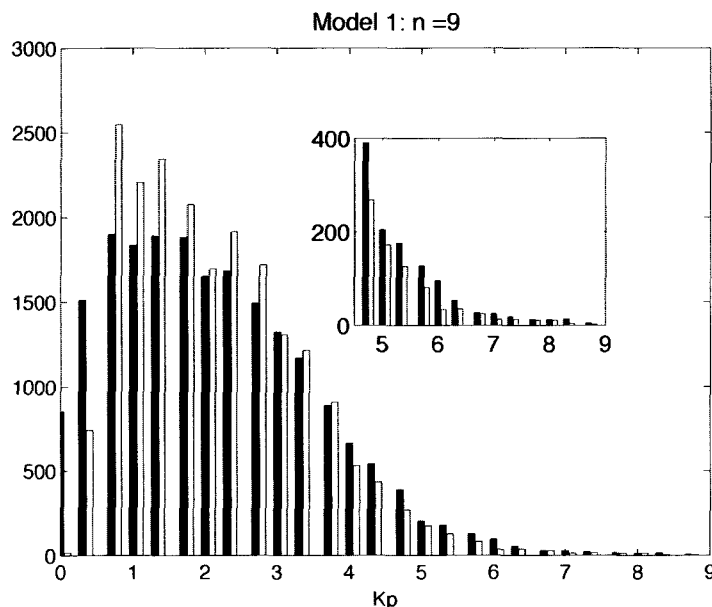


Figure 5.11 : Kp distribution for model 1 using the best function, $n=9$.

minimize the cost function. One possibility is that the time evolution of the data and the time correlations between the solar wind and geomagnetic activity, as seen as a steeply declining trend of cross-correlation functions, is limiting the prediction efficiency. It offers a practical alternative to provide Kp proxies in a timely manner, without concerns about nowcast Kp availability. This model has been running in real time mode for over 2 years now and the predictions can be obtained from <http://space.rice.edu/ISTP/wind.html>.

As the new solar cycle 24, which started in December 2008, is ramping up, it threw a surprise on 22 July 2009 when the 1-hour BI almost reached 200 kV and the corresponding 3-hr BI exceeded 160 kV. The success of our model 1 predictions is shown in figure 5.14 as *white* histograms against the *black* histograms, which represent the measured 1-hr Kp oversampled over the same interval. While the high Kp (>4) are well predicted by the network, it overpredicts Kp below 3, in agreement with the

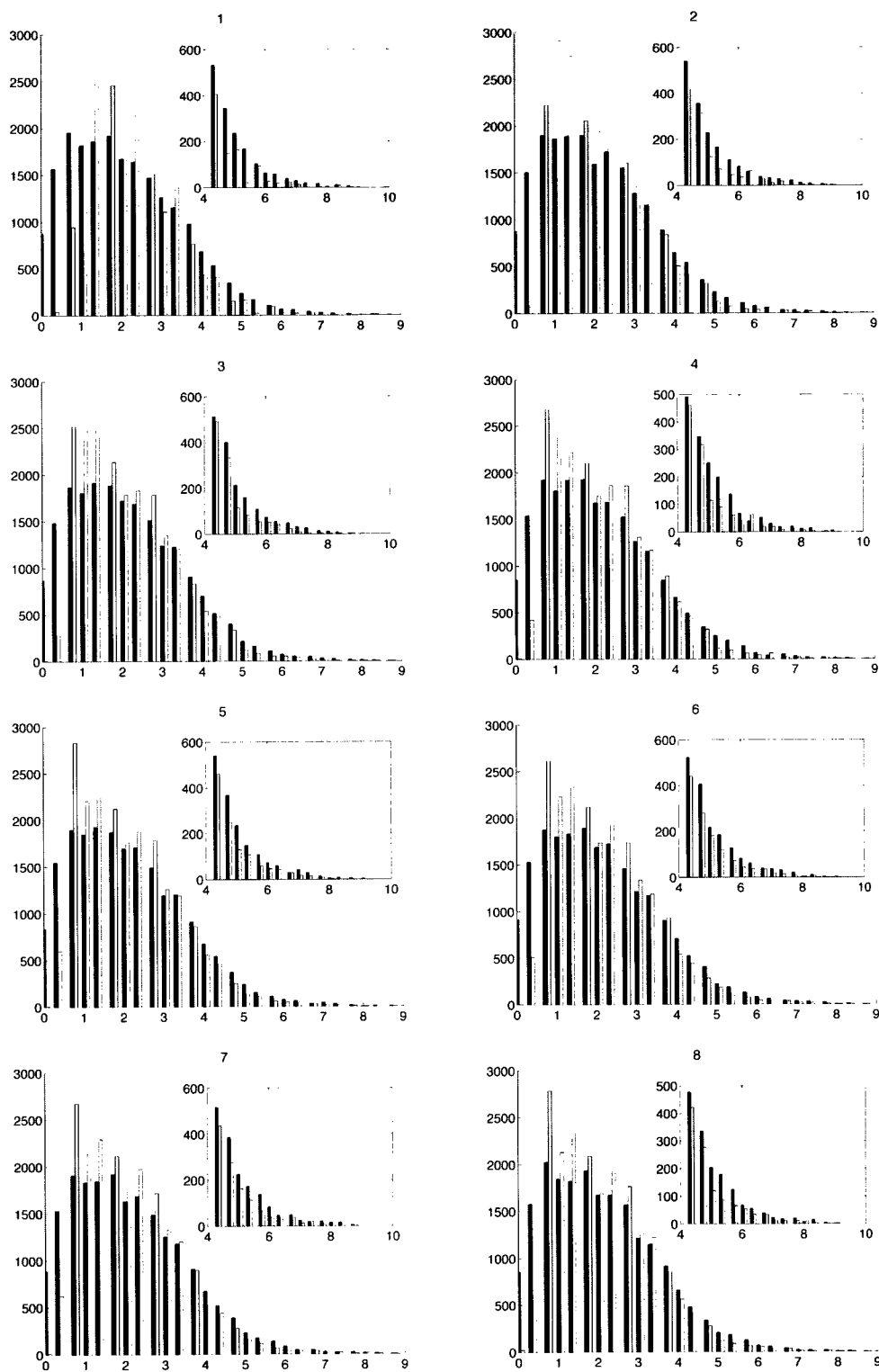


Figure 5.12 : Model 1 test results : Measured (black) and predicted (white) 1-hour Kp distribution with each number representing the number of inputs.

Table 5.1 : Model 1 summary of results: Kp_{t+1}^* from BI

#	Model 1 Inputs	Results				
		Time History	Linear Corr.	RMSE Valid.	HSS (Kp > 4)	ARE (Kp > 4)
1	BI_t	1	0.769	0.896	0.479	0.243
2	BI_t, BI_{t-1}	2	0.825	0.798	0.560	0.194
3	BI_t, BI_{t-1}, BI_{t-2}	3	0.844	0.749	0.569	0.181
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}$	4	0.855	0.729	0.588	0.176
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}$	5	0.859	0.717	0.577	0.182
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}$	6	0.864	0.715	0.595	0.174
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}$	7	0.865	0.713	0.588	0.178
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}$	8	0.859	0.711	0.595	0.193
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}$	9	0.863	0.714	0.609	0.177
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}$	10	0.865	0.703	0.609	0.173

test results in figure 5.10. The statistics for the time interval shown in the figure are: $r = 0.859$, $HSS = 0.81$, $RMSE = 0.82$. The corresponding results based on the 3-hour predictions will be discussed in the following section.

5.3.2 Model 2: 3-hour lead time Kp predictions using only the BI

We saw earlier through 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 difficult to capture and forecast any impending changes shorter than 3-hour duration (as evidenced by the cross-correlation

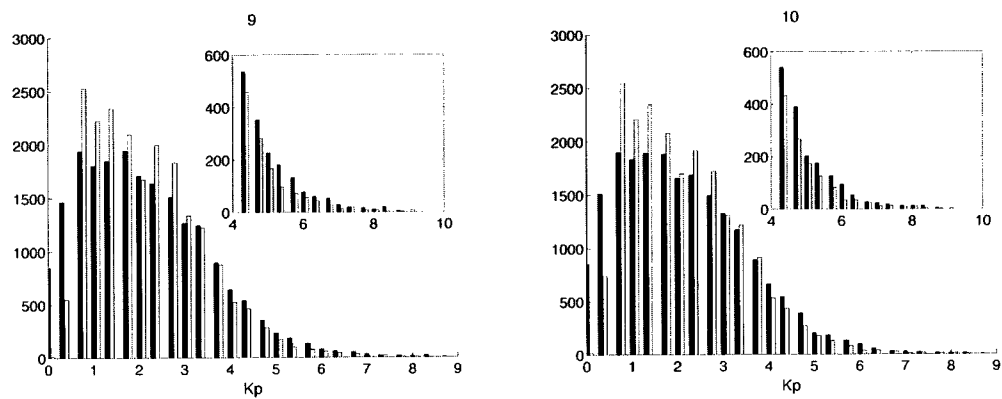


Figure 5.13 : Model 1 test results contd.: Measured (black) and predicted (white) 1-hour Kp distribution with each number representing the number of inputs.

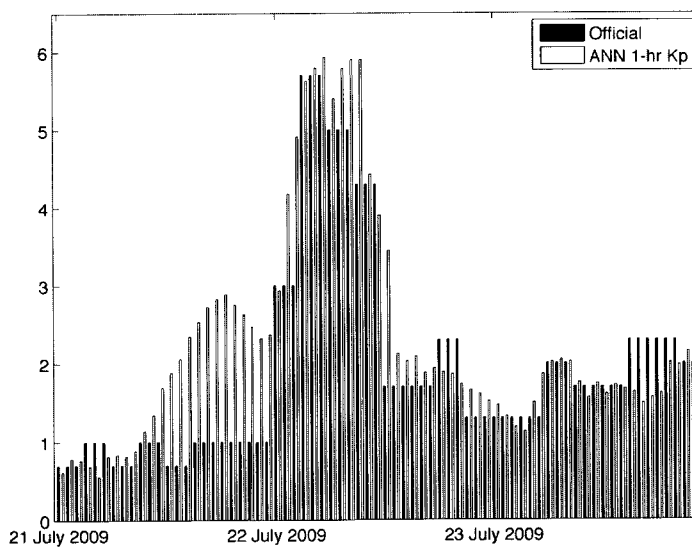


Figure 5.14 : Results of our model 1 predictions from a recent activity.

peaks of Kp, Dst and AE in figures 5.7, 5.8 and 5.9 found between 2-3 hours regardless of the cadence chosen), and therefore, it might seem a potential downside to this model is, perhaps, its time resolution. However, a more practical approach to any advanced warning system is to look for a longer lead-time without compromising too much on the accuracy.

The second model, similar to model 1, inputs solar wind parameters to derive the BI and predicts Kp 3 hours ahead, hereafter called model 2. This model not only extends the forecast range of model 1 but also offers a full 3-hour Kp prediction capability every one hour since the BI is generally available at near-real time. The network was trained using 3-hour averages of the BI to forecast Kp a full 3 hours ahead. The best function is given by,

$$\mathbf{Kp}_{t+3}^* \equiv \mathbf{f}(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-18}) \quad (5.2)$$

Table 5.2 displays the model summary and Kp histograms in figures 5.17 and 5.18. Because of the 3-hour scheme, the longest time history the network was able to look back was 30 hours. However, the network's prediction performance gradually decreased after peaking at 18 hours. n in this case was found to be 5. Therefore, using only 7 inputs and 21 hours of time history, we are able to obtain the best results. The scatter plot of the official Kp versus the ANN Kp test results (5524 points) is shown in figure 5.15, with $r = 0.819$ and $\text{RMSE} = 0.823$ overall, and $\text{HSS} = 0.533$ and $\text{ARE} = 0.218$ for $\text{Kp} > 4$. The corresponding Kp distribution is shown in figure 5.16. From these results, and while the cross-correlations between the BI and Kp are significantly high, it can be said that a true 3-hour Kp can be predicted just as well as model 1. Our retrospective test shows that when Kp exceeds 6 (HSS is 0.571, $\text{FAR} = 0.174$), the hits to miss ratio is 38 to 48. This model may be widely useful, offering an ideal substitute for models 3 and 4 which may need special handling during times

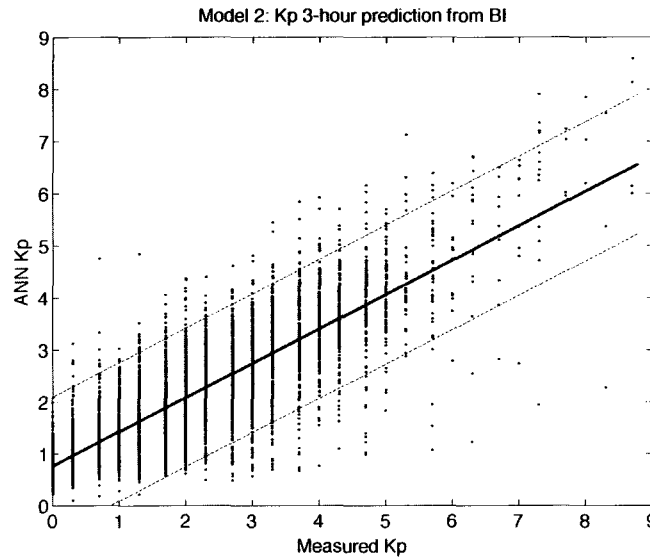


Figure 5.15 : Model 2: 3-hr ANN Kp vs official Kp, $r = 0.819$. The dashed lines are the prediction bounds with 95% confidence limit and the thick line is the linear fit. The model uses 21 hours of solar wind history to make a prediction.

of non-availability of the nowcast Kp. This model is also running in real time mode now at <http://space.rice.edu/ISTP/wind.html> along with the model 1.

The magnetospheric activity observed on 22 July 2009 was also predicted well by our 3-hour BI model, clearly demonstrating its capability as a true forecasting tool. Figure 5.19 describes the results in 3 separate panels. The top panel shows the history of the 1-hour averaged BI, the second panel compares the model 2 predictions against the NOAA/SWPC nowcast Kp, which is issued in near-real time, and lastly, the bottom panel compares the official Kps against our model predictions. The statistics for the time interval shown in the figure are: $r = 0.842$ and $RMSE = 0.84$ overall, and $HSS = 1.0$ for $Kp > 4$.

The next two subsections introduce a new set of models that also incorporates the trailing Kp history which may be obtained from any nowcasting algorithms. We will further discuss their advantages and disadvantages.

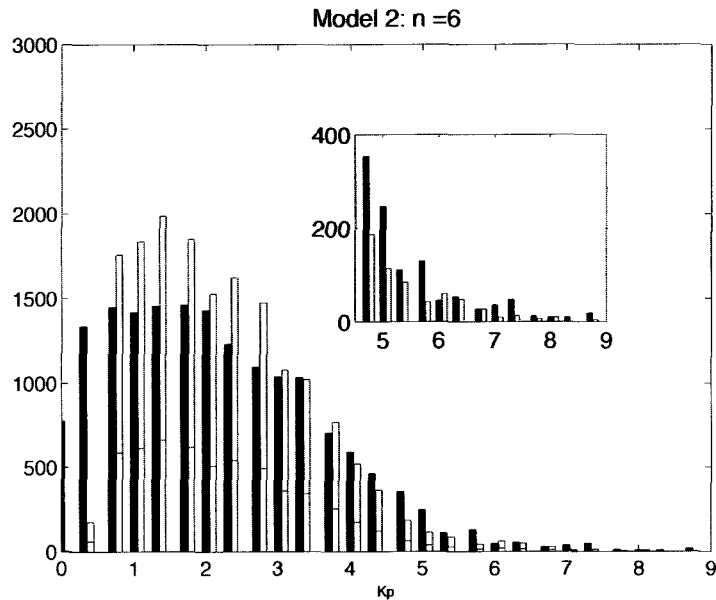


Figure 5.16 : Kp distribution for model 2 using the best function, $n=6$.

5.3.3 Model 3: 1-hour lead time Kp predictions using the BI and Kp history

One of the models I developed takes the time history of Kp and the derived BI as its inputs, hereafter model 3, to predict Kp approximately one hour ahead. A similar model was discussed earlier: the APL model 1 which takes the Takahashi Kp algorithm (section 1.6.4), and which was shown to have a great success. As motivated earlier, the purpose of having a 1-hour Kp prediction model is that shorter time resolution 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. 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, which can

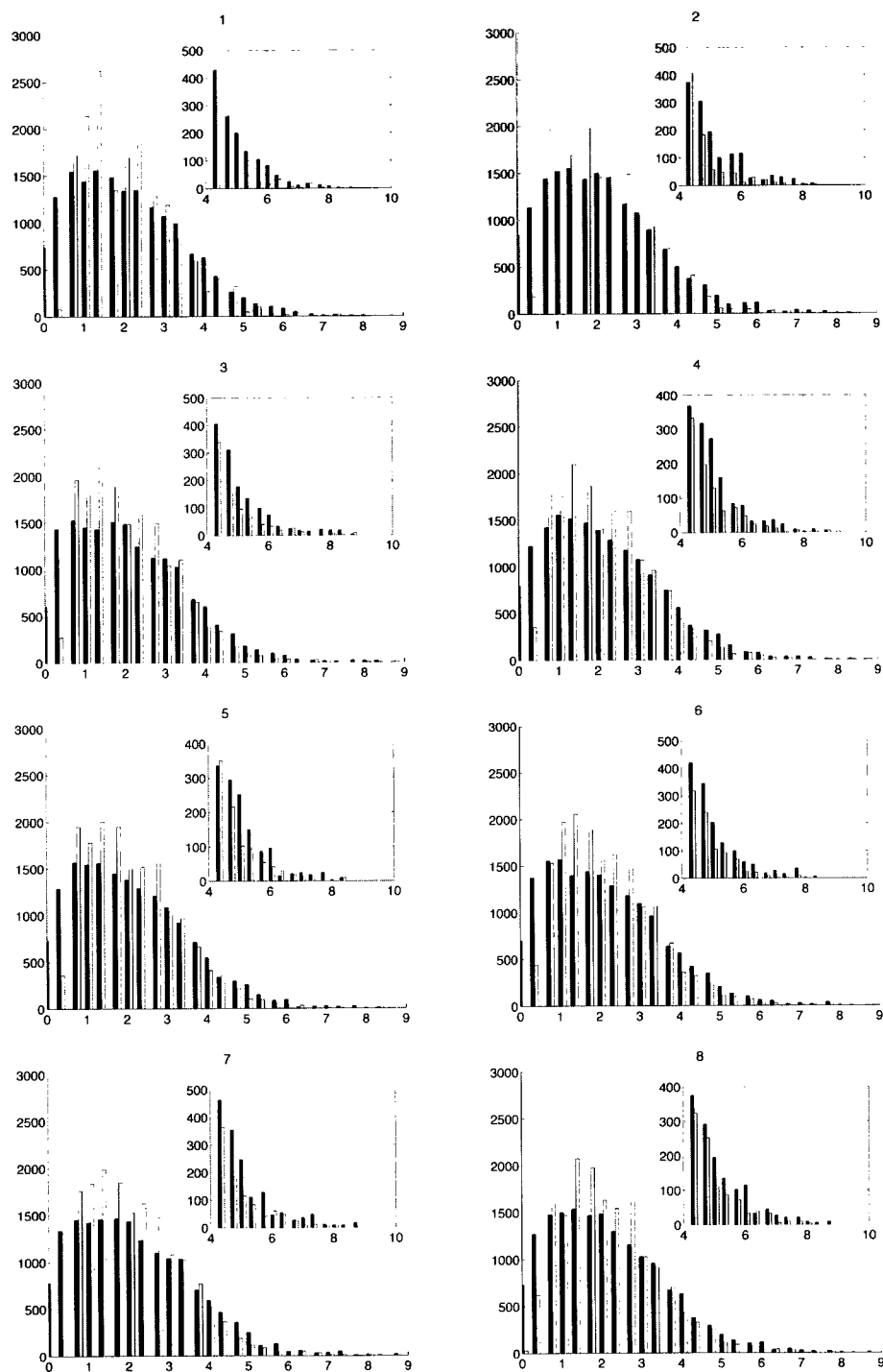


Figure 5.17 : Model 2 test results: Measured (black) and predicted (white) 3-hour Kp distribution with each number representing the number of inputs. Note: since these are 3-hour averages, the distribution has been multiplied by a factor of 3 for easy plot comparison.

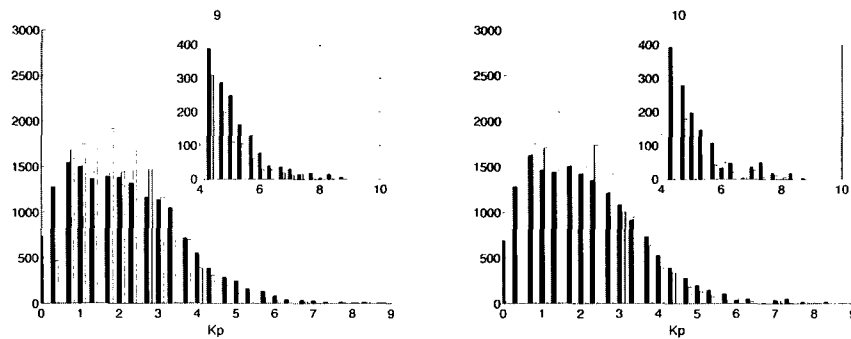


Figure 5.18 : Model 2 test results: Measured (black) and predicted (white) 3-hour Kp distribution with each number representing the number of inputs. Note: since these are 3-hour averages, the distribution is multiplied by a factor of 3 for easy plot comparison.

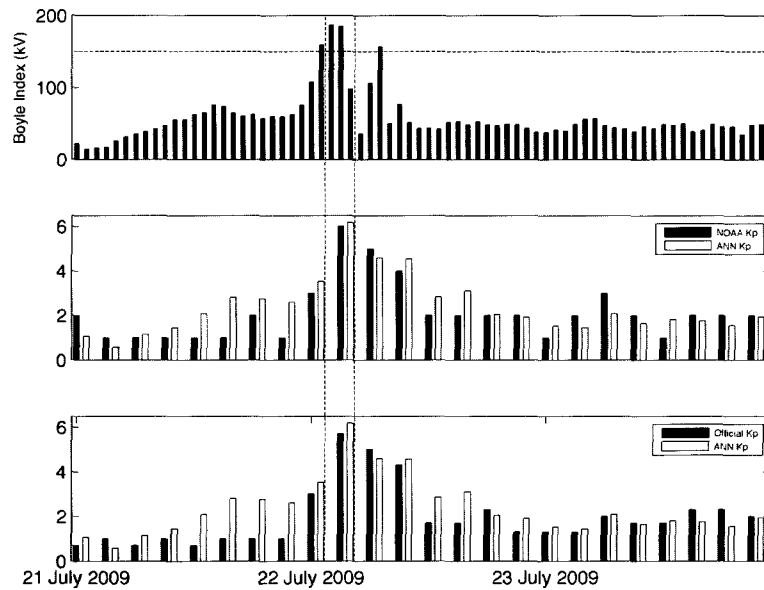


Figure 5.19 : Results of our model 2, 3-hour ahead predictions, from a recent activity. The bottom panels compare the NOAA Kp and the official Kp with the model 2 predictions (white histograms).

Table 5.2 : Model 2 summary of results: Predict Kp_{t+3}^* from BI

#	Model 2 Inputs	Results				
		Time History	Linear Corr.	RMSE Valid.	HSS (Kp > 4)	ARE (Kp > 4)
1	BI_t	3	0.792	0.858	0.490	0.226
2	BI_t, BI_{t-3}	6	0.804	0.842	0.499	0.239
3	BI_t, BI_{t-3}, BI_{t-6}	9	0.820	0.813	0.503	0.219
4	$BI_t, BI_{t-3}, \dots, BI_{t-9}$	12	0.818	0.822	0.514	0.217
5	$BI_t, BI_{t-3}, \dots, BI_{t-12}$	15	0.819	0.808	0.545	0.224
6	$BI_t, BI_{t-3}, \dots, BI_{t-15}$	18	0.815	0.823	0.511	0.218
7	$BI_t, BI_{t-3}, \dots, BI_{t-18}$	21	0.819	0.846	0.533	0.201
8	$BI_t, BI_{t-3}, \dots, BI_{t-21}$	24	0.820	0.823	0.521	0.219
9	$BI_t, BI_{t-3}, \dots, BI_{t-24}$	27	0.814	0.838	0.509	0.233
10	$BI_t, BI_{t-3}, \dots, BI_{t-27}$	30	0.811	0.827	0.519	0.253

be obtained from http://www.swpc.noaa.gov/rt_plots/kp_3d.html; these are not true depictions of the actual Kp. There is, however, a 30-40 minute lag before NOAA makes them 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 target Kp follows the *best* function below

$$Kp_{t+1}^* \equiv f(BI_t, BI_{t-1}, \dots, BI_{t-5}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-8}) \quad (5.3)$$

where, again, the time t , $t-1$, etc. in each case represents the end of the integration time period. Therefore, equation 5.3 is very similar to equation 5.1 except for the

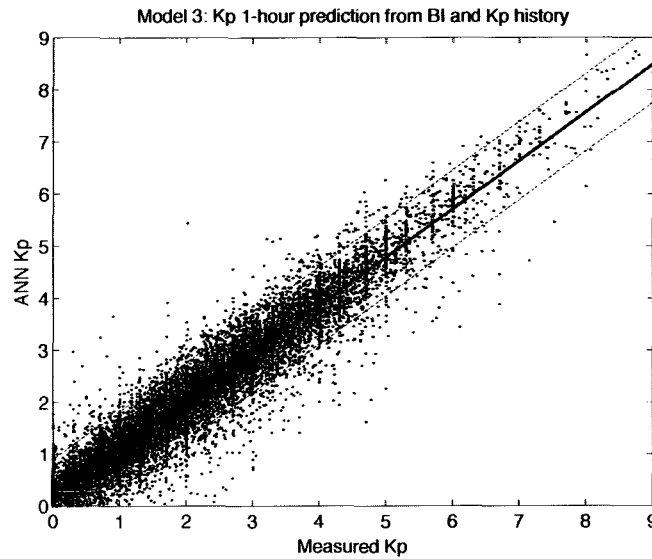


Figure 5.20 : Model 3: 1-hour ahead ANN predicted Kp vs measured Kp. Linear correlation, $r = 0.960$. The network uses 6 hours of input history of each BI and Kp. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

Kp inputs. The BI and Kp inputs are delayed in time by three hours and the ANN has been trained to handle them as individual inputs i.e., for a set of BI, there is a corresponding set of Kp that *lags* the BI by 3 hours. 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 from ACE. With the first used Kp clearly delayed relative to the BI, the BI serves a precursor of rises or fall in the Kp to come leading up to the predicted time. The most recent Kp value used is $(t-3)$, assuring that the network is not trained to require a value of Kp which is not available in real time. The same scheme in equation 5.3 may be used to carry out real time predictions.

The network is designed such that it looks 6 hours into the past to make the best predictions (refer tables 3.4 and 5.3 and figures 5.22 and 5.23). It was trained, validated and tested using hourly averages of BI and Kp covering an 11 year period

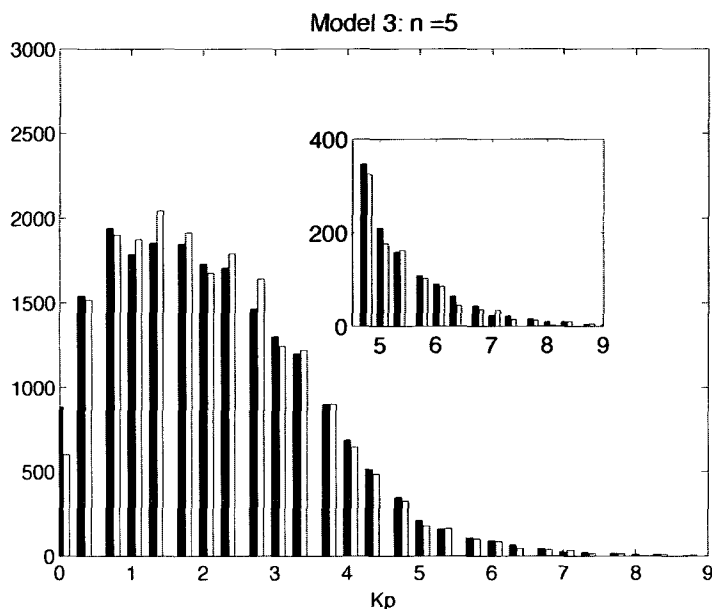


Figure 5.21 : Kp distribution for model 3 using the best function, $n=5$.

from 1997-2007 with data sets classified to avoid overlap. *In all cases, 20461 randomly chosen data points were not part of the training set.* In dealing with any missing values of Kp or the BI, I 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 test data that is completely new to the network along with the time period over which it is tested.

In figure 5.20, I plot the network responses as a function of the official Kp (linear correlation coefficient of $r = 0.960$ and overall $RMSE = 0.393$) with the 95% confidence bounds shown as the dashed lines. The network has, in fact, learned to reproduce over 92% of the variance of the data presented to it through the retrospective test over the whole set. Evidently, adding Kp to the inputs certainly enhances the predictions.

This test strictly follows the training in that it does not consider the most recent K_p . The histogram shown in figure 5.21 complements the scatter plot.

Table 5.3 : Model 3 summary of results: Predict Kp_{t+1}^* from BI and Kp

#	Model 3 Inputs	Results				
		Time History	Linear Correlation	RMSE Valid.	HSS ($Kp > 4$)	ARE ($Kp > 4$)
1	$BI_t; Kp_{t-3}$	3	0.857	0.725	0.602	0.175
2	$BI_t, BI_{t-1}; Kp_{t-3}, Kp_{t-4}$	6	0.883	0.670	0.625	0.160
3	$BI_t, BI_{t-1}, BI_{t-2}; Kp_{t-3}, Kp_{t-4}, Kp_{t-5}$	9	0.900	0.611	0.651	0.148
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-6}$	12	0.907	0.588	0.686	0.142
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-7}$	15	0.916	0.570	0.707	0.130
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-8}$	18	0.960	0.393	0.793	0.089
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-9}$	21	0.924	0.536	0.708	0.132
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-10}$	24	0.918	0.558	0.709	0.134
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-11}$	27	0.940	0.476	0.750	0.111
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}; Kp_{t-3}, Kp_{t-4}, \dots, Kp_{t-12}$	30	0.903	0.607	0.702	0.131

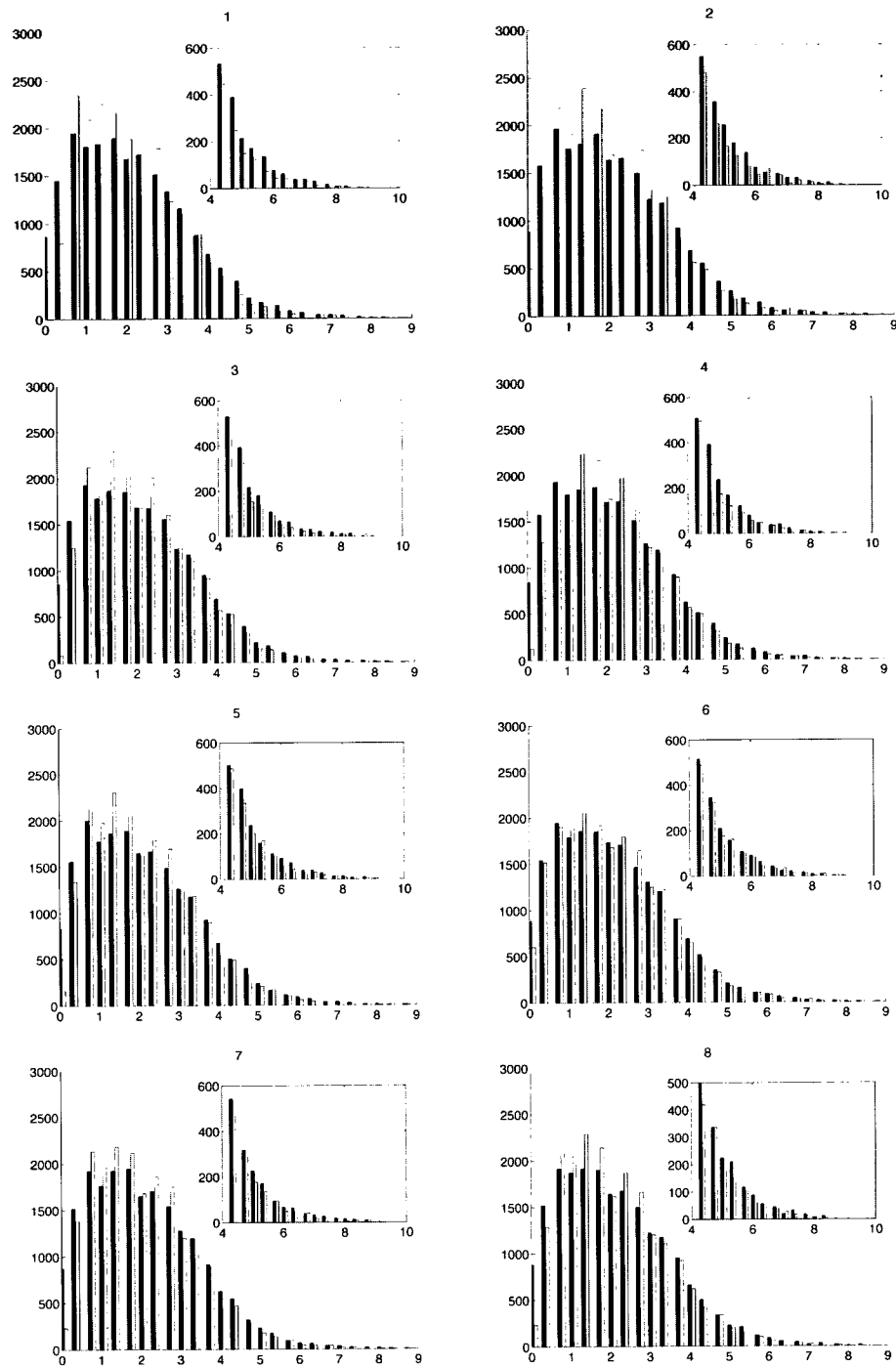


Figure 5.22 : Model 3 test results: Measured (black) and predicted (white) 1-hour Kp distribution. Each number here denotes the number of inputs each of BI and Kp.

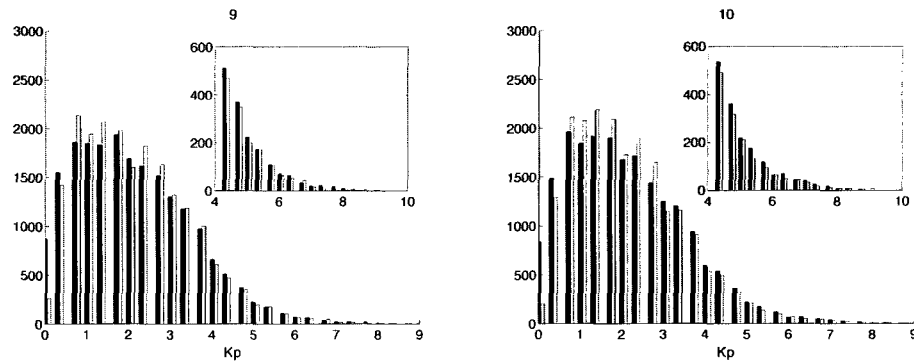


Figure 5.23 : Model 3 test results cont.: Measured (black) and predicted (white) 1-hour Kp distribution. Each number here denotes the number of inputs each of BI and Kp.

Furthermore, from table 5.3, when the observed Kp exceeds 4, the HSS and ARE are found to be 0.793 and 0.089 respectively with hits to misses ratio of 1563:500 (20461 total points). However, when the Kp discriminator is increased to 6, the HSS is found to be 0.748, better than what was seen for model 1 for the same threshold. In this case, the hits to miss ratio is 170:87. The histogram of Kp distribution corresponding to each individual input series is shown in figure 5.17.

Caution must be used while running model 3 in real time because of the following 2 factors: (1) Crucial time is lost towards obtaining the nowcast Kp, and (2) the real time model uses the NOAA nowcasted Kp, which is only estimated to the nearest unit (varies from 0 to 9 in 9 increments unlike the official Kp which increments in steps of 0.3), and since the network training and model evaluations were based on the official Kp record, the network performance is likely to diminish slightly. For example, on 22 July 2009, during the 3-hour interval when the measured Kp peaked at 5.7, the NOAA Kp was measured to be 6.0. Therefore, models utilizing the NOAA Kp can give out predictions which may be an under- or overestimate.

5.3.4 Model 4: 3-hour lead time Kp predictions using the BI and Kp history

According to the tests performed in the preceding section, it is apparent that the ANN has a better prediction efficiency while using a 1-hour lead-time that includes the time history of Kp as opposed to the model that does not. We have seen earlier that the BI correlates with Kp better for 3-hour averages (figure 5.1). Therefore, just like model 1, it is a viable and reasonable option to train the network to predict Kp approximately 2 hours ahead, in spite of the fact that we lose 45 minutes to an hour approximately owing to processing delays.

The network inputs and training are similar to the description in section 3.4.5 and the best working function is written as:

$$\mathbf{Kp}_{t+3}^* \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-18}; \mathbf{Kp}_{t-3}, \mathbf{Kp}_{t-6}, \dots, \mathbf{Kp}_{t-27}) \quad (5.4)$$

where t denotes the current epoch, and $t-3$ etc. means actual 3 hours behind the current epoch t and so on. Given the availability of 3-hour nowcast Kp and live updates of the BI, this design should have the network deliver a prediction for the next upcoming 3-hour Kp. Assuming a delay of approximately 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 3-6 UT, using the BI up to 03:55 and the previous Kps up through 0-3 UT. If for any reason the 0-3 UT Kp is not available, we duplicate the previous Kp; however, this has not been necessary at least until now and since October 2007, the time we began using the real time Kp from NOAA. Table 5.4 lists the model 4 prediction summary and figures 5.26 and 5.27 shows the individual histograms. A direct comparison of the model outputs with the true Kp shows a drop in the prediction efficiency when the lead-time is extended from 1 to 3 hours.

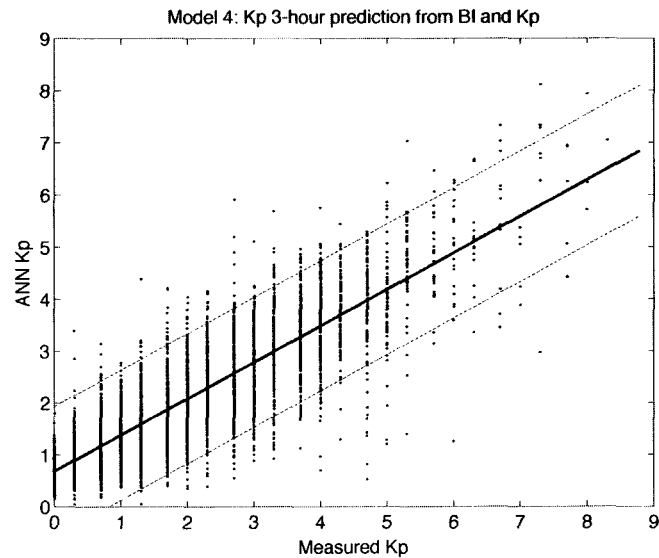


Figure 5.24 : Model 4: 3-hour ahead ANN predicted Kp vs measured Kp. Linear correlation, $r = 0.839$. The ANN uses 27 hours of input history of each BI and Kp. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

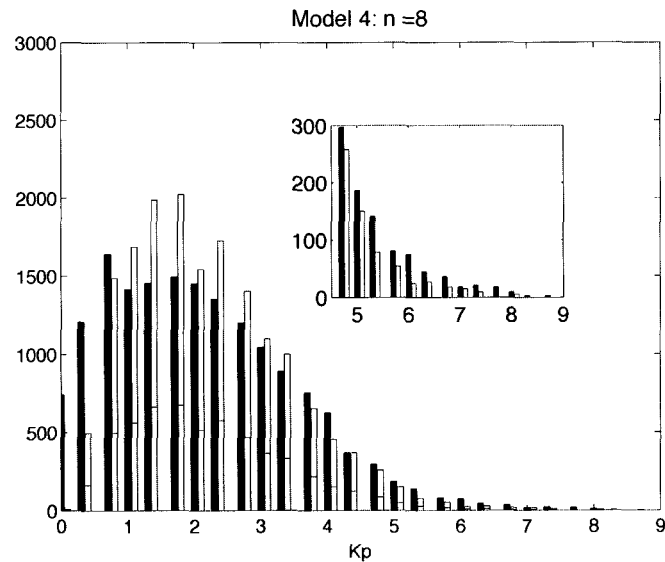


Figure 5.25 : Kp distribution for model 3 using the best function, $n=8$.

Table 5.4 : Model 4 summary of results: Predict Kp_{t+3}^* from BI and Kp

#	Model 4 Inputs	Results				
		Time History	Linear Correlation	RMSE Valid.	HSS ($Kp > 4$)	ARE ($Kp > 4$)
1	$BI_t; Kp_{t-3}$	3	0.837	0.785	0.513	0.192
2	$BI_t, BI_{t-3}; Kp_{t-3}, Kp_{t-6}$	6	0.828	0.799	0.529	0.211
3	$BI_t, BI_{t-3}, BI_{t-6}; Kp_{t-3}, Kp_{t-6}, Kp_{t-9}$	9	0.835	0.777	0.560	0.209
4	$BI_t, BI_{t-3}, \dots, BI_{t-9}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-12}$	12	0.835	0.781	0.551	0.212
5	$BI_t, BI_{t-3}, \dots, BI_{t-12}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-15}$	15	0.839	0.771	0.518	0.210
6	$BI_t, BI_{t-3}, \dots, BI_{t-15}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-18}$	18	0.842	0.767	0.544	0.203
7	$BI_t, BI_{t-3}, \dots, BI_{t-18}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-21}$	21	0.839	0.769	0.549	0.200
8	$BI_t, BI_{t-3}, \dots, BI_{t-21}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-24}$	24	0.844	0.771	0.566	0.207
9	$BI_t, BI_{t-3}, \dots, BI_{t-24}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-27}$	27	0.839	0.767	0.565	0.195
10	$BI_t, BI_{t-3}, \dots, BI_{t-27}; Kp_{t-3}, Kp_{t-6}, \dots, Kp_{t-30}$	30	0.840	0.769	0.574	0.204

The linear correlation coefficient between the real and ANN Kp is found to be 0.839 with a prediction efficiency of 0.92 (figure 5.24 and corresponding histogram in figure 5.25), outperforming the Kp self predictor ($r=0.790$); a self predicting Kp algorithm (see section 5.3.6) is an ANN-based algorithm that predicts Kp from the time history of Kp only, providing an extra benchmark for the models based on BI, and BI and Kp. As expected, the network estimates did not improve with longer time histories.

Setting the Kp threshold at 4, the HSS is 0.565. Now, if we went from Kp 4 to Kp 6, our test indicates a drop in the HSS (0.487) and FAR (0.133) over model 2 (HSS is 0.571 and FAR = 0.174). Therefore, while there are no significant difference in the prediction performances in models 2 and 4, it is safe to say that models using only the BIs as inputs are just as efficient as those that uses the time history of both BI and Kp (figure 5.19 also substantiates). In the absence of nowcast Kp data, model 2 is a great substitute for model 4. Even though these results may not comply with what were expected from the cross-correlation analysis, the predictions from a 3-hour average of the BI are only slightly worse than the predictions from 1-hour averages. What is really important is their accuracy in predicting geomagnetic activity and how the results can be quantified. Figure 5.28 summarizes the overall results of the four Kp models. The following subsection will lead us into discussing “ideal response” scenarios in two separate and interesting ways.

Finally, as another reminder, a similar caution to the one noted in model 3 is in order while running model 4 in real time i.e., the official Kps are different from the NOAA nowcast Kps which may skew the predictions and there are processing delays in obtaining the NOAA Kp.

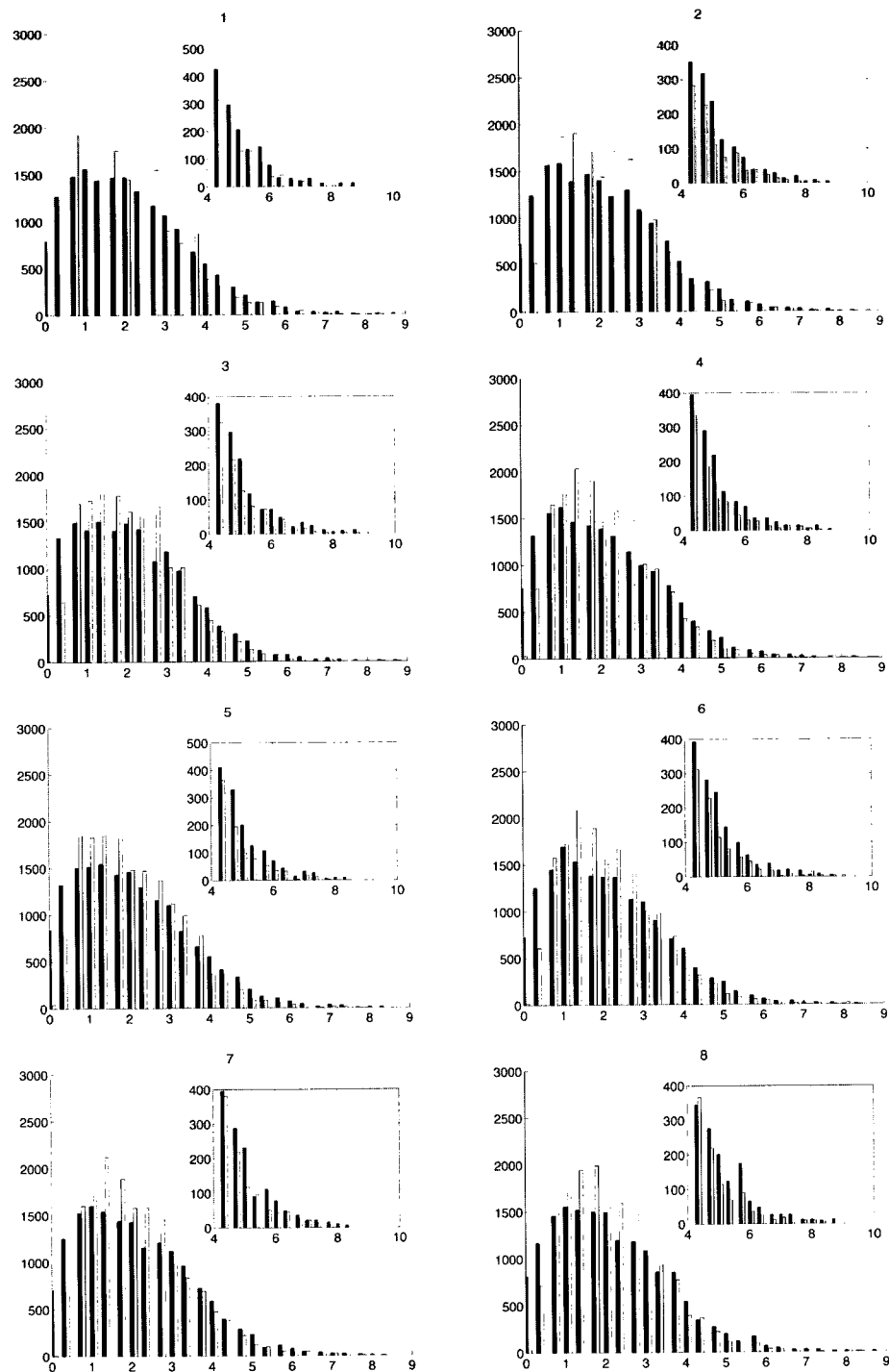


Figure 5.26 : Model 4 test results: Measured and predicted 3-hour Kp distribution with each number here denotes the number of inputs each of BI and Kp. Note: the distribution is multiplied by 3 for easy of plot comparisons.

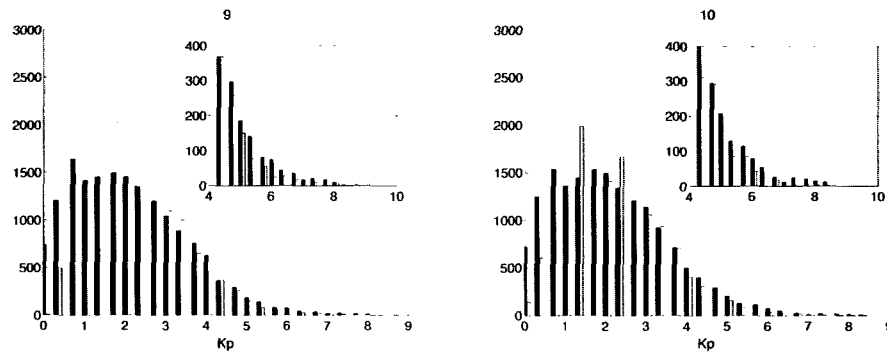


Figure 5.27 : Model 4 test results contd.: Measured and predicted 3-hour Kp distribution. Note: the distribution is multiplied by 3 for the sake of plot comparisons.

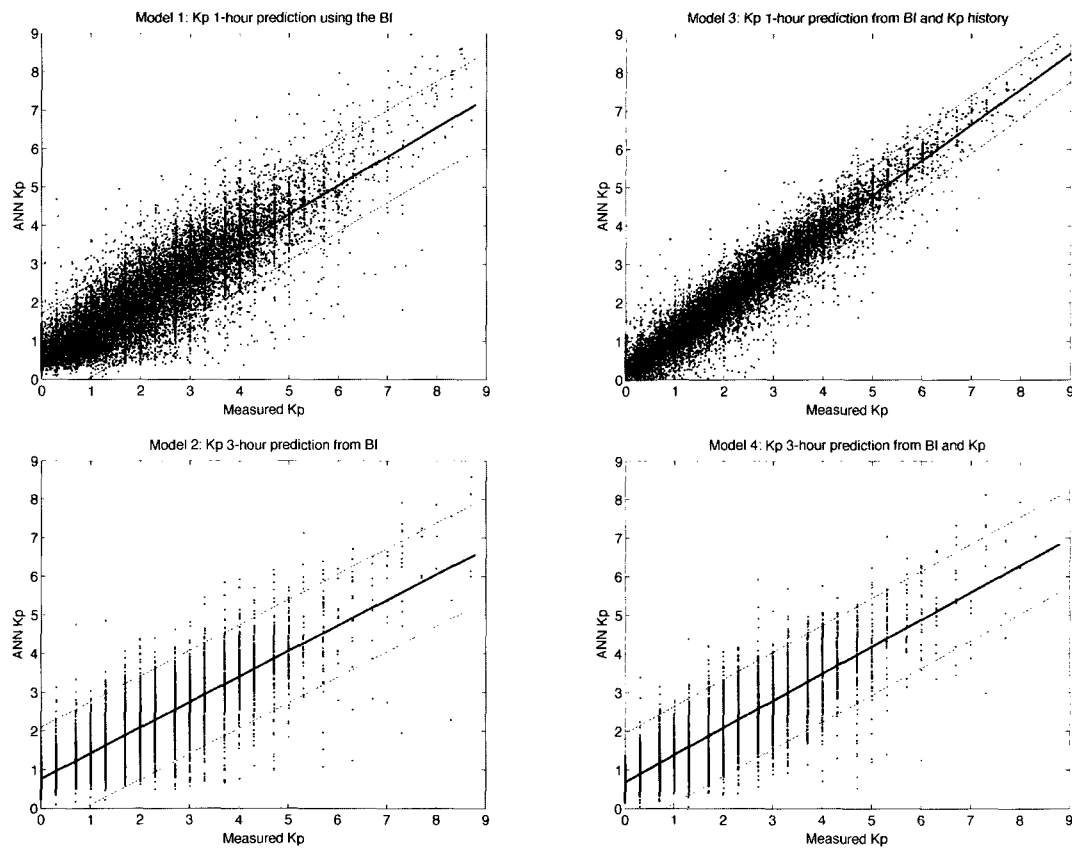


Figure 5.28 : Summary of the best results from the 4 Kp models. Model 1: $r=0.865$, Model 2: $r=0.819$, Model 3: $r=0.960$, Model 4: $r=0.839$. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

5.3.5 Kp models: Understanding accuracy

It is often difficult if one were to make judgments regarding the ANN predictions based on the scatter plots alone. Sometimes, such descriptions would be inept and inaccurate, given the scatter in the results. To look at results more closely, figures 5.29 and 5.30 presents a different perspective of the model performance by displaying the predicted and official values over-plotted for randomly chosen 30 day periods each from 2006 (December) and 2001 (April). Actually, I deliberately chose these two periods to study the network responses as they provide different instances to study a few major storms; April 2001 can be classified under increasing and maximum phases of solar cycle 23 [Cane and Richardson, 2003] while December 2006 (the Kp index reached 8 while the Dst index dipped low to a value of -146 nT) occurs during the depths of quiet solar minimum conditions of solar cycle 23. Affected customers, because of the December 2006 activity driven by a series of solar flares, included numerous satellites (e.g., GOES 13, ACE, Cassini), NASA (astronauts aboard the International Space Station and shuttle Discovery remained inside a protected area of their spacecraft on 5 Dec. 2006 to mitigate the risk), commercial airlines (significant impacts to the network of air traffic control radars in Canada, causing interference and false targets of approximately 150 miles in length) and power grids [Source: NOAA/SWPC].

For the purpose of model verification, and because a good prediction is what is sought after, I have listed the results of models 1 through 4 individually. For a measured Kp of 6, at a certain time interval 0300-0400 UTC, if the models had predicted a value 6 or higher at 0300 UTC then it is considered as a “hit”. I have shown the results of the model predictions from the two chosen time intervals in tables 5.5 (April 2001) and 5.6 (December 2006), and the actual predictions themselves in

Table 5.5 : Geomagnetic Storm Alerts ($K_p > 6$).

Period Covered	April 2001			
Kp Model	1	2	3	4
Hits	19	6	27	6
Misses	20	8	12	8
False Alarms	2	0	6	1
Right Rejections	709	236	705	235
Total	750	250	750	250
Maximum Lead Time	1 hours	3 hours	1 hour	3 hours
Heidke Skill Score	0.620	0.586	0.738	0.555
Probability of Detection	0.487	0.429	0.692	0.429
False Alarm Rate	0.095	0.000	0.182	0.143

figures 5.29 and 5.30. In describing predictions, the χ^2 equivalent of goodness-of-fit can be characterized using the set of parameters defined for a 2x2 contingency table (section 3.5.1).

For the time period covered in April 2001, the best outcome is seen for model 3. There have been 27 hits to 12 misses for the base interval of 1 hour. Out of those misses, at least 8 of them have been predicted to be K_p 5 or above and some of them resulted in an extended alert because the previous predicted K_p value had been over 5. The overall skill score is found to be 0.738 with a false alarm probability of 0.095. For the period covered in December 2006, the best results is seen for model 3. The overall skill scores are 0.796 for the HSS with a false alarm probability of 0.091. On

Table 5.6 : Geomagnetic Storm Alerts ($K_p > 6$).

Period Covered	December 2006			
	1	2	3	4
Kp Model	1	2	3	4
Hits	7	2	10	2
Misses	7	3 (4-, 5+, 5)	4	3 (4-, 6-, 6-)
False Alarms	1	0	1	1
Right Rejections	685	230	685	229
Total	700	235	700	235
Maximum Lead Time	1 hours	3 hours	1 hour	3 hours
Heidke Skill Score	0.631	0.566	0.796	0.492
Probability of Detection	0.500	0.400	0.714	0.400
False Alarm Rate	0.125	0.000	0.091	0.333

the other hand, the performance level diminishes slightly for the base interval of 3 hours ($HSS = 0.429$ and 0.566 for the two periods in April 2001 and December 2006 respectively), but we get a maximum lead time of 3 hours.

Judging from the network predictions, one can infer that the storms (with $K_p \geq 5$ of interest) have been well predicted both during the sudden commencement of the storm and the recovery phases (refer to the time interval between the two vertical dashed lines in figure 5.30). Furthermore, we can also quantify the rate of success by graphically representing the hits versus misses using absolute error ($|K_p - K_p^*|$) or relative error ($\frac{|K_p - K_p^*|}{|K_p|}$) measures, though neither of them can distinguish between overshoots and undershoots i.e., predicted values falling above or below the measured.

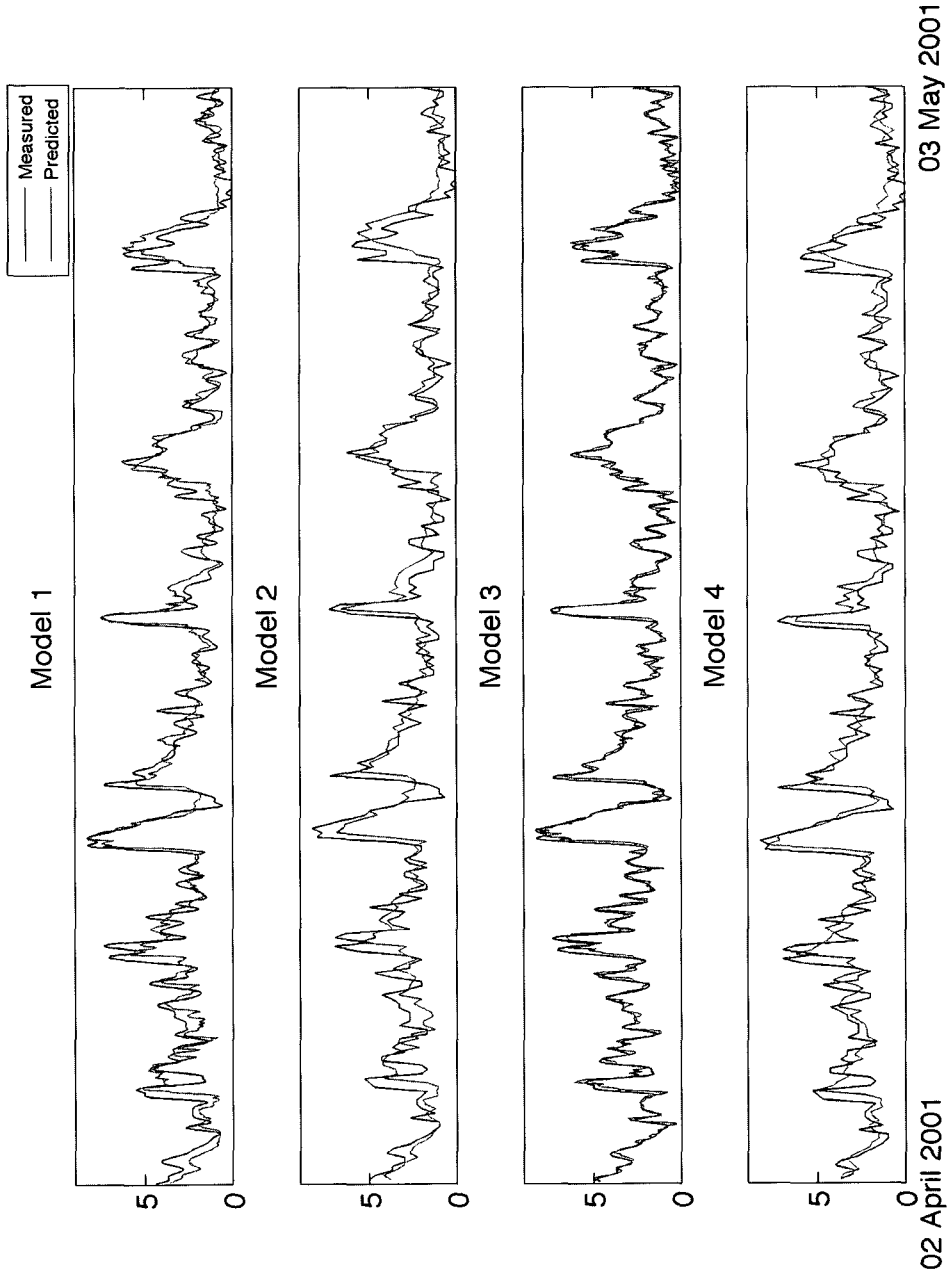


Figure 5.29 : Model vs. data comparison: Predictions from all the Kp models are overplotted with the measured Kp. Data shown here covers all of April 2001.

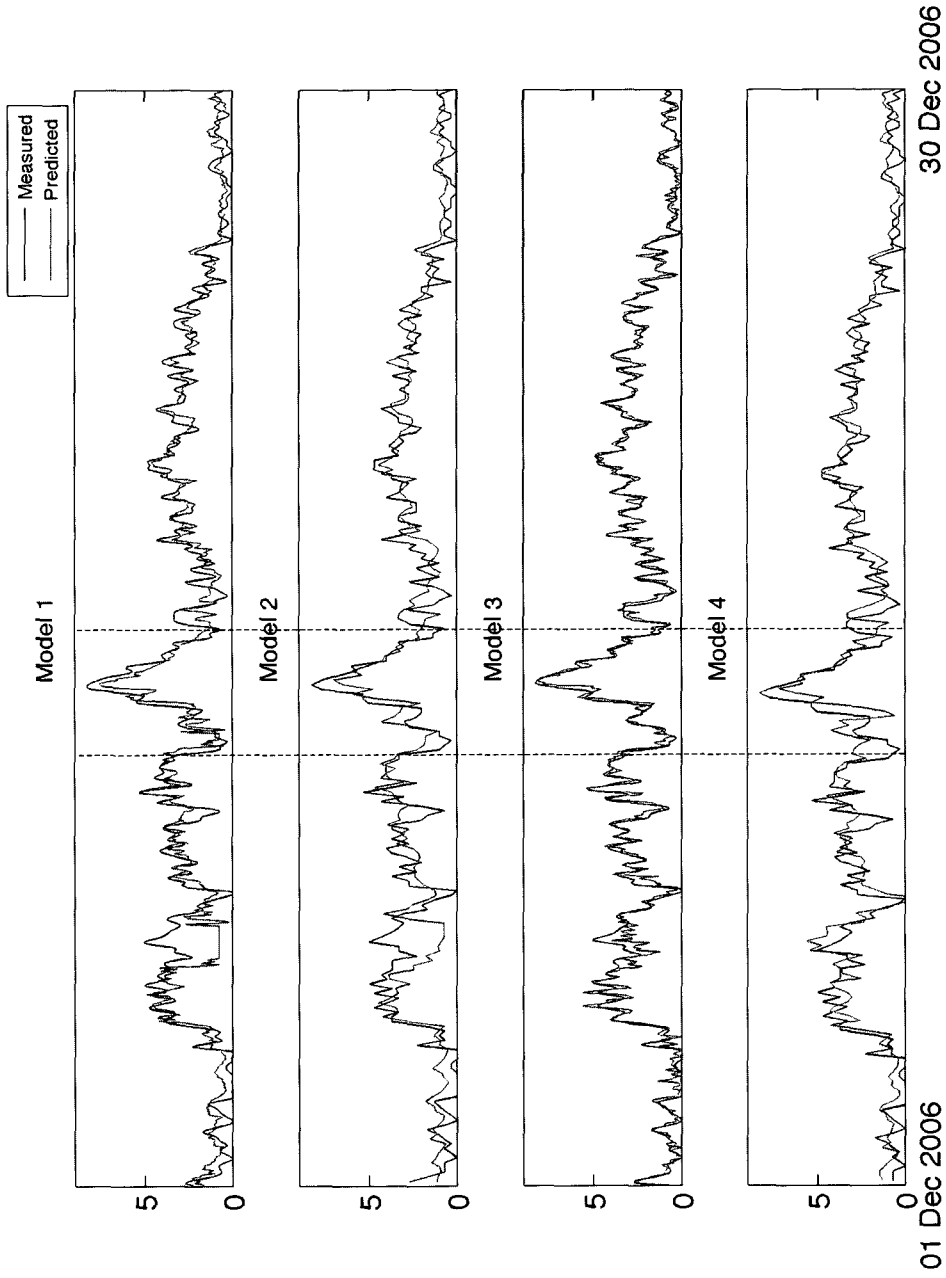


Figure 5.30 : Model vs. data comparison: Predictions from all the Kp models are overplotted with the measured Kp. Data shown here covers all of December 2006. See appendix (page 208) for a closer view of the data within the dotted lines.

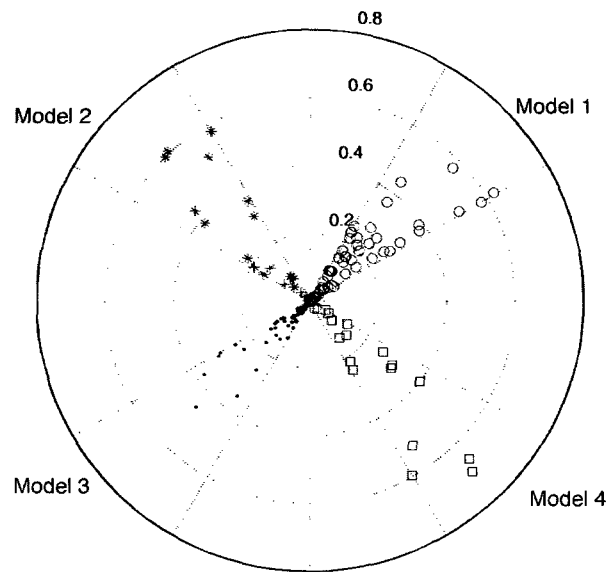


Figure 5.31 : The “bulls-eye” plot: Shown here are the relative errors, measured for every measured $K_p > 6$. The results from the models are shown in separate sectors.

Figure 5.31 combines the results the model predictions during the time intervals of April 2001 and December 2006. We see a clear difference in the prediction performances of the two 1-hour K_p models. Model 3, which includes the K_p history, is significantly better than model 1. On the other hand, for the 3-hour models, interestingly, performance statistics highly favor model 2 i.e., the BI only model. In general, barring a few pitfalls, all these models have demonstrated their prediction capabilities to cater to different needs and we constrained well at the mid- to higher-range of K_p s. However, one cannot clearly make a call as to whether these are “true predictions/forecasts” or a mere “persistence”. The following section investigates “persistence” in detail.

5.3.6 “Persistence” vs. “true” forecasting

We introduced the idea of persistence earlier in chapter 1. To recapitulate, persistence is defined as the statistical dependence among successive terms of the same variable or an event [Wilks, 1995]. Let us study the following two cases, both using 3-hour averages: (1) a Kp time series, and (2) a time series involving the BI. At zero lag, the autocorrelation is of the data stream with itself is always exactly unity (figure 5.32). For the Kp in panel 2 of figure 5.32, at the next time step of 3 hours, it shows an autocorrelation at 3 hours of 0.80. That is, any predicting algorithm can predict with 80% efficiency the upcoming Kp by just saying “it will the same as the last one”. Thus any prediction algorithm that uses prior knowledge of Kp must have a correlation coefficient of greater than 0.80 or it does not beat the “persistence” prediction. On the other hand, for those algorithms that do not use prior Kp data, an 80% correlation coefficient is good. Thus, our one-hour prediction of $r=0.888$

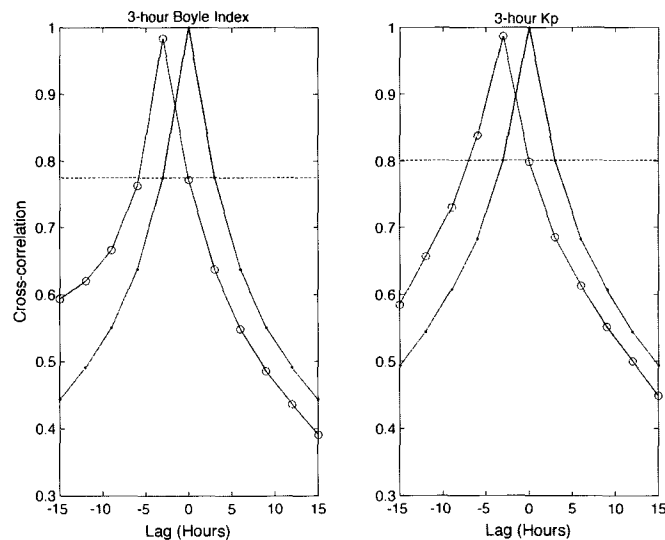


Figure 5.32 : Auto-correlation for 3-hour BI (left panel) and 3-hour Kp (right panel) and cross-correlation for the self predicting BI and Kp models are shown here. Note that both the models (circles) lag the auto-correlation curves (smooth curve).

(model 3) and three-hour ahead predicted Kp with $r=0.830$ (model 4) are significantly better than the “persistence prediction”. This can be demonstrated using the tests of significance (see section 5.5).

On the other hand, if someone has an algorithm that produces persistence-based predictions and then does a cross-correlation between their predictions and the real Kp, then their “predictions” will show a maximum cross-correlation at -3 hours. For example, a self predicting ANN-based Kp algorithm, an algorithm that uses the time history of the Kp index (here, previous 5 Kp values) to predict the upcoming Kp, has a correlation of 0.987 at -3 hours and 0.80 at the zero-lag (shown as the lagged curve in panel 2); the self predicting Kp algorithm may be a slight improvement over the linear predictor (section 1.3), but it is only as good as persistence forecasting. That is, their prediction will “lag” the real data, because their predictions are too strongly influenced by the value of the previous Kp. This effect on the model outputs is known as “persistence contamination”. In the second example, we investigate the effect of persistence through the BI, by constructing a time series of the solar wind. The maximum correlation for this ANN-based self predicting BI algorithm, using the previous 5 BI values, occurs at a time lag of -3 hours (0.983) and the zero-lag, the correlation is 0.771 (panel 1, figure 5.23), which equals the estimated limit for persistence. Again, the predictions clearly “lag” the true values. When it happens they are really *not* true predictions.

5.3.7 Post-test Correlation Analysis

Model evaluations and performance indicators based on statistical tests such as skill scores, error-bars, linear regression analysis are often sufficient to describe a forecast. They are as good as their ability to answer questions such as how far a “miss” strayed away from a “hit”? or, how close was a hit to getting a perfect score. However, having

seen what persistence implies, models using the known values of its target index should not rule out the possibility of a prediction weighing too much on the last known value. In a time series forecasting, only post-test analysis based on cross-correlation can offer deeper insights. Many existing forecast models found in the literature tend to have ignored this fact, however (e.g., the APL model 1 described earlier). This section is intended to show how my solar wind-only driven models, and my analysis thereof, improves the understanding of the existing work.

I show the cross-correlation of the predictions versus the observations in terms of its time lags for the models 2 and 4 in figure 5.33. It also shows the autocorrelation 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 high (0.80) autocorrelation at 3 hours, showing the persistence of Kp from one 3-hour measurement to the next, setting the standard for effectiveness of prediction algorithms. The results of the cross correlation of our model 2 and model 4 predictions with the official Kp are shown here. Both of my model predictions at lag=0 are significantly above 0.80 (0.819 for model 2 and 0.839 for model 4), showing effectiveness in beating the persistence hypothesis i.e., they are effective in predicting faster and slower rate of changes between any two successive measurements. For those curves, a lag=0 means that we are correctly predicting the following Kp value at the proper time, assuming the cross-correlation coefficients are higher than the persistence value at that lag.

Digression: Statistically, one of the ways to ascertain the significance of a single correlation coefficient is through a t-test. For example, in model 2, for a sample size of 5821, the t-test yields a good significance at the 1% probability level for both $r = 0.819$ and the auto-correlated Kp at $r = 0.800$. Alternatively, probabilities can also be used to show their significance and, given the sample size of this example, the probability that they are uncorrelated is a mere $< 0.5\%$. In other words, their

single correlation coefficient is statistically significant. Now, in order to show the significance of $r = 0.819$ from a specified value of, say, $r = 0.800$, we use the Z-test. A Z-test was performed on model 2 to investigate the significance of its correlation coefficient ($r = 0.819$) from the auto-correlated Kp at $r = 0.800$; both the values are derived from the same distribution with independent variance. We start with the “null hypothesis” i.e., the two correlations are not significantly different from each other. Performing the Z-test, we get a Z-value of 4.10 from 5530 test samples. To reject the null hypothesis, assuming 95% confidence limit, we need a Z-value of 1.96. Obviously, with a value of 4.10 we can reject the null hypothesis to claim that the two correlations are “significantly” different. In fact, setting the confidence level at 99%, we still beat the null hypothesis.

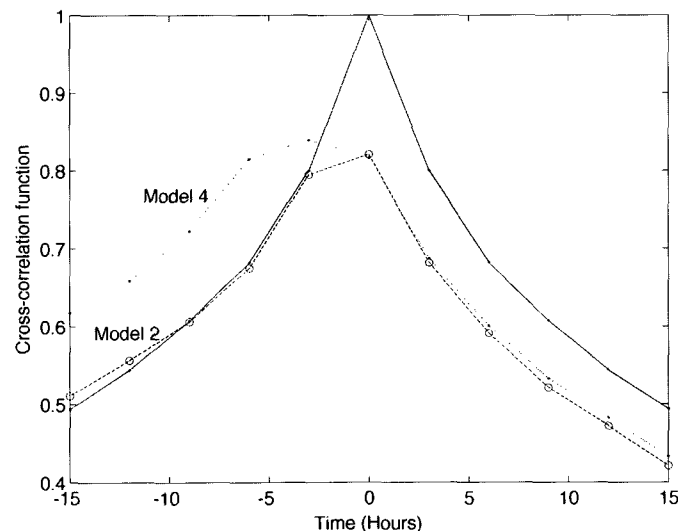


Figure 5.33 : Auto-correlation for 3-hour Kp and cross-correlation for the official Kp vs ANN predicted 3-hour Kp is shown here. Note that the model which includes Kp history as an input (Model 4) appears to lag the real data.

However, for my model which uses Kp history (model 4), 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 4) appears to overrely on the previous Kp value. Thus a prediction using Kp history significantly lags a true prediction. The physical implication of this behavior can be felt during times of non-availability of nowcast Kp. In which case, the predictions could be an overestimate and inaccurate because we will have to duplicate the previous Kp. The model 3 “prediction peaks” clearly lags in time relative to the measured Kp (shown between the two vertical dashed lines in figure 5.34), whereas the trend is not visible in model 1 (top panel in figure 5.34). This effect can also be seen in other papers which use Kp history (e.g. see Fig 5 (g) and (h) of Wing et al. [2005] showing their models 1 and 2, shown earlier in figure 1.11). But their model (APL model 3), which is purely driven by the solar wind alone, does not exhibit this feature. We can now further claim that in spite of a small (not significant) drop in linear correlation coefficient and the prediction efficiency, model 2 really does forecast (not just duplicate) Kp, almost 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. The importance of our models 1 and 2, by eliminating the Kp in the input, not only eliminates the persistence effects and the fact that the realtime Kp is only given to one digit precision, but they also outperform some of the existing models.

5.3.8 Kp models compared

In chapter 1, I presented a survey of 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 of them use the solar wind to predict Kp approximately one hour ahead and therefore, easily compare with our own model 1. Even though I did not attempt to re-evaluate them model by model, I did conduct a head-to-head with the Costello Kp model over the data I found that was common

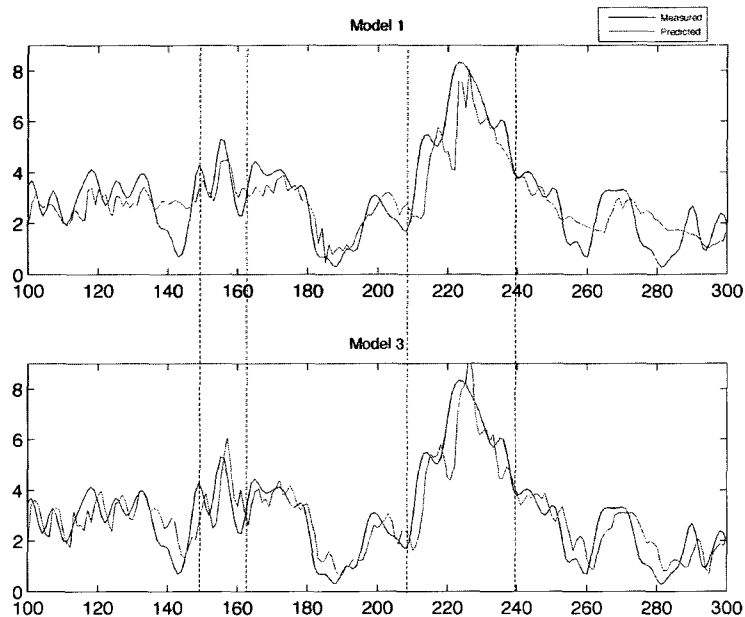


Figure 5.34 : Plot to demonstrate the effect of persistence visible in the models using Kp inputs. Prediction “peaks” are clearly shifted relative to the measured values in the bottom panel (model 3 which uses Kp as an input).

to us. As far as the other models, I have used 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. My model 1 outperforms the Costello NN and the Boberg et al. NN models. But the slight edge my model 1 has over the APL model 1 could be attributed to the fact that the latter was evaluated over a sufficiently longer time period. On the other hand, my model 1 has a lower correlation coefficient ($r = 0.826$) than the APL model 1 ($r = 0.92$). Furthermore, the APL models 1 and 3 and Costello NN Kp model predict Kp approximately 1 hour ahead every 15 minutes, which partly explains the differences in the prediction performance. As mentioned earlier, correlation coefficients alone are not good enough for model evaluations when

it comes to predictions, but since it is difficult for the actual models to be able to get ported for testing, they become the sole metric for evaluations.

In a head-to-head test of my model 1 with the Costello's ANN Kp, performed towards a true comparison, we find that our model 1 predicted Kp versus the real Kp is found to be $r = 0.84$ while the Costello ANN predicted Kp versus the real Kp is found to be $r = 0.81$ for the same time frame (figure 5.35). In order to conduct a fair test, I have chosen the data 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, I reran my model to ensure that the training set does not include the test data. 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].

5.3.9 “Spacalrt” real-time warning system

The models that have been just developed are ready for implementation. In fact, models 1 and 2 have already been routinely running in real time and the results of the predictions are all available from <http://space.rice.edu/ISTP/wind.html>. As mentioned earlier, the BI plot (figure 1.11) was launched into real-time mode in Oct. 2003, purely for a scientific and educational motive. We have nearly 600 subscribers coming from a variety of backgrounds as of date.

One of NOAA's Space Weather scales for geomagnetic is based on Kp as a physical measure. Based on this measure, a minor geomagnetic “G1” storm alert is issued if Kp reaches 5. Alerts and warnings are issued based on two factors: (1) predicted

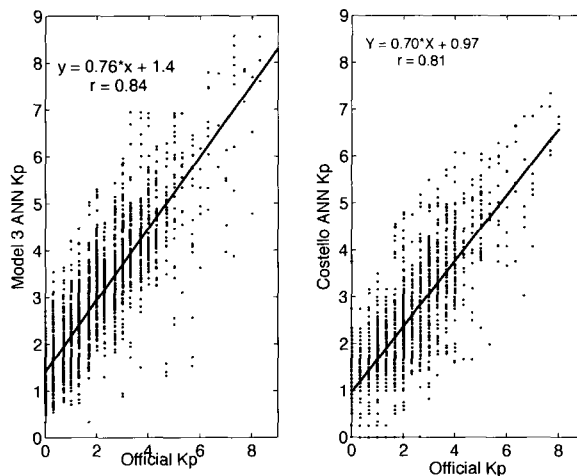


Figure 5.35 : Head-to-head performance of Model 1 and Costello ANN Kp: (a) Model 1 predicted Kp versus measured Kp ($r = 0.84$), and (b) Costello ANN predicted Kp versus measured Kp ($r = 0.81$) [Bala et al., 2009].

parameter exceeding a certain threshold, and (2) based on persistence i.e., if the predicted value continues to be high above a certain threshold for extended periods. At present, “Red” alerts are being issued to our subscribers whenever the 10-minute Boyle index exceeds 200 kV, provided the preceding 10-minute Boyle index had been at least 150 kV. This will provide a rapid response. However, the magnitude of the storm to come depends, of course, on the time history of the solar wind. The new ANN-based models have been trained to use the solar wind history. Our new algorithms now successfully predicts both a 1-hour and a 3-hour Kp, updated each half-hour. Accordingly, we will also issue a “Red” alert if either the 1-hour or 3-hour Kp is predicted to be 6 or greater, and will be issued continuously until the predicted value of Kp subsides.

We have shown that ANNs are excellent tools for temporal prediction problems and established that Kp can be forecasted with good accuracy. The following subsections introduce new formulae that predicts the Dst and the AE with lead times of

1 and 3 hours.

5.3.10 Model 5: 1-hour lead time Dst predictions from BI

We have already established that the solar wind coupling function in the Boyle Index correlates well with the Dst, and that their correlations are significant up to a 6-hour time range, within which the series is predictable. This model was easy to construct, because, unlike the Kp index, inherently, the Dst is an hourly measure. Since nowcast Dst data are somewhat unreliable, we prefer simpler functions to avoid any concerns. We set out to find the value of n in equation 3.8. The best predicting function was derived based on the various performance indicators and statistics listed in table 5.7. Our best function can be written as

$$\mathbf{Dst}_{t+1}^* \equiv \mathbf{f}(\mathbf{BI}_t, \mathbf{BI}_{t-1}, \dots, \mathbf{BI}_{t-9}) \quad (5.5)$$

Following a thorough and systematic study, we find that it takes 10 hours of Dst time history for the ANN to get the best prediction overall ($r = 0.802$ and $\text{RMSE} = 14.46$ nT \forall values of Dst; $\text{HSS} = 0.549$ and $\text{ARE} = 0.376$ when $\text{Dst} < -40$ nT). I set the Dst threshold to -40 nT based on a study conducted by Cane and Richardson, [2003] that reported the CMEs between 1996 and 2002 (see figure 5.36) and the corresponding minimum Dst values recorded. This data span (1996-2002) covers roughly about 50% of the data volume used in this study.

Figure 5.37 shows the overall results from test samples with the model outputs plotted against the measured Dst. Evidently, the model overestimates (large negative values indicate severity) Dst below 0 nT. For the Dst threshold set at -40 nT, the hits versus miss ratio is 186 to 123 with a 0.505 POD. If the threshold is raised to -80 nT, we get a 248:178 hits to miss ratio with a 0.660 HSS. The overall RMSE of 14.56 nT (22398 hours of test data) found in this model is better than the Wu and Lundstedt,

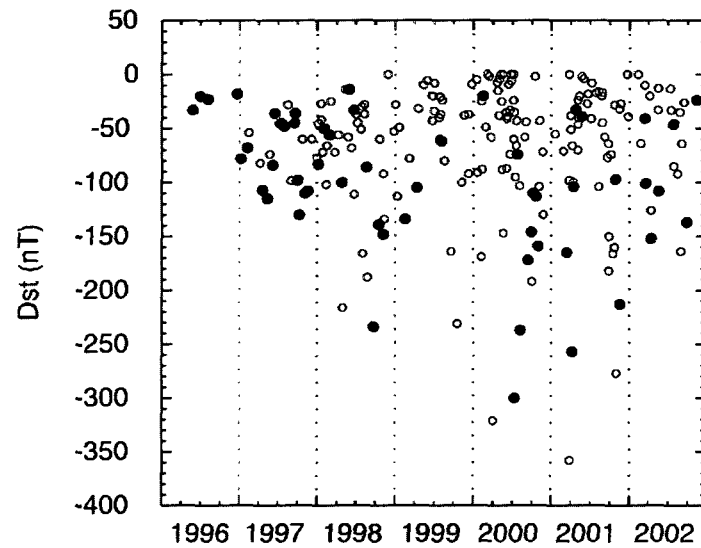


Figure 5.36 : Distribution of minimum Dst value due to the passage of various CMEs or related sheath regions versus the time (Adopted from Cane and Richardson, [2003]).

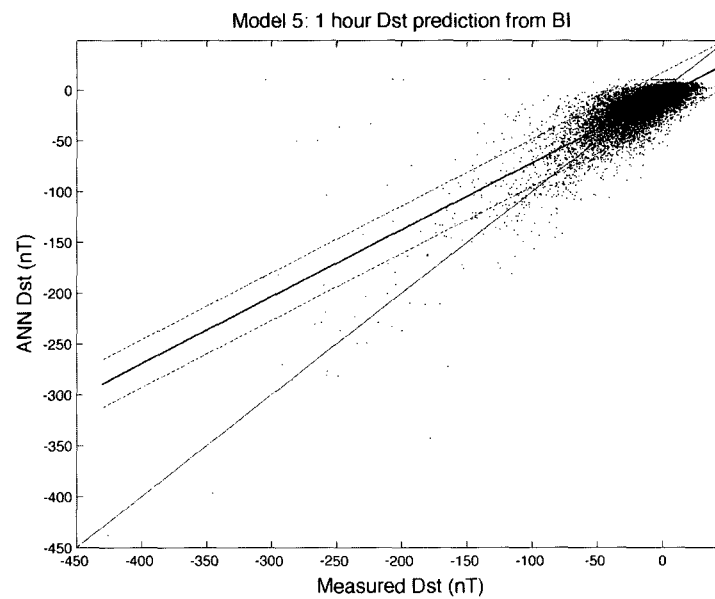


Figure 5.37 : Model 5: ANN predicted Dst vs measured Dst, $r = 0.802$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

[1997] model value of 14.7 nT (1957 hours of test data), but slightly behind Temerin and Li, [2002] value of 7.9 nT (4320 hours of test data). This model's performance during some severe events will be discussed in the next section. Furthermore, since this model does not include the time history of Dst in the inputs, it offers true prediction without the issue of persistence; the peak (0.802) of the cross-correlation function is observed at the time lag 0.

Table 5.7 : Predict Dst_{t+1}^* from solar wind

#	Model 5 Inputs	Results				
		Time History	Linear Corr.	RMSE Valid.	HSS (Dst < -40)	ARE (Dst < -40)
1	BI_t	1	0.653	18.23	0.381	0.485
2	BI_t, BI_{t-1}	2	0.711	16.84	0.418	0.455
3	BI_t, BI_{t-1}, BI_{t-2}	3	0.696	17.52	0.441	0.441
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}$	4	0.755	15.85	0.484	0.417
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}$	5	0.764	15.85	0.495	0.401
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}$	6	0.775	15.45	0.520	0.388
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}$	7	0.797	14.63	0.517	0.368
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}$	8	0.801	14.76	0.528	0.366
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}$	9	0.801	14.45	0.530	0.360
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}$	10	0.802	14.56	0.549	0.351

5.3.11 Model 6: 3-hour lead time Dst predictions from BI

We predict Dst 3 hours in advance in the same manner as model 2, which predicts Kp 3 hours ahead. Since Dst is a 1-hour average, we compare our results against the measured Dst compiled with 3 hour averages. We list the results of the various test functions in table 5.8. The best performing function is written as:

$$\mathbf{Dst}_{t+3}^* \equiv \mathbf{f}(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-18}) \quad (5.6)$$

for its higher skill score (HSS = 0.613 for Dst < -40 nT) and a low RMSE (13.42 nT) overall (5988 total points). Generally speaking, all our best representations have been achieved using 6 to 8 inputs of the solar wind only or with the combination of the target index; model 5 has been the only exception to this general trend, which has 10 inputs. The ANN limits the model inputs to achieve the best generalization of the data presented. The actual predicted Dst values versus the measured values are shown in figure 5.38 ($r = 0.853$). Interestingly, this result indicates that this model performs better than model 5 in addition to extending the forecast lead time. Furthermore, when the Dst threshold is set at -40 nT, the overall hits:miss ratio is 303:233 and 68:55 when the threshold is lowered to -80 nT. Overall, this model does well compared to model 5.

Let's take a closer look at the predictions from both the Dst models through figures 5.39 and 5.40, which illustrates two unique cases, December 2006 and April 2001 respectively. The Earth-directed CME, which occurred on December 13, 2006, impacted the Earth's magnetosphere at ~ 1500 UT 14 December, 2006 (velocity was measured at ~ 900 km/sec). The Dst index dipped to -45 nT at around 0000 UT December 15, 2006 before intensifying further to a value of -145 nT at ~ 0700 UT December 15, 2006. The recovery phase of the storm started around ~ 1200 UT 15 December 2006. The April 2001 data presents a different scenario with a series of

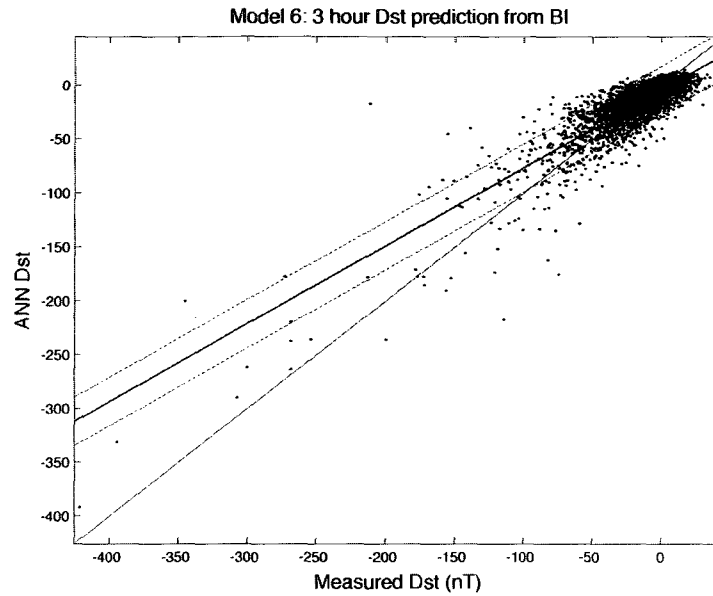


Figure 5.38 : Model 6: ANN predicted Dst vs measured Dst, $r = 0.853$. This model gives a lead time of 3 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

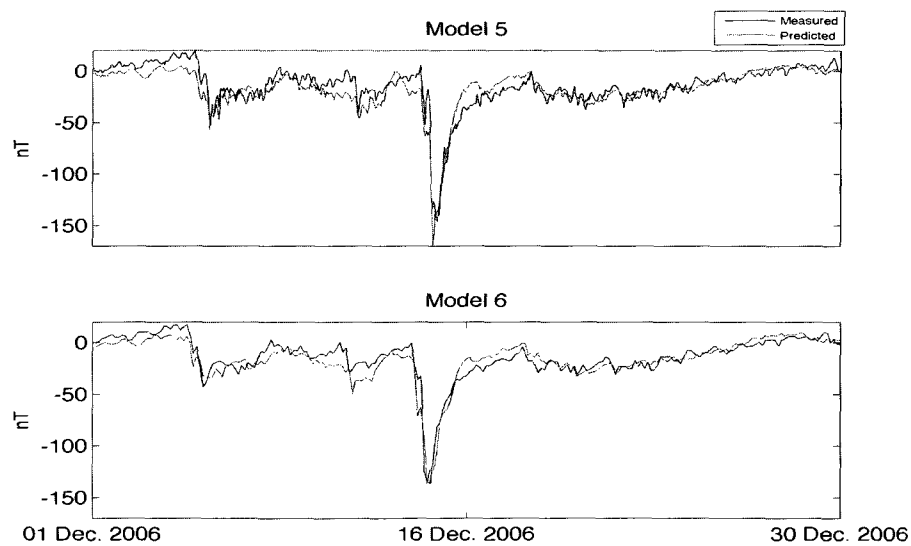


Figure 5.39 : Predictions from models 5 and 6 are overplotted with the measured Dst. Data shown here covers all of December 2006.

Table 5.8 : Predict Dst_{t+3}^* from solar wind

#	Model 6 Inputs	Results				
		Time History	Linear Corr	RMSE Valid.	HSS (Dst < -40)	ARE (Dst < -40)
1	BI_t	3	0.741	15.33	0.491	0.403
2	BI_t, BI_{t-3}	6	0.769	14.44	0.511	0.389
3	BI_t, BI_{t-3}, BI_{t-6}	9	0.805	14.20	0.523	0.354
4	$BI_t, BI_{t-3}, \dots, BI_{t-9}$	12	0.820	13.72	0.549	0.366
5	$BI_t, BI_{t-3}, \dots, BI_{t-12}$	15	0.846	13.34	0.560	0.337
6	$BI_t, BI_{t-3}, \dots, BI_{t-15}$	18	0.846	13.08	0.567	0.335
7	$BI_t, BI_{t-3}, \dots, BI_{t-18}$	21	0.853	13.42	0.613	0.335
8	$BI_t, BI_{t-3}, \dots, BI_{t-21}$	24	0.828	12.99	0.587	0.315
9	$BI_t, BI_{t-3}, \dots, BI_{t-24}$	27	0.844	12.28	0.605	0.293
10	$BI_t, BI_{t-3}, \dots, BI_{t-27}$	30	0.840	13.06	0.611	0.307

geomagnetic storms, the most intense one occurring on April 11, 2001 (Dst dipping to -271 nT around 2300 UT). It can be seen that both the models are fairly accurate in predicting the initial and recovery phases, implying that the magnetospheric response to the change in solar wind conditions happens well within the time limits of the predicted interval, and in agreement with the cross-correlation function. However, based on the overall prediction summary covering the two events (table 5.9), it can be said that our 3-hr Dst model does indeed offer better predictions, given its lead time. Finally, this model's cross-correlation function peaks at the zeroth lag. The next two subsections provides an overview of the AE prediction models.

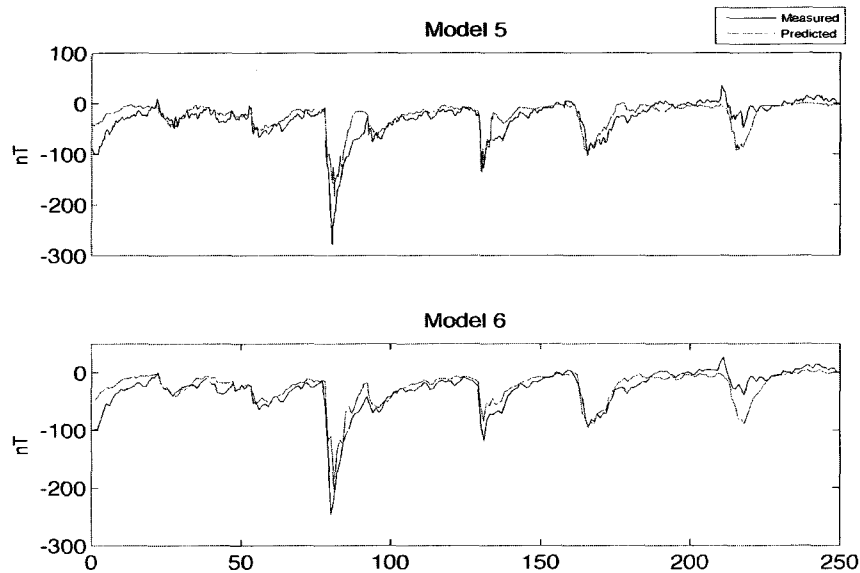


Figure 5.40 : Predictions from models 5 and 6 are overplotted with the measured Dst. Data shown here covers all of April 2001.

5.3.12 Model 7: 1-hour lead time AE predictions from BI

We trained our 1-hour AE prediction model in the manner such that linear scale was chosen to handle the BI inputs and the corresponding AE values at the output. The AE index, characterizing the auroral activity in the auroral zone, varies over a free range with minimum at 0 nT, while the BI, an approximation for the PCP potential drop, also takes non-negative values with no pre-defined maximum. Table 5.10 lists the results summary based on different time histories in the input. The best function describing our BI-AE relationship can be defined by the following form:

$$\mathbf{AE}_{t+1}^* \equiv \mathbf{f}(\mathbf{BI}_t, \mathbf{BI}_{t-1}, \dots, \mathbf{BI}_{t-6}) \quad (5.7)$$

The model predictions versus the measured values is shown in figure 5.41, with $r = 0.74$. Even though the value of the correlation coefficient might look significant, it is quite apparent that the RMSE and the average relative error are large numbers

Table 5.9 : Geomagnetic Storm Alerts (Dst threshold: < -40 nT).

Period Covered	April 2001		December 2006	
	5	6	5	6
Dst Model	5	6	5	6
Hits	115	47	26	9
Misses	71	17	8	1
False Alarms	31	10	13	4
Right Rejections	533	175	653	223
Total	750	249	700	237
Maximum Lead Time	1 hours	3 hours	1 hour	3 hours
Heidke Skill Score	0.607	0.706	0.707	0.771
Probability of Detection	0.618	0.734	0.765	0.900
False Alarm Rate	0.212	0.175	0.333	0.308

(as a comparison, the maximum ARE observed in our other models was $\sim 28\%$). The steep decline in the prediction performance of the AE model when compared to the Kp and Dst models can be attributed to the fact that AE has two components to it. As has been addressed earlier, the two components are the solar wind energy input into the magnetosphere and the tail magnetic field reconfiguration, resulting in geomagnetic substorms. Therefore, the network's ability to learn input-output patterns is impaired because of the effect of the latter, which, perhaps, is somewhat causally decoupled i.e., the instantaneous AE response to the solar wind is captured well whereas the component of AE due to tail-side reconfiguration is rather slow in its response to be modeled by the network.

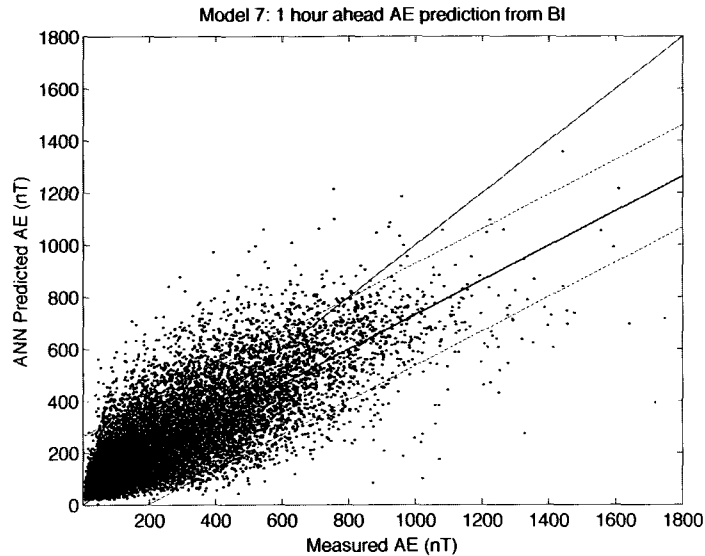


Figure 5.41 : Model 7: ANN predicted AE vs measured AE, $r = 0.810$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

5.3.13 Model 8: 3-hour lead time AE predictions from BI

This model is very similar in architecture to its Kp and Dst counterparts. Since AE is a 1-hour index, we construct 3-hour averages of the AE to perform one-step ahead predictions. The best function to predict AE with a lead time of 3 hours is given by:

$$\mathbf{AE}_{t+3}^* \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-15}) \quad (5.8)$$

Figure 5.42 shows the scatterplot of ANN predicted AE versus the measured AE ($t = 0.740$, $\text{HSS} = 0.466$, $\text{ARE} = 0.332$). The 3-hour ahead predictions are poor compared to the baseline set by the 1-hour model, and the quality of the predictions are far from being accurate. We again choose to use the time intervals of April 2001 (figure 5.43) and December 2006 (figure 5.44) to represent our results for closer investigation. In general, in both these examples, it can be seen that the network certainly does have the ability to predict the rises and fall in the AE but clearly

Table 5.10 : Predict AE_{t+1}^* from solar wind

#	Model 7 Inputs	Results				
		Time History	Linear Corr.	RMSE Valid.	HSS (AE > 500)	ARE (AE > 500)
1	BI_t	1	0.803	125.32	0.571	0.289
2	BI_t, BI_{t-1}	2	0.805	125.31	0.538	0.291
3	BI_t, BI_{t-1}, BI_{t-2}	3	0.805	123.47	0.535	0.286
4	$BI_t, BI_{t-1}, \dots, BI_{t-3}$	4	0.810	124.18	0.554	0.289
5	$BI_t, BI_{t-1}, \dots, BI_{t-4}$	5	0.801	125.18	0.560	0.289
6	$BI_t, BI_{t-1}, \dots, BI_{t-5}$	6	0.797	126.77	0.544	0.297
7	$BI_t, BI_{t-1}, \dots, BI_{t-6}$	7	0.810	122.04	0.562	0.281
8	$BI_t, BI_{t-1}, \dots, BI_{t-7}$	8	0.805	125.37	0.551	0.294
9	$BI_t, BI_{t-1}, \dots, BI_{t-8}$	9	0.799	124.83	0.538	0.302
10	$BI_t, BI_{t-1}, \dots, BI_{t-9}$	10	0.810	121.92	0.546	0.292

lacks the accuracy with the predicted values falling short of the measured values by a huge margin (RMSE values of 157.2 nT, 175.8, 162.3 nT and 167.6 nT). Barring a few cases, the network's inability to predict peaks larger than ~ 750 nT can also be noticed. Nevertheless, the results are quantifiable and significant (see table 5.12 for the prediction summary). Our 3-hour AE model offers a new benchmark over other existing algorithms which predicts with lead times of 1 hour or less (e.g., Gleisner and Lundstedt, 1997; Palocchia et al., 2008).

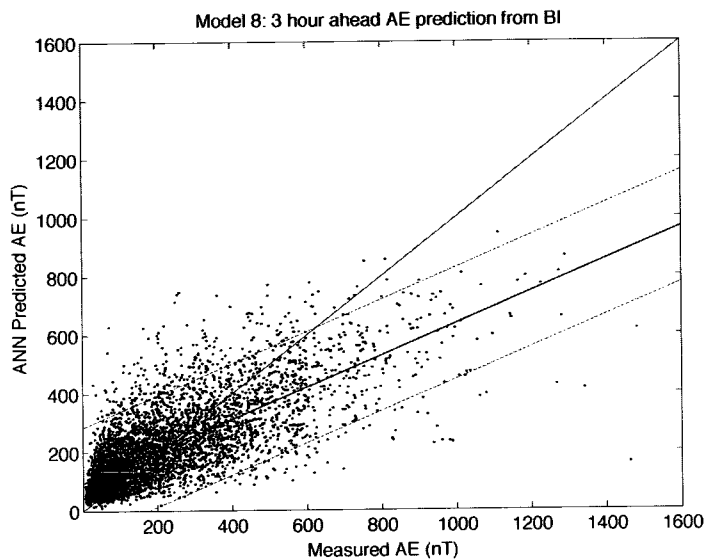


Figure 5.42 : Model 8: ANN predicted AE vs measured AE, $r = 0.74$. This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

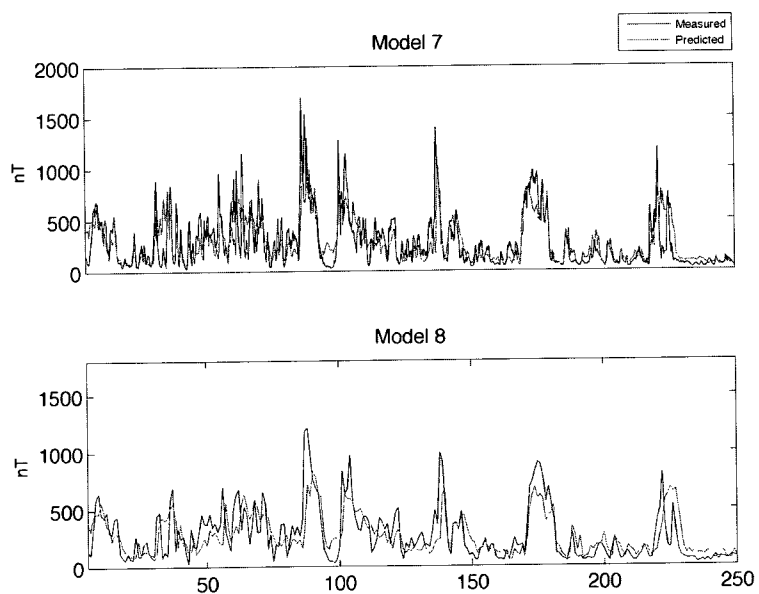


Figure 5.43 : Predictions from models 7 and 8 are overplotted with the measured AE. Data shown here covers all of April 2001.

Table 5.11 : Predict AE_{t+3}^* from solar wind

#	Model 8 Inputs	Results				
		Time History	Linear Corr	RMSE Valid.	HSS (AE > 500)	ARE (AE > 500)
1	BI_t	3	0.718	135.19	0.392	0.380
2	BI_t, BI_{t-3}	6	0.711	137.96	0.350	0.374
3	BI_t, BI_{t-3}, BI_{t-6}	9	0.728	133.86	0.367	0.363
4	$BI_t, BI_{t-3}, \dots, BI_{t-9}$	12	0.734	132.22	0.428	0.343
5	$BI_t, BI_{t-3}, \dots, BI_{t-12}$	15	0.723	133.42	0.350	0.362
6	$BI_t, BI_{t-3}, \dots, BI_{t-15}$	18	0.740	131.98	0.446	0.332
7	$BI_t, BI_{t-3}, \dots, BI_{t-18}$	21	0.708	136.45	0.404	0.372
8	$BI_t, BI_{t-3}, \dots, BI_{t-21}$	24	0.717	133.44	0.421	0.349
9	$BI_t, BI_{t-3}, \dots, BI_{t-24}$	27	0.728	135.36	0.446	0.360
10	$BI_t, BI_{t-3}, \dots, BI_{t-27}$	30	0.741	129.59	0.390	0.330

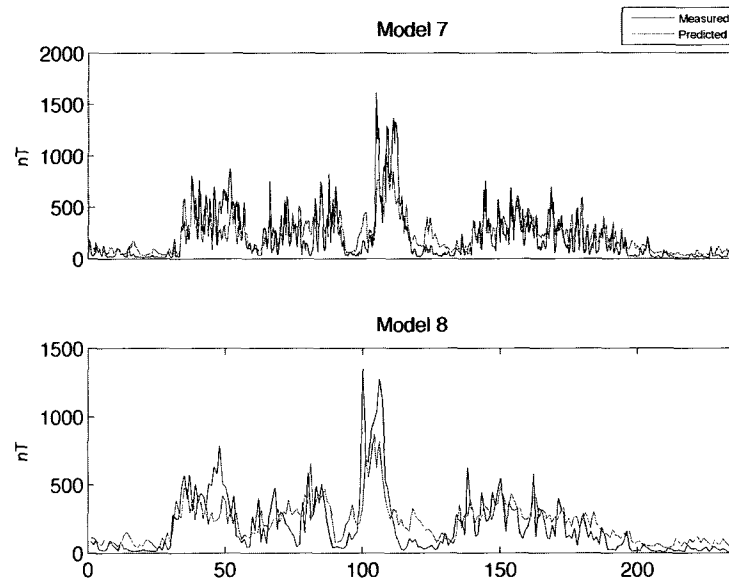


Figure 5.44 : Predictions from models 7 and 8 are overplotted with the measured AE. Data shown here covers all of December 2006.

Table 5.12 : Geomagnetic Storm Alerts (AE threshold: > 500 nT).

Period Covered	April 2001		December 2006	
AE Model	7	8	7	8
Hits	79	20	38	7
Misses	51	19	43	15
False Alarms	46	14	12	1
Right Rejections	584	197	607	214
Total	760	250	700	237
Maximum Lead Time	1 hour	3 hours	1 hour	3 hours
Heidke Skill Score	0.543	0.471	0.539	0.439
Probability of Detection	0.608	0.513	0.469	0.318
False Alarm Rate	0.368	0.412	0.240	0.125

5.3.14 Green's function test

We saw earlier that all inputs to the ANN architecture followed a simple rule that the most recent value of the BI and Kp receives the maximum weight of unity, the second most recent value receiving 90% of the previous, and so on (equation 3.1), thus forcing the network to count on the most recent values of the solar wind; however, this scheme was an arbitrary pick rather than a standard format.

A simple test, using Green's function to study the network response, demonstrates the effectiveness of this procedure. Let's send a short, yet strong, impulse function to the network such as the one defined below:

$$G(t) = \delta(t - t_0) = \begin{cases} 200 \text{ kV}, & t \rightarrow t_0, \\ 20 \text{ kV}, & \text{otherwise.} \end{cases}$$

The impulse function is simply a solar wind driver, which is a function of the BI. We have chosen a value of 20 kV to represent the average minimum solar wind conditions, since the models run either in 1- or 3-hour cadences; BI under 10 kV have been recorded but they are 10-minute averages. Figure 5.45 show the responses of the 6 BI-only models (2 of each Kp, Dst and AE). Obviously, the network does not respond until the onset of the 200 kV spike, as indicated by extremely low values of Kp, Dst and the AE. But, as soon as the impulse strikes the magnetosphere, the network responds instantaneously. This can be seen as the sudden jump in predictions on all 3-hour models and the 1-hour AE model. However, the prediction peaks of the 1-hour Kp and Dst models is not seen at the next subsequent time step, rather at the second time step following the onset of the impulse. This, of course, is because a 3-hour average is much stronger than a 1-hour average. Interestingly, the predictions do not decay before producing a lower secondary peak. Thus, the network predictions

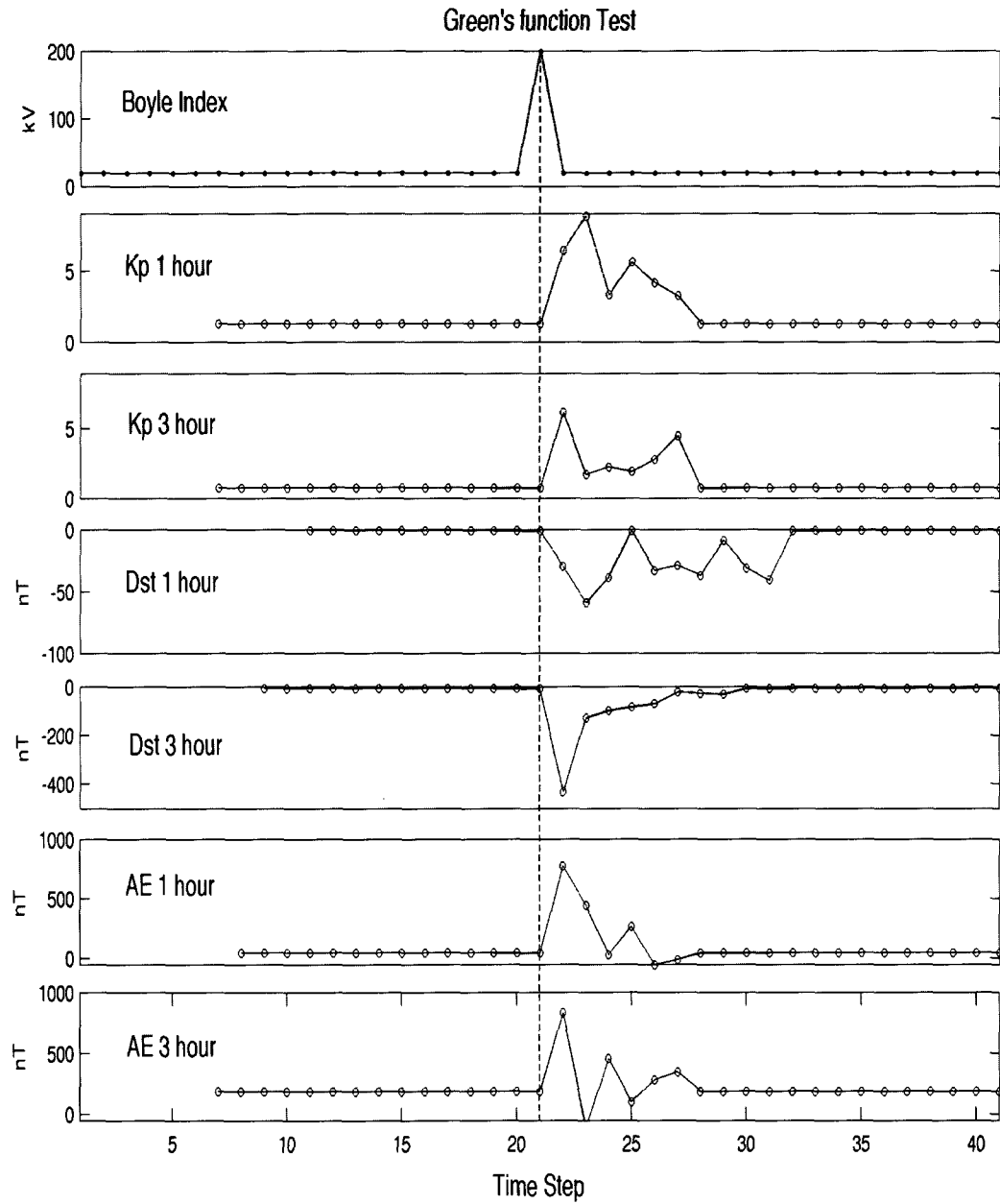


Figure 5.45 : A Green's function test reveals the network responding instantly, and decaying slowly thereafter, to the short solar wind impulse.

are strongly influenced by the most recent solar wind values. Furthermore, proving that the network can predict faster rate of change.

Table 5.13 gives the combined prediction summary of all the 8 models. We can claim that the geomagnetic indices (Kp, Dst and AE) are well predictable through solar wind coupling functions as inputs, given the scope of neural networks for temporal predictions. We were able to analyze the time variability of the BI using different time resolutions and time histories. Our study reveals that Kp and Dst are more predictable than the AE. However, despite having rigorously explored various possibilities and exploited the ANN to its capacity, a few questions still remains unanswered. The best HSS we were able to achieve was close to 80% (model 3). A clear one-to-one correspondence between the BI and these indices has not been drawn yet, needing a more closer look at the physics of the solar wind-magnetosphere coupling. The next section deals with building a new network by exploring the possibility of including the dynamic pressure term in the inputs to complement the BI because the BI does not contain an explicit solar wind density term.

Table 5.13 : Table showing the prediction summary. Thresholds: Kp > 4; Dst < -40 nT; AE > 500 nT.

Forecast Model	Linear Corr.	RMSE Test	HSS	ARE
Linear 1-hour Kp predictor using BI (18239 points)	0.712	-	-	0.844
Model 1: ANN 1-hour Kp predictor using BI (18225 points)	0.865	0.703	0.609	0.173
Linear 3-hour Kp predictor using BI (6079 points)	0.770	-	-	0.878
Model 2: ANN 3-hour Kp predictor using BI (6075 points)	0.819	0.846	0.533	0.201
Model 3: ANN 1-hour Kp predictor using BI & Kp (18225 points)	0.960	0.393	0.793	0.089
Model 4: ANN 3-hour Kp predictor using BI & Kp (6075 points)	0.839	0.767	0.565	0.195
Model 5: ANN 1-hour Dst predictor using BI (6075 points)	0.802	14.56 nT	0.549	0.351
Model 6: ANN 3-hour Dst predictor using BI (6075 points)	0.853	13.42 nT	0.613	0.335
Model 7: ANN 3-hour AE predictor using BI (6075 points)	0.810	122.04 nT	0.562	0.281
Model 8: ANN 3-hour AE predictor using BI (6075 points)	0.740	131.98 nT	0.446	0.332

5.4 Pressure term inclusion

We have investigated and shown that the BI can best characterize the magnetospheric activity over three distinct ground-based geomagnetic indices, namely Kp, Dst, and AE. The network architecture and free parameters of the ANN have been carefully chosen to improve the predictability of these indices. We have further demonstrated that these indices can be predicted well without the knowledge of their time histories.

The basic principle of this work has been to use the v_{sw}^2 as the viscous term and the $11.7B \sin^3(\theta/2)$ as the magnetic merging term and combine them into a single function, the BI, before using as inputs to the ANN. However, this function does not contain an explicit solar wind density term. Although it is widely believed that, after merging and viscous terms, the next term in the order of importance is the pressure term [Newell et al., 2008]. In order to increase the baseline further, we have extended the study to accommodate an exclusive dynamic pressure term (equation 5.9) in the network inputs.

$$\text{Dynamic pressure (Dp), } P_{sw} = m_p n_p v_{sw}^2 (1 + 4n_a/n_p) \quad (5.9)$$

where n_p is the number density of the protons and n_a/n_p is the alpha to proton ratio; P_{sw} is one of the most commonly used viscous terms besides v_{sw} . Here, we propose to bring an extra term into the network in the form of solar wind pressure rather than actually modifying the BI i.e., for a set of BI in the inputs there is a corresponding set of dynamic pressure term.

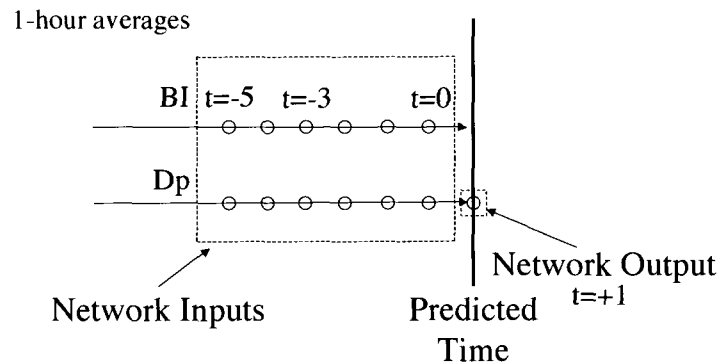
5.4.1 Network Training

We pair the input vectors and its corresponding target in the same manner as models 3 and 4, except that we replace the Kp with dynamic pressure. Recall, model 3 predicts Kp 1 hour ahead using hourly averages and model 4 predicts Kp 3 hours

ahead using three hour averages. Data for training were sampled using the “sliding window” method described for models 1 through 8. For example, 6 hours of solar wind input history means having a set of 6 BIs and 6 Dps as inputs to the network, and for which, the desired target is Kp_{t+1} (figure 5.46). We name this a training pattern. Training patterns are shuffled randomly and then chosen and designated for training, testing and validation ($\approx 47\%$, $\approx 36\%$, and $\approx 17\%$ and $\approx 59\%$, $\approx 22\%$, and $\approx 19\%$ for 1-hour and 3-hour models respectively).

The motivation for this technique owes to the success we had developing models 3 and 4, both of which had two separate input streams. The logistics and the right network design, of two input time series, have been explored already. It is enough that we experiment with 5 to 10 sets of inputs and evaluate the network performance in each case.

1-hour lead time Kp predictions using BI and Dp



$$Kp_{t+1}^{\text{Ram}} = f(BI_{t=0}, BI_{t=-1}, \dots, BI_{t=-5}; Dp_t, Dp_{t=-1}, \dots, Dp_{t=-5})$$

Figure 5.46 : An example: an input-output training pattern for the 1-hour Kp model that includes the pressure term. The same idea may be extended for different inputs and time cadences and, for the Dst and AE models.

5.4.2 The “Ram” functions

We evaluated how well different input histories helped the predictions. The following equations summarize the best performing functions.

$$\mathbf{Kp}_{t+1}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-1}, \dots, \mathbf{BI}_{t-5}; \mathbf{Dp}_t, \mathbf{Dp}_{t-1}, \dots, \mathbf{Dp}_{t-5}) \quad (5.10)$$

$$\mathbf{Kp}_{t+3}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-24}; \mathbf{Dp}_t, \mathbf{Dp}_{t-3}, \dots, \mathbf{Dp}_{t-24}) \quad (5.11)$$

$$\mathbf{Dst}_{t+1}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-1}, \dots, \mathbf{BI}_{t-9}; \sqrt[3]{\mathbf{Dp}}_t, \sqrt[3]{\mathbf{Dp}}_{t-1}, \dots, \sqrt[3]{\mathbf{Dp}}_{t-9}) \quad (5.12)$$

$$\mathbf{Dst}_{t+3}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-21}; \sqrt[3]{\mathbf{Dp}}_t, \sqrt[3]{\mathbf{Dp}}_{t-3}, \dots, \sqrt[3]{\mathbf{Dp}}_{t-21}) \quad (5.13)$$

$$\mathbf{AE}_{t+1}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-1}, \dots, \mathbf{BI}_{t-6}; \sqrt[3]{\mathbf{Dp}}_t, \sqrt[3]{\mathbf{Dp}}_{t-1}, \dots, \sqrt[3]{\mathbf{Dp}}_{t-6}) \quad (5.14)$$

$$\mathbf{AE}_{t+3}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-18}; \sqrt[3]{\mathbf{Dp}}_t, \sqrt[3]{\mathbf{Dp}}_{t-3}, \dots, \sqrt[3]{\mathbf{Dp}}_{t-18}) \quad (5.15)$$

Equations 5.10 to 5.15 are quite reminiscent of the Kp models (models 3 and 4) that included the Kp time history. The best performing “dynamic pressure” term for the new Kp models is the Dp itself. However, for the new Dst and AE models, the best predictor is not the Dp term but Dp with power reduced to a third $\sqrt[3]{Dp}$. We considered different powers of Dp before using $\sqrt[3]{Dp}$. While the motivation for this approach was conceived from literature examples, my arrival to this term is merely through experimentation.

The prediction efficiency of the new models is better or at least as good as the standalone BI models (table 5.14). The biggest advantage of including the dynamic pressure term in inputs can be seen in the Kp models (figures 5.47 and 5.48). The overall statistics have improved “significantly” in both the models. On the other hand,

the results from the new AE models (figures 5.49 and 5.50) are identical to models 7 and 8. We did not see a clear winner here as we saw in Kp. On the other hand, the Dst models (figures 5.51 and 5.52) are special in that we did see a slight improvement in the 1-hour model but the results have worsened for the 3-hour model. It is worth mentioning that the Dst index have been corrected for pressure term contamination using the BMR equation (chapter 1). To summarize, Kp is best predicted using the BI and dynamic pressure and Dst is best predicted an hour ahead using hourly cadences of the BI and dynamic pressure; the predictions are also clear of persistence contamination. As far the AE indices, the results are inconclusive about using the dynamic pressure term, probably making it easier to use the BI as standalone for real time predictions.

Finally, having investigated the effectiveness of the BI in raising the baseline further in short-term (3 hours or less) geomagnetic activity index forecasting, we now turn our attention, as we originally hypothesized, towards further investigating whether legitimate forecasting is plausible beyond lead times of over 3 hours, perhaps up to 6 hours.

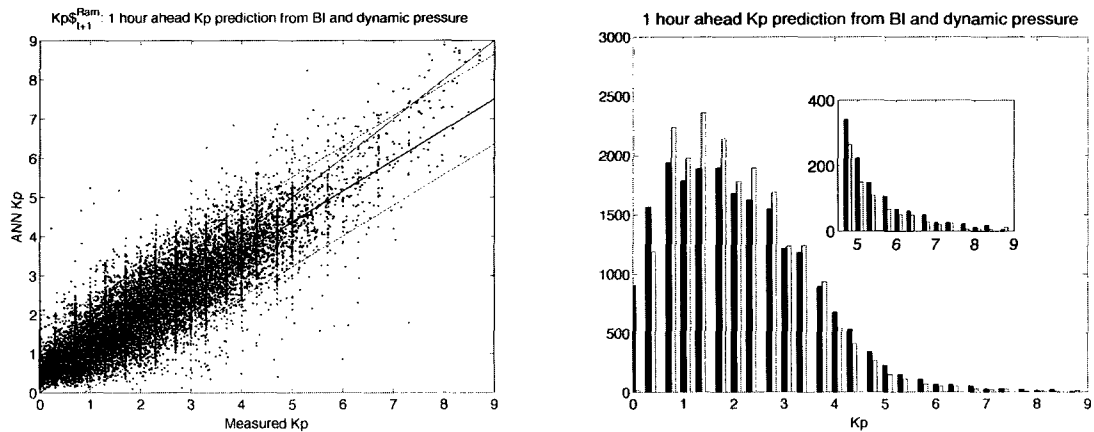


Figure 5.47 : BI and Dp included: ANN predicted K_p vs measured K_p , $r = 0.884$ (left panel). This model gives a lead time of 1 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit. K_p histogram is shown to the right.

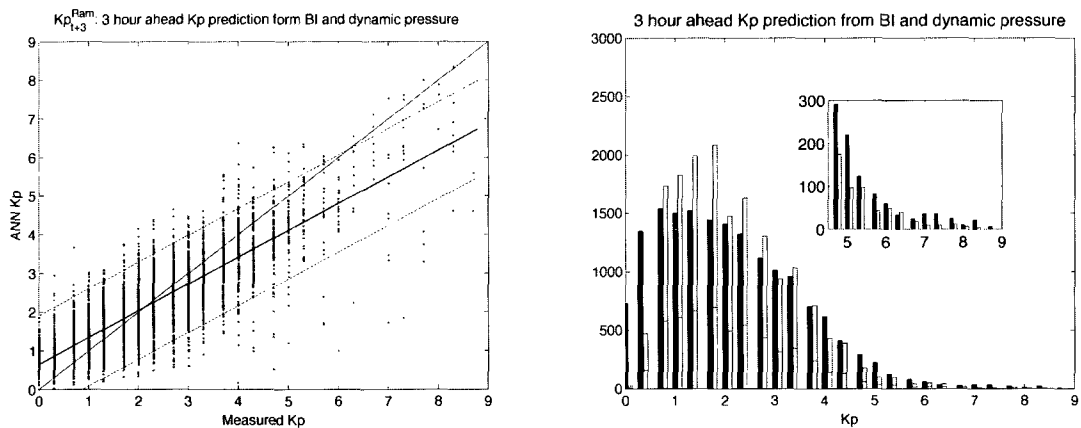


Figure 5.48 : BI and Dp included: ANN predicted K_p vs measured K_p , $r = 0.841$ (left panel). This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit. K_p histogram is shown to the right.

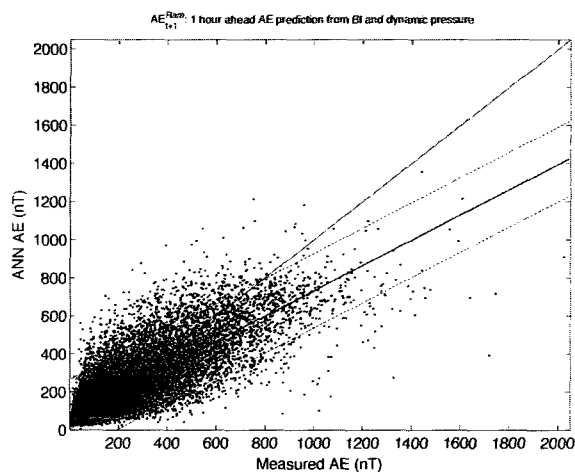


Figure 5.49 : BI and Dp included: ANN predicted AE vs measured AE, $r = 0.810$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

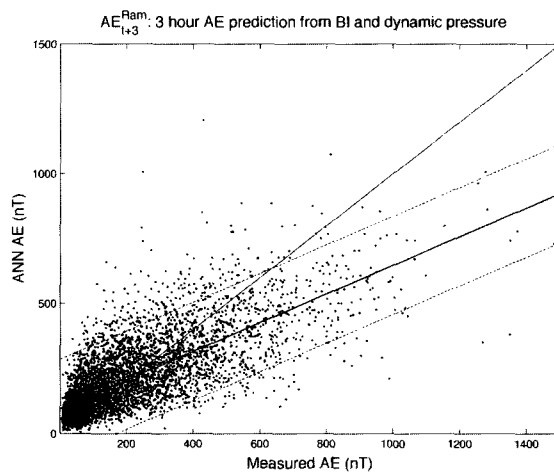


Figure 5.50 : BI and Dp included: ANN predicted AE vs measured AE, $r = 0.744$. This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

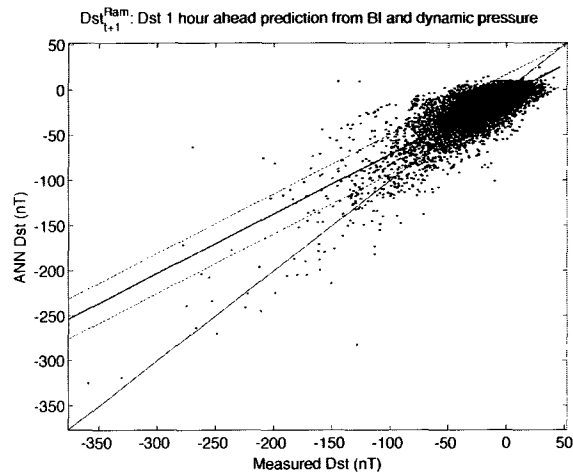


Figure 5.51 : BI and Dp included: ANN predicted Dst vs measured Dst, $r = 0.814$. This model gives a lead time of 1 hour. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

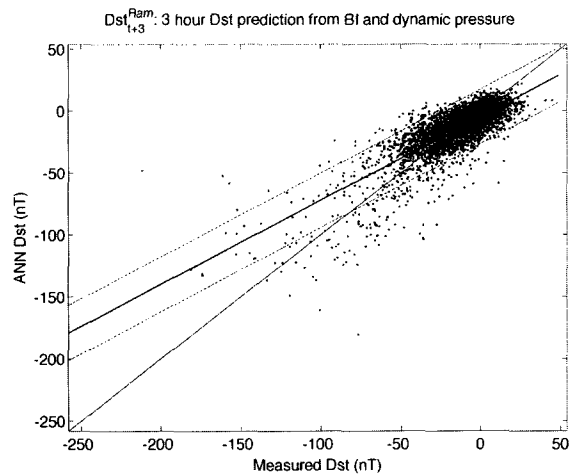


Figure 5.52 : BI and Dp included: ANN predicted Dst vs measured Dst, $r = 0.811$. This model gives a lead time of 3 hours. The dashed lines are the 95% confidence bounds. Thick black line is the linear fit.

Table 5.14 : Table showing the prediction summary of the models including dynamic pressure term. The numbers inside the parenthesis are the original BI only models. Thresholds: Kp > 4; Dst < -40 nT; AE > 500 nT.

Forecast Model: Pressure term included	Linear Corr.	RMSE Test	HSS	ARE
Kp_{t+1}^{Ram} : ANN 1-hr Kp predictor using BI & Dp	0.884 (0.865)	0.659 (0.703)	0.626 (0.609)	0.160 (0.173)
Kp_{t+3}^{Ram} : ANN 3-hr Kp predictor using BI & Dp	0.841 (0.819)	0.779 (0.846)	0.511 (0.533)	0.198 (0.201)
Dst_{t+1}^{Ram} : ANN 1-hr Dst predictor using BI & Dp	0.814 (0.802)	13.91 nT (14.56)	0.570 (0.549)	0.345 nT (0.351)
Dst_{t+3}^{Ram} : ANN 3-hr Dst predictor using BI & Dp	0.811 (0.853)	13.47 nT (13.42)	0.566 (0.613)	0.346 nT (0.335)
AE_{t+1}^{Ram} : ANN 1-hr AE predictor using BI & Dp	0.810 (0.810)	123.48 nT (122.04)	0.556 (0.562)	0.284 nT (0.281)
AE_{t+3}^{Ram} : ANN 3-hr AE predictor using BI & Dp	0.744 (0.740)	129.72 nT (131.98)	0.422 (0.446)	0.347 nT (0.332)

5.5 Longer range predictions

The two important issues relevant to space weather forecasting are accuracy and lead time. Thus far, we focussed on the former, improving the accuracy of short-term forecasts in the range 1-3 hours. Our results suggests that we have fair deal of accuracy in this range and that they are better than some of the better known models available. This section deals with the latter, increasing the lead time of forecasts without giving up the accuracy.

The network design and training are similar to the 3-hour models. The only novelty is that we predict “2” steps ahead instead of just “1”, using three hour averages of all the parameters (BI, Dp, Kp, Dst and AE). Figure 5.53 is a generic architecture of the 6-hour lead time models. We chose to include the Dp term in the inputs, given its performance earlier. The predictions from the 3 “6-hour” models will be called Kp_{t+6}^{Ram} , Dst_{t+6}^{Ram} , and AE_{t+6}^{Ram} . The best functions are:

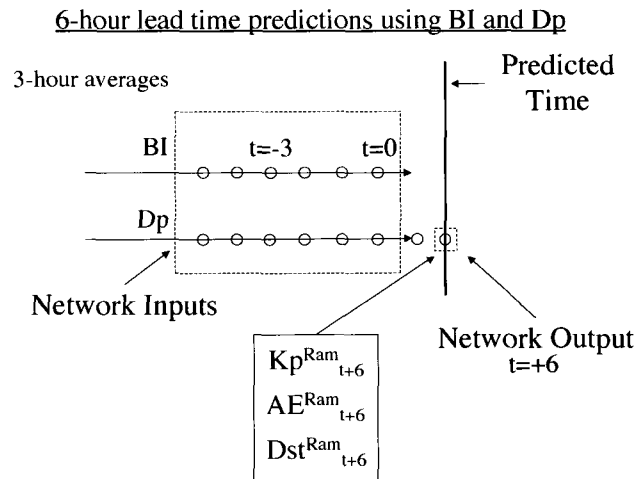


Figure 5.53 : An example: an input-output training pattern for the 6-hour ahead Kp, AE and Dst models that includes the pressure term.

Table 5.15 : Summary of models with 6 hour lead time from BI and Dp.

Forecast Models	Linear Correlation	RMSE Test	HSS [†]	ARE [†]
ANN 6-hr Kp predictor	0.730	0.890	0.285	0.292
ANN 6-hr Dst predictor	0.692	10.97	0.347	0.408
ANN 6-hr AE predictor	0.553	125.87	0.070	0.548

[†]Thresholds: Kp > 4; Dst < -40 nT; AE > 500 nT.

$$\mathbf{Kp}_{t+6}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-15}; \mathbf{DP}_t, \mathbf{DP}_{t-3}, \dots, \mathbf{DP}_{t-15}) \quad (5.16)$$

$$\mathbf{Dst}_{t+6}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-15}; \mathbf{DP}_t, \mathbf{DP}_{t-3}, \dots, \mathbf{DP}_{t-15}) \quad (5.17)$$

$$\mathbf{AE}_{t+6}^{\text{Ram}} \equiv f(\mathbf{BI}_t, \mathbf{BI}_{t-3}, \dots, \mathbf{BI}_{t-15}; \mathbf{DP}_t, \mathbf{DP}_{t-3}, \dots, \mathbf{DP}_{t-15}) \quad (5.18)$$

Table 5.15 list the summary of the models. We use the same reference thresholds here: Kp > 4, Dst < -40 nT, and AE > 500 nT. It can be seen that the predictions have gone down significantly (the best HSS is less than 35%). A deeper analysis of the results reveals the statistics shown in table 5.16 (thresholds: Kp > 6, Dst < -40 nT, and AE > 500 nT). Furthermore, I have investigated the predictions from these 6-hour models against the two storms in April 2001 and December 2006, our reference baseline. Interestingly, the Kp and the Dst have been modeled well (figure 5.54). However, the AE predictions are extremely poor. The HSS is over 74% and 55% while predicting Kp and Dst respectively, for the period covered in December 2006.

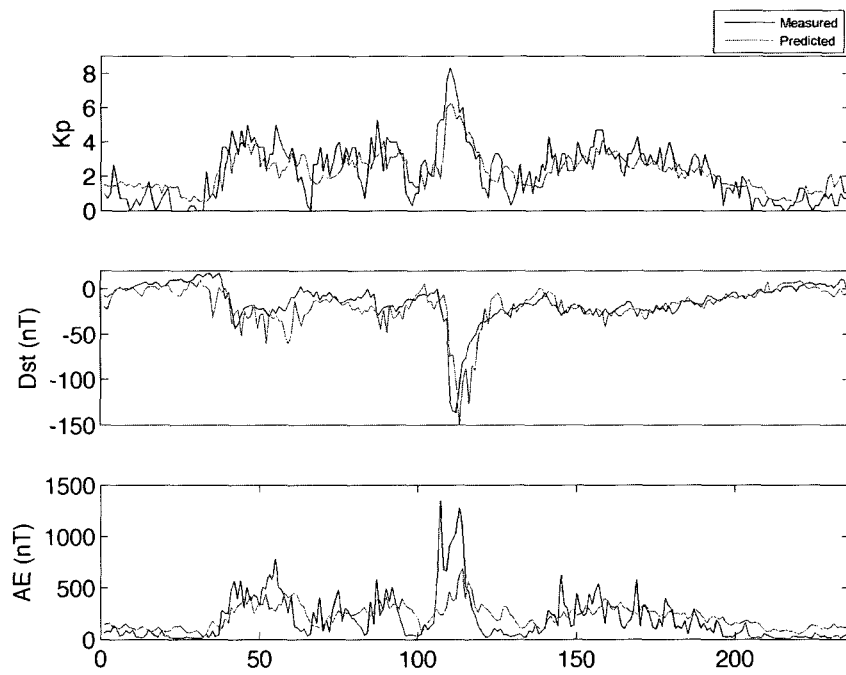


Figure 5.54 : Predictions from 6-hour models are overplotted with the measured values. Data shown here covers all of December 2006.

Table 5.16 : Geomagnetic Storm Alerts (Thresholds: $K_p > 6$; $Dst < -40$ nT; $AE > 500$ nT).

Period Covered	April 2001			December 2006		
Model	Kp_{t+6}^{Ram}	Dst_{t+6}^{Ram}	AE_{t+6}^{Ram}	Kp_{t+6}^{Ram}	Dst_{t+6}^{Ram}	AE_{t+6}^{Ram}
Hits	3	34	4	3	9	2
Misses	11	30	31	2	1	20
False Alarms	1	17	6	0	12	1
Right Reject.	235	169	209	231	214	213
Total	250	250	250	235	236	236
Max Lead Time	6 hours	6 hours	6 hours	6 hours	6 hours	6 hours
HSS	0.316	0.471	0.123	0.746	0.555	0.141
POD	0.214	0.531	0.114	0.600	0.900	0.090
FAR	0.250	0.333	0.600	0.000	0.571	0.333

Chapter 6

Discussion and Conclusion

In this study, I have investigated the behavior of the Boyle Index, an empirical form to approximate the polar cap potential, and its role as a forecasting parameter. After examining the BI using multiyear data covering a solar cycle (1997-2007), I draw the following two conclusions: (1) the BI can characterize the strength and magnitude of the magnetospheric convection system as it was originally conceived to be, and (2) offers an improvement in the predictability of one of the traditional geomagnetic indices Kp, Dst and AE.

I have used the ACE level 2 data of SWEPAM, IMP-8 and WIND extending from 1995 to 2007 to derive BI and thus investigated its statistical correlation with three magnetic indices (Kp, Dst and AE). I 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, I 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 lead time around 3 hours for both the one-hour and three-hour predictions. My studies have also shown that the BI is strongly correlated with Dst and AE, and that they are predictable in the shorter time range (6 hours). 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. I have further analyzed a few preconditioning mechanisms and the magnetosphere's response to such mechanisms by using artificial neural networks. In-

corporating sufficiently long weighted input averages into the training process helped the network learn the signatures of extended activity, and the evidences were noticed in my confirmatory tests.

Motivated by the extensive amount of in situ data available for research, the space physics community is keen on looking for ways to connect various magnetospheric processes with solar wind parameters. Therefore, it is not surprising that several similar solar wind-magnetosphere coupling functions which can also represent Kp and the state of the magnetosphere have been identified in the literature. A recent effort by Newell et al. [2007], with their comprehensive list of solar wind-magnetosphere coupling functions, is certainly noteworthy. Much more recently, unlike empirical approximations like the BI and Newell functions, Borovsky [2008] derived a new solar wind-magnetosphere coupling function called as the *control function* weighing heavily on the rudiments of the physics of dayside reconnection.

Recent research results tend to agree that the benefits of applying the concept of coupling functions to space weather is derived from the presence of a solar wind-dependent “viscous” interaction term in addition to the IMF-dependent “merging” term, the two most dominant processes driving magnetospheric convection, thereby, making them good candidates to describe the state of the magnetosphere system. The results borne out of this dissertation further corroborate to this. I also demonstrated that from the timeline of the speed of the solar wind and its IMF direction using information obtained from an upstream solar wind monitor, one can effectively increase the lead-time of forecasts before an ensuing storm and subsequent auroral substorms to provide short-term, yet close-to accurate, forecasts. Furthermore, I have shown that the forecast accuracies obtained from the ANNs are “significantly” better than persistent forecasting. My algorithms have better accuracy and lead time over some popular algorithms, especially the Costello ANN Kp model.

6.1 Summary

I have now successfully constructed and evaluated computer algorithms based on an artificial neural network to forecast the geomagnetic activity indices Kp, Dst and AE in real time from a solar wind coupling function called the *Boyle Index*. My analyses indicate a slight run-down in the prediction accuracy with increasing lead-time; the best prediction efficiency was achieved for an 1-hour lead time. However, if the BI is coupled with a dynamic pressure term (nv_{sw}^2), the predictive power increases.

My retrospective tests and network performance (based on randomly chosen test data out of 1997-2007) indicate that our algorithm can give a reliable Kp forecast with a lead-time of 1-hour within ± 0.7 Kp on an average for $Kp > 4$, when using the past information of the BI and Kp. For predictions with a lead-time of 3 hours, our models can predict within ± 1.0 Kp on an average for $Kp > 4$. I have also shown that we can predict within ± 0.9 of the true Kp ($Kp > 4$) for a lead-time of an hour and within ± 1.1 for a lead-time of 3-hours, all while using the solar wind alone. Finally, running our models 3 and 4 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. Furthermore, our 1-hour and 3-hour Dst models can offer predictions within ± 13.91 and 13.42 nT RMS uncertainties respectively. Finally, our 1-hour and 3-hour AE models can offer predictions of the upcoming AE values within ± 122.04 and 129.72 nT RMS uncertainties respectively.

Two of the Kp-prediction models discussed in this dissertation (models 1 and 2) have culminated into real time “space weather” forecast systems already. We have been running these two purely-solar wind-driven models since October, 2007

with good success. We are in the process of implementing a modified version of the existing alert system to send out “red-alerts” and “warnings” to the subscribers of our “spacalrt” by giving key considerations to the new results of this dissertation. The thresholds for email alerts have been modified based on model validations, specifically, to include model predictions. Soon, our subscribers will receive notices and warnings through this new system in the event of any severe geomagnetic activity that may be ongoing or imminent i.e., whenever the 10-minute BI exceeds 200 kV, and whenever the predicted Kp exceeds 6.

To sum up, I have successfully developed real time Kp, Dst and AE forecast models operational in the following modes:

(1*) a model that takes only solar wind and magnetospheric data from ACE to derive the Boyle index and predicts Kp 1 hour ahead;

(2*) a model that takes only the solar wind data and predicts Kp for a full 3 hours ahead;

(3) a model that takes solar wind and magnetospheric data from ACE and Kp from the NOAA (http://www.swpc.noaa.gov/rt_plots/kp_3d.html) and predicts Kp 1 hour ahead;

(4) a model that takes the same input as (3) but predicts Kp 2 hours ahead;

(5) a model that takes only the solar wind data and predicts Dst 1 hour ahead;

(6) a model that takes only the solar wind data and predicts Dst full 3 hours ahead;

(7) a model that takes only the solar wind data and predicts AE 1 hour ahead;

** Models 1 and 2 are considered the best among the Kp models because of lack of reliance on Kp history and therefore, free of persistence contamination, and also because they do reasonable well at predictions.*

(8) and a model that takes only the solar wind data and predicts AE full 3 hours ahead.

6.2 Kp models: Possible applications

Relativistic electrons trapped in the Earth’s radiation belt can influence a variety of phenomenon posing radiation hazards to spacecraft and humans on the Earth. Relativistic Electron Forecast Model (REFM), currently operational at NOAA, predicts electron fluxes using solar wind inputs from ACE to produce a three day forecast and warning customers of such “killer” electrons. Currently, the REFM is driven by the Costello ANN Kp model which has a lead time of approximately 1 hour. One of the model deficiencies, however, lies in its forecast uncertainties.

The Dynamic Radiation Environment Assimilation Model (DREAM) is an ongoing research thrust, based on data assimilation, aimed at improving the radiation belt forecasting at the Los Alamos National Laboratory (LANL), of which the physics-based radial diffusion code is a component. One of the key ingredients of this model is the Kp dependent radial diffusion coefficient. As an integral part of the DREAM effort at LANL, the radial diffusion model describes the radial evolution of the highly energetic electrons in geosynchronous orbits. Presently, this physics model is running in 1-dimensional mode with Kp as one of its inputs, through the Costello ANN Kp model. The specific form of the equation used in the radiation belt model, derived from the original Fokker-Plank equation, is given by:

$$\frac{\partial f}{\partial t} = L^2 \frac{\partial}{\partial L} \left(\frac{D_{LL}}{L^2} \frac{\partial f}{\partial L} \right) - \frac{f}{\tau(L, t)}, \quad (6.1)$$

where $f(L, t)$ is the phase space density, $D_{LL}(L, Kp)$ is the radial diffusion coefficient and $\tau(L, t)$ is the loss lifetime. L denotes the L-shell values. The model adopts the empirical form of the diffusion coefficient [Brautigam and Albert, 2000] described by

the following power law:

$$D_{LL}(L, Kp) = 10^{(0.506Kp-9.325)} L^{10} \quad (6.2)$$

The scope of this model is to study the radiation belt electron enhancements so that we can predict the response of the radiation belt given the solar wind conditions. While the DREAM effort at LANL is still evolving, data assimilation and tool testing can be facilitated with new and advanced procedures.

Now that it has been established that my models perform better than the Costello NN Kp model, in addition to its success in providing a longer lead time of up to 3 hours, my Kp-prediction models which are all global physics-based models can be used to drive the radial diffusion code and the REFM for better accuracy and lead time.

6.3 Future Possibilities

Contrary to the age-old belief that southward IMF is the only condition for which plasma can enter the magnetosphere, recent results from THEMIS spacecraft suggests that Earth's plasma sheet becomes colder and denser for the plasma to force their way into the magnetosphere all happening when the IMF is northward. If so, the new solar cycle might pose a new problem challenging the best algorithms that were trained using preceding solar cycles. Since the effects of preconditioning and mass loading are critical to predicting the onset of storms, my neural network algorithm with its ability to look back up to 18 hrs in time should be effective at being able to capture these effects. Therefore, looking near term, a modified algorithm or an algorithm based on modified form of the Boyle potential is inevitable as the new solar cycle begins to ramp up.

6.3.1 Borovsky Function

Borovsky [2008] derived another coupling function R , called the solar wind “control function” (CF), by taking the dayside reconnection rate parameters into account. The control function R is given by the following expression,

$$R = 0.4\mu_o^{1/2}\sin(\theta/2)\rho_o v_o^2(1 + 0.5M_{ms}^{-2})(1 + \beta_s)^{-1/2} \cdot \left[C\rho_o + (1 + \beta_s)^{-1/2}\rho_m \right]^{-1/2} \left[1 + \beta_s \right]^{1/2} + 1 \Big]^{-1/2} \quad (6.3)$$

where ρ_o is the mass density of the solar wind upstream of the bow shock, v_o is the velocity of the solar wind upstream of the bow shock, C is the compression ratio of the bow shock, β_s is the plasma- β value of the magnetosheath plasma near the nose, and M_{ms} is the magnetosonic Mach number of the solar wind. Supplements to equation 6.1 are:

$$\beta_s = 3.2 \times 10^{-2} M_A^{1.92}; \quad (6.4)$$

$$C = \left[(1/4)^6 + [1/1(1 + 1.38\log_e(M_A))]^6 \right]^{-1/6} \quad (6.5)$$

$$M_{ms} = v_o \left((B_o^2/\mu_o\rho_o) + 5P_o/3\rho_o \right)^{1/2} \quad (6.6)$$

$$M_A = v_o(\mu_o\rho_o)^{1/2}/B_o \quad (6.7)$$

A simple linear plot (figure 6.1) shows the difference between the two functions, the BI ($r = 0.72$) and the CF (0.81). Clearly, the CF gives a large improvement over the existing baseline, at least with respect to predicting the Kp index. Further studies are needed to test this function as a forecasting tool.

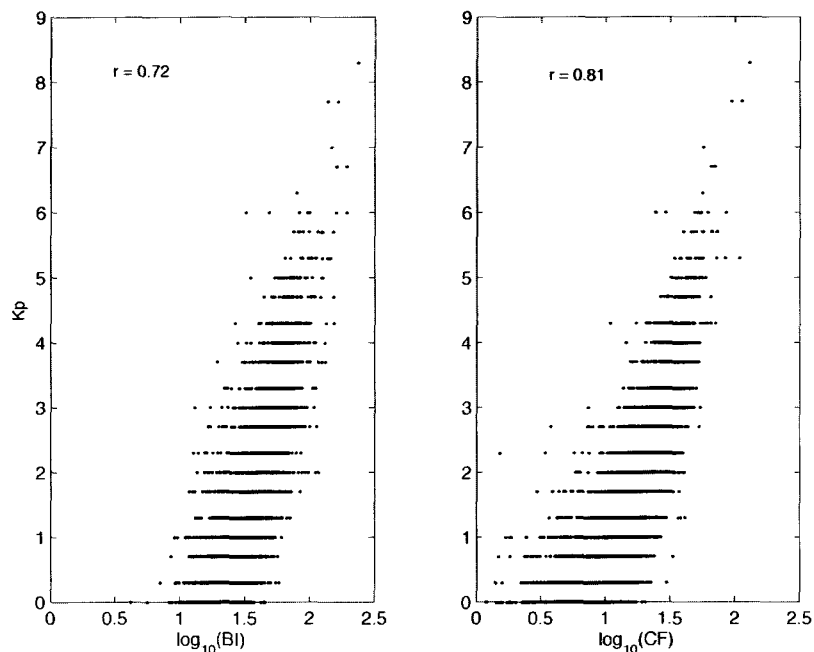


Figure 6.1 : Scatter plots of the 3-hour averages of the BI and the CF versus the Kp is shown here.

6.3.2 Improving long term forecasts

One of my future interests also lies in finding techniques and methodologies to improve long-term predictions. Recent trends in space weather research include long-term end-to-end forecast models that are currently being developed. However, one of the key ingredients still unclear in these models is the polarity of the IMF B_Z at the leading edge of Coronal Mass Ejections (CMEs) emanating from the Sun, which are critical to simulating the solar wind magnetospheric interactions, and therefore to space weather forecasts. In the absence of reliable methods as yet, either to predict the strength of CMEs or to the direction of IMF B_Z , photospheric signatures on the surface of the Sun offer initial clues (e.g., The Wang-Sheeley-Arge Model). But forecasts based on such observations may produce false alarms due to the lack of knowledge of the exact

geometry of the CMEs.

One plausible approach which might offer some valuable insights to the direction of the IMF would be identifying the reversal of polarity from northward to southward IMF and vice-versa. This can be done through neuro-computing using a Green's function input of the solar wind parameters to see how they differ. I would like to conduct temporal studies of the magnetic vector fields of solar wind following, for example, a CME, using data from the ACE or the VSTO (Virtual Solar Terrestrial Observatory), so the reversal of the field directions and its timeline during such transient events can be investigated in detail.

An advanced non-linear approach like neural networks is relatively new and has delivered good results, especially when traditional statistical approaches based on regression analysis have not performed well. In order to leverage the merits of ANN, I would also like to pursue and encourage collaborative efforts within the various disciplines of space physics including studies related to solar transient events and long-term solar activity.

Finally, the best possibility for long range predictions might be the establishment of a more sunward monitor, perhaps solar sails, to increase the solar wind lead time from ≈ 45 minutes to 3-4 hours.

Appendix

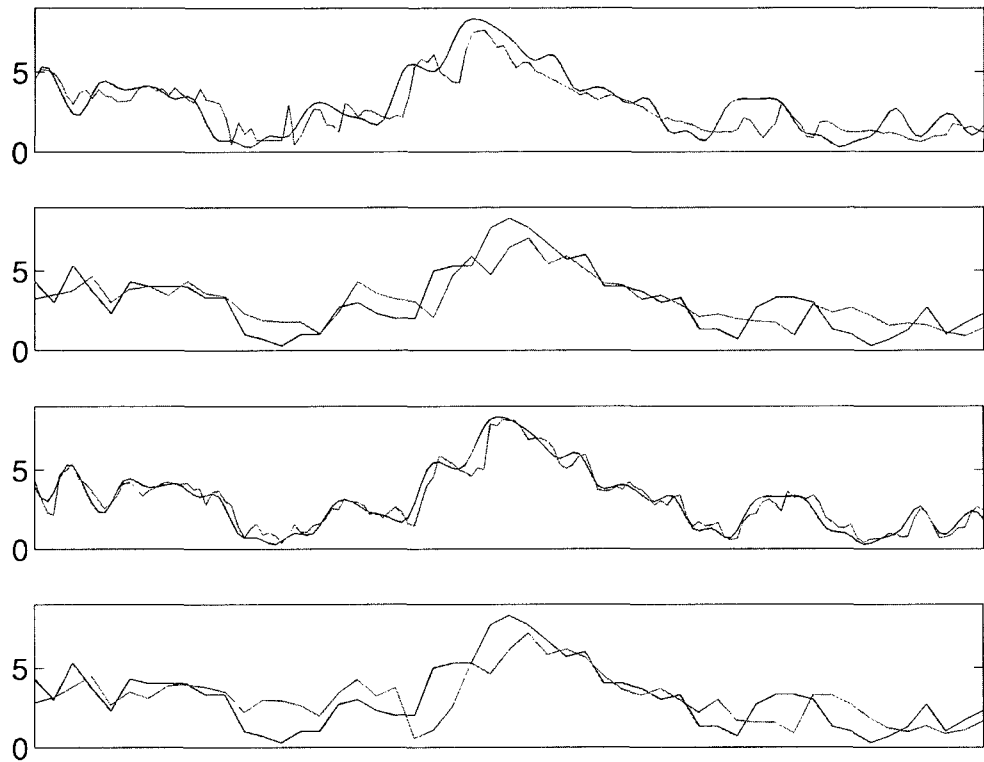


Figure 6.2 : Supplement plot to figure 5.30 (page 160). Model vs. data comparison: Predictions from all the Kp models are overplotted with the measured Kp.

References

- [Baker, 1998] Baker, D. N. (1998), What is Space Weather?, *Adv. Space Res.*, 22, 1, 7-16, 1998
- [Bala et al., 2009] Bala, R., P. H. Reiff, and J. E. Landivar (2009), Real-time prediction of magnetospheric activity using the Boyle Index, *Space Weather*, 7, No. 4, S04003, 2009.
- [Bartels et al., 1939] Bartels, J., N.H. Heck, and H. F. Johnston (1939), The three-hour-range index measuring geomagnetic activity, *J. Geophys. Res.*, 44, 411, 1939.
- [Baumjohann and Paschmann, 1987] Baumjohann, W., and Gotz Paschmann (1987), Solar Wind-Magnetosphere Coupling: Process and Observations, *Physica Scripta*, T18, 61, 1987.
- [Behnke, 2008] Behnke, R. A. (2008), Space Weather as a hybrid of basic research and applied science, *Space Weather*, 6, S09007, 2008
- [Berthelier, 1978] Berthelier, A. (1978), Etude des influences du vent solaire sur l'activite magnetique terrestre, particulierement aux hautes latitude doctorat d'etat, Univ. Pierre et Marie Curie, 1979
- [Boberg et al., 2000] Boberg, F., P. Wintoft, and H. Lundstedt (2000), Real-time Kp prediction from solar wind data using neural networks, *Phys. Chem. Earth*, 25, 275, 2000

- [Borovsky et al., 2006] Borovsky, J. E., and M. H. Denton (2006), Differences between CME-driven storms and CIR-driven storms, *J. Geophys. Res.*, *111*, A07S08, 2006
- [Borovsky, 2008] Borovsky, J. E. (2008), The rudiments of a theory of solar wind/magnetosphere coupling derived from first principles, *Journal of Geophysical Research*, Volume 113, Issue A8.
- [Boyle et al., 1997] Boyle, C. B., P. H. Reiff, and M. R. Hairston (1997), Empirical polar cap potentials, *J. Geophys. Res.*, *102*, 111, 1997
- [Cane and Richardson, 2003] Cane, H. V., and I. G. Richardson, Interplanetary coronal mass ejections in the near-Earth solar wind during 1996–2002, *J. Geophys. Res.*, *108*, A009817, 2003.
- [Carbary, 2004] Carbary, J. F. (2004), Aurora Boundaries Quantified by Geomagnetic Index, American Geophysical Union, Fall Meeting 2004.
- [Costello, 1997] Costello, K. A. (1997), Moving the Rice MSFM into a real-time forecast mode using solar wind driven forecast models, *Ph.D. dissertation, Rice university, Houston, Texas*.
- [Detman and Joselyn, 1999] Detman, T., and J. A. Joselyn (1999), Real-time Kp predictions from ACE real time solar wind, in *Solar Wind Nine*, edited by S. R. habbal et al., *AIP Conf. Proc.*, 471, 729-732
- [Feldstein and Starkov, 1967] Feldstein, Y., I and G. V. Starkov (1967), Dynamics of auroral belt and polar geomagnetic disturbances, *Planetary and Space Science*, Vol. 15, p.209.

- [Fok et al., 2001] Fok, M.-C., Wolf, R. A., Spiro, R. W., and Moore, T. E. (2001), Comprehensive computational model of Earth's ring current, *Journal of Geophysical Research*, Volume 106, Issue A5, p. 8417-8424, 2001.
- [Hairston et al., 2003] Hairston, M. R., T. W. Hill, and R. A. Heelis (2003), Observed saturation of the ionospheric polar cap potential during the 31 March 2001 storm, *J. Geophys. Res. Lett.*, *30(6)*, 1325, 2003
- [Haykin, 1999] Haykin, Simon (1999), Neural Networks: A Comprehensive Foundation, *McMillan, New Jersey*, 1999
- [Hill et al., 1976] Hill, T. W., A. J. Dessler, and R. A. Wolf (1976), Mercury and Mars: The role of ionospheric conductivity in the acceleration of magnetospheric particles, *J. Geophys. Res. Lett.*, *3*, 429, 1976
- [Hill, 1983] Hill, T. W., (1983), Solar-wind Magnetosphere Coupling, *Solar-Terrestrial Physics*, 261-302, 1983
- [Hsu and McPherron] Hsu, Tung-Shin; McPherron, Robert L. (2003), Occurrence frequencies of IMF triggered and nontriggered substorms, *Journal of Geophysical Research Space Physics*, Volume 108, Issue A7, pp. SMP 23-1.
- [Johnson and Wing, 2005] Johnson, J. R., and S. Wing (2005), A solar cycle dependence of nonlinearity in magnetospheric activity, *J. Geophys. Res.*, *110*, A04211, 2005
- [Kamide, 1988] Kamide, Y. (1988), Electrodynamic Processes in the Earth's Ionosphere and Magnetosphere, Kyoto Sangyo University Press, Kyoto, Japan, 1988.
- [Kanji, 1999] Kanji, Gopal K. (1999), 100 Statistical Tests, *SAGE Publications*, 1999

- [Koons and Gorney, 1991] Koons, H. C. and Gorney, D. J. (1991), A neural network model of the relativistic electron flux at geosynchronous orbit, *Journal of Geophysical Research* (ISSN 0148-0227), vol. 96, April 1, 1991, p. 5549-5556.
- [Lavraud et al., 2006] Lavraud, B., M. F. Thomsen, J. E. Borovsky, M. H. Denton, and T. I. Pulkkinen (2006), Magnetosphere preconditioning under northward IMF: Evidence from the study of coronal mass ejection and corotating interaction region geoeffectiveness, *J. Geophys. Res.*, 111, A09208, 2006
- [Li et al., 2007] Li, X., K. S. Oh, and M. Temerin (2007), Prediction of the AL index using solar wind parameters, *J. Geophys. Res.*, 112, A06224
- [Lyons et al., 1997] Lyons, L. R.; Blanchard, G. T.; Samson, J. C.; Lepping, R. P.; Yamamoto, T.; Moretto, T. (1997), Coordinated observations demonstrating external substorm triggering, *Journal of Geophysical Research*, Volume 102, Issue A12, p. 27039-27052.
- [MacDougall et al., 2006] MacDougall, J. W., and P. T. Jayachandran (2006), Polar cap voltage saturation, *J. Geophys. Res.*, 111, A12306, 2006
- [Manvielle et al., 1991] Manvielle, M., and A. Berthelier (1991), The K-derived planetary indices: Description and availability, *Rev. Geophys.*, 29, 415, 1991
- [Masters, 1995] Masters, Timothy (1995), *Neural, Novel & Hybrid Algorithms for Time Series Prediction*, John Wiley and Sons, Inc, 1995
- [Mayaud, 1978] Mayaud, P. N. (1978), Morphology of the transient irregular variation of the terrestrial magnetic field, and their main statistical laws, *Ann. Geophys.*, 34, 243, 1978.

- [Mehrotra et al., 1996] Mehrotra K, Chilukuri K. Mohan and Sanjay Ranka (1996), Elements of Artificial Neural Networks, The MIT press, 1996.
- [Moldwin, 2008] Moldwin, M (2008), An Introduction to Space Weather, Cambridge University Press, 2008.
- [Nagai, 1993] Nagai, Akira (1993), Prediction of magnetospheric parameters using artificial neural networks (1993), *Ph.D. dissertation, Rice University, Houston, Texas.*
- [Newell et al., 2002] Newell, P. T., Sotirelis, T., Ruohoniemi, J. M., Carbary, J. F., Liou, K., Skura, J. P., Meng, C.-I., Deehr, C., Wilkinson, D., and Rich, F. J. (2002), OVATION: Oval variation, assessment, tracking, intensity, and online nowcasting, *Annales Geophysicae*, vol. 20, Issue 7, pp.1039-1047, 2002.
- [Newell et al., 2007] Newell, P. T., T. Sotirelis, K. Liou, C.-I. Meng, and F. J. Rich (2007), A nearly universal solar wind-magnetosphere coupling function, *J. Geophys. Res.*, 112, A01206, 2007
- [O'Brien and McPherron, 2000] O'Brien, T. P.; McPherron, R. L. (2000), An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay *Geophysical Research Letters*, Volume 105, Issue A4, p. 7707-7720.
- [Papitashvili et al., 2000] Papitashvili, V. O., N. E. Papitashvili, and J. H. King (2000), Solar cycle effects in planetary geomagnetic activity: Analysis of 36-year long OMNI dataset, *J. Geophys. Res. Lett.*, 98, 2797, 2000
- [Parks, 2004] Parks, G. K. (2004), Physics of Space Plasmas, *Westview Press*, 2004
- [Reiff, 1999] Reiff, P. H. (1990), The Use and Misuse of Statistics in Space Physics, *J. Geomag. Geoelec.*, 42, 1145-1174, 1990

- [Reiff and Luhmann, 1986] Reiff, P. H. and J. G. Luhmann (1986), Solar Wind Control of the Polar-Cap Voltage, *Solar Wind-Magnetosphere Coupling*, edited by Y. Kamide and J. A. Slavin, 453-476, 1986
- [Rumelhart et al., 1986] Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986), Learning representations by back-propagation errors, *Letters to Nature*, 323, 533, 1986
- [Russell, 2007] Russell, C. T. (2007), The Coupling of the solar wind to the Earth's magnetosphere, *Space Weather*, Springer Berlin Heidelberg, 103, 2007
- [Singer et al., 2001] Singer, H. J., G. R. Heckman, and J. W. Hirman (2001), Space Weather Forecasting: A Grand Challenge, *Space Weather, Geophysical Monograph*, 125, 23, 2001
- [Siscoe et al., 2002a] Siscoe, G. L., N. U. Crooker, G. M. Erickson, B. U. O. Sonnerup, N. C. Maynard, J. A. Schoendorf, K. D. Siebert, D. R. Weimer, W. W. White, and G. R. Wilson (2002a), Hill model of transpolar potential saturation: Comparisons with MHD simulations, *J. Geophys. Res.*, 107(A6), 1075.
- [Takahashi et al., 2001] Takahashi, K., B. A. Toth, and J. V. Olson (2001), An automated procedure for near-real-time Kp estimates, *J. Geophys. Res.*, 106, 017, 2001
- [Temerin and Li, 2002] Temerin, M., and X. Li (2002), A new model for the prediction of Dst on the basis of the solar wind, *J. Geophys. Res.*, 107(A12), 1472, doi:10.1029/2001JA007532.
- [Thomsen, 2004] Thomsen, M. F. (2004), Why Kp is such a good measure of magnetospheric convection, *Space Weather*, 2, S11004, 2004.

- [Vassiliadis, 2007] Vassiliadis, Dimitris, Forecasting space weather in *Space Weather - Physics and Effects*, *Springer Berlin Heidelberg*, 14, 403, 2007
- [Vasyliunas et al., 1982] Vasyliunas, V. M., J. R. Kan, G. L. Siscoe, and S.-I. Akasofu (1982), Scaling relations governing magnetospheric energy transfer, *Planet. Space Sci.*, 30, 359.
- [Vogt et al., 2006] Vogt, M. F., C. M. S. Cohen, P. Puhl-Quinn, V. K. Jordanova, C. W. Smith, and R. M. Skoug (2006), Space weather drivers in the ACE era, *Space Weather*, 4, S09001, 2006
- [Wasserman, 1989] Wasserman, P. D., Neural Computing, *Van Nostrand Reinhold*, *New York*, 1989
- [Wilks, 1995] Wilks, D. S., Statistical Methods in the Atmospheric Sciences, *Academic Press, Inc.*, 1995
- [Wing et al., 2005] 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, 2005
- [Wu et al., 1997] Wu, Jian-Guo and Henrik Lundstedt (1997), Geomagnetic storm predictions from solar wind data with the use of neural networks, *J. Geophys. Res.*, 102, 255, 1997
- [Wu et al., 2002] Wu, C.-C., and R. P. Lepping, Effect of solar wind velocity on magnetic clod-associated magnetic storm intensity, *J. Geophys. Res.*, 107, A11, 1346, 2002
- [Wygant et al., 1983] Wygant, J. R., Torbert, R. B., and Mozer, F. S. (1983), Comparison of S3-3 polar cap potential drops with the interplanetary magnetic field

and models of magnetopause reconnection, *Journal of Geophysical Research* (ISSN 0148-0227), vol. 88, July 1, 1983, p. 5727-5735.