

6-25-2010

# A stochastic ensemble forecast model for geosynchronous relativistic electron fluxes

Steven Nelson

Follow this and additional works at: [https://digitalrepository.unm.edu/ece\\_etds](https://digitalrepository.unm.edu/ece_etds)

---

## Recommended Citation

Nelson, Steven. "A stochastic ensemble forecast model for geosynchronous relativistic electron fluxes." (2010).  
[https://digitalrepository.unm.edu/ece\\_etds/191](https://digitalrepository.unm.edu/ece_etds/191)

This Thesis is brought to you for free and open access by the Engineering ETDs at UNM Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

Nelson, Steven G.

*Candidate*

Electrical & Computer Engineering

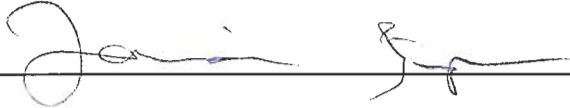
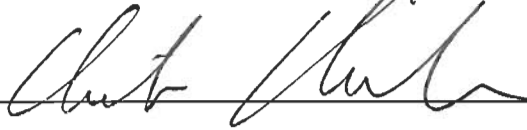
*Department*

This thesis is approved, and it is acceptable in quality and form for publication:

*Approved by the Thesis Committee:*



Chairperson



**A STOCHASTIC ENSEMBLE FORECAST MODEL FOR  
GEOSYNCHRONOUS RELATIVISTIC ELECTRON FLUXES**

**BY**

**STEVEN G. NELSON**

**B.S., VIRGINIA POLYTECHNIC INSTITUTE AND STATE  
UNIVERSITY, 2009**

**THESIS**

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

**Master of Science  
Electrical Engineering**

The University of New Mexico  
Albuquerque, New Mexico

**May, 2010**

**A STOCHASTIC ENSEMBLE FORECAST MODEL FOR  
GEOSYNCHRONOUS RELATIVISTIC ELECTRON FLUXES**

**BY**

**STEVEN G. NELSON**

**ABSTRACT OF THESIS**

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

**Master of Science  
Electrical Engineering**

The University of New Mexico  
Albuquerque, New Mexico

**May, 2010**

# **A Stochastic Ensemble Forecast Model for Geosynchronous Relativistic Electron Fluxes**

by

**Steven G. Nelson**

B.S., Virginia Polytechnic Institute and State University, 2009

M.S., University of New Mexico, 2010

## **Abstract**

A stochastic ensemble model composed of three functional forecasting models has been developed to forecast  $>2$  MeV electron flux at geosynchronous (GEO) orbit. The REFM model is based on a statistical link between electron flux and solar wind speed using empirically derived linear filter coefficients, the Li model solves a radial diffusion equation with a diffusion coefficient that is a function of the solar wind velocity and interplanetary magnetic field, and the FLUXPRED model is a multi-layer feed-forward neural network with electron flux and summed  $K_p$  as input. Individual model results were combined using a multivariate regression to produce significantly better predictive results than any of the individual models alone. A stochastic model is then developed to forecast the probability that a fluence threshold will be exceeded. The regression technique, model optimization, and calculation of forecast probability will be discussed in reference to the ensemble model.

# Contents

<b>1 Introduction</b> .....	<b>1</b>
1.1 Motivation .....	1
1.2 Background.....	2
1.3 Input Parameters for Forecasting Models.....	7
<b>2 Methods and Models</b> .....	<b>11</b>
2.1 Data Preprocessing .....	11
2.2 REFM Model.....	15
2.3 Fluxpred Model .....	18
2.4 Li Model.....	24
2.5 Ensemble Model .....	26
<b>3 Results</b> .....	<b>31</b>
3.1 Historical Ensemble Model .....	31
3.2 Predictive Ensemble Model .....	33
3.3 Stochastic Ensemble Model .....	39
<b>4 Conclusions</b> .....	<b>55</b>
<b>References</b> .....	<b>57</b>

# Chapter 1

## Introduction

### 1.1 Motivation

Charged particle radiation is known to be deleterious to spacecraft electronics, leading to temporary and sometimes permanent degradation. Sustained periods of high energy electron fluxes ( $>2$  MeV) can cause an electric field to form within an insulating material internal to the spacecraft. Once this field exceeds the material's breakdown voltage, a discharge may occur that couples to other internal electronics. This phenomenon, known as deep dielectric charging, can lead to considerable side effects such as spacecraft shutdown, thruster ignition, or control anomalies, to name a few. Particularly of interest is the radiation environment at geosynchronous (GEO) orbit where a large number of important commercial and military satellites operate. In fact, anomalous control was seen during momentum wheel malfunction on two Canadian ANIK communications spacecraft in January 1994 as a result of sustained high energy electron flux for seven days prior to the event. February 1994 is another famous example of deep dielectric charging; the satellite feed relaying the Winter Olympic Games on the BS-3a spacecraft was interrupted during a final historical ski jump for the Japanese team due to a prior two-week period of high energy electron flux [*Shea and Smart, 1998*].

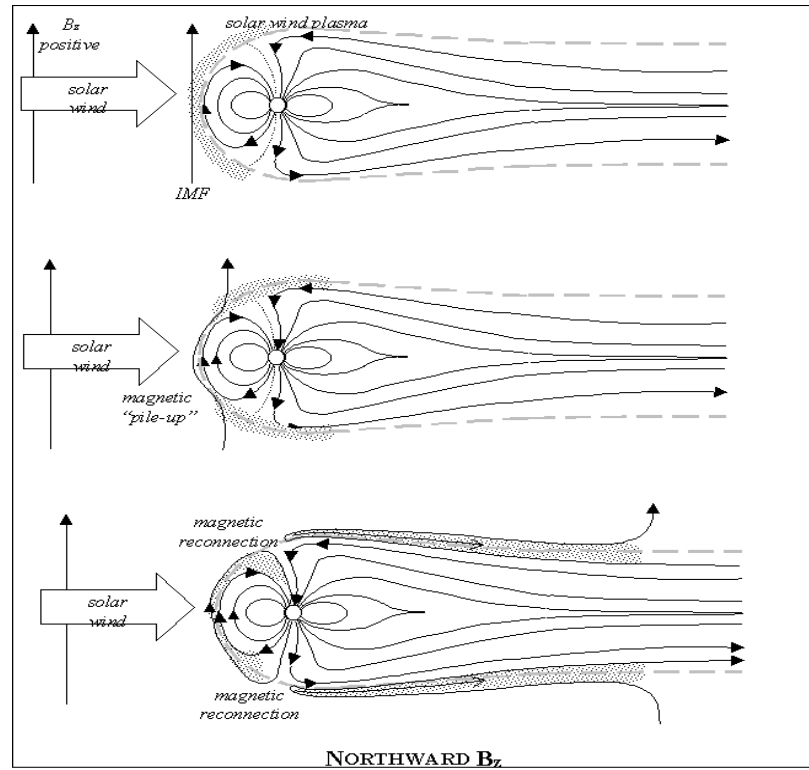
Thus, the usefulness in predicting when the radiation environment becomes hazardous is

apparent for protecting both commercial and military assets in space. A variety of attempts have been made to forecast high energy flux ( $>2$  MeV) at GEO [Nagai 1988; Baker et. al 1990; Koons and Gorney 1991; Li 2004; Ling et. al 2010]. An ensemble forecasting model is proposed with the idea that contributing the strengths of these models should result in a better answer than any one of the models alone. The three models chosen for this ensemble each use separate approaches to model the relativistic geosynchronous electron flux. The Relativistic Electron Flux Model (REFM) model is based on a linear predictive filter [Baker et. al, 1990], the Li model is based on radial diffusion [Li, 2004], and the Fluxpred model employs a neural network [Ling et. al 2010].

## **1.2 Background**

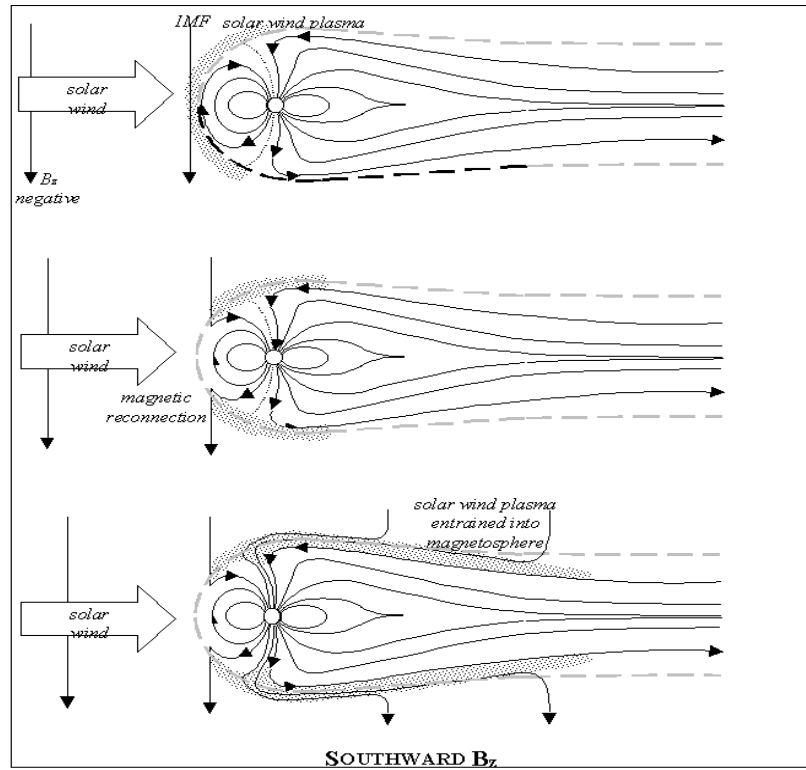
The earth's magnetosphere forms a teardrop-shaped cavity in which the earth resides. The shape of this cavity is a result of the deflection of solar wind particles from the intrinsic magnetic field generated around the earth. The solar wind energy input into the magnetosphere is greatly dependent on the direction of the  $z$ -component of the interplanetary magnetic field (IMF), also known as  $B_z$ . When the  $B_z$  points northward, magnetic field lines tend to "pile up" on the earth's magnetopause (the boundary between the earth's magnetic field and the solar wind) and magnetic reconnection occurs in the tail region, as shown in Figure 1.





**Figure 1.** Magnetic field line configuration during a northward pointing  $B_z$  (IMF). Excerpted from [Crown *et al.*, 2000].

When the  $B_z$  points southward, magnetic reconnection (a process by which separate magnetic field lines splice together to form a new field line) occurs on the “nose” of the magnetopause. This allows the solar wind to penetrate closer to the earth and entrain much of the solar wind plasma into the night-side magnetosphere, some of which appears on the earth as aurora. The configuration of this process is illustrated in Figure 2.

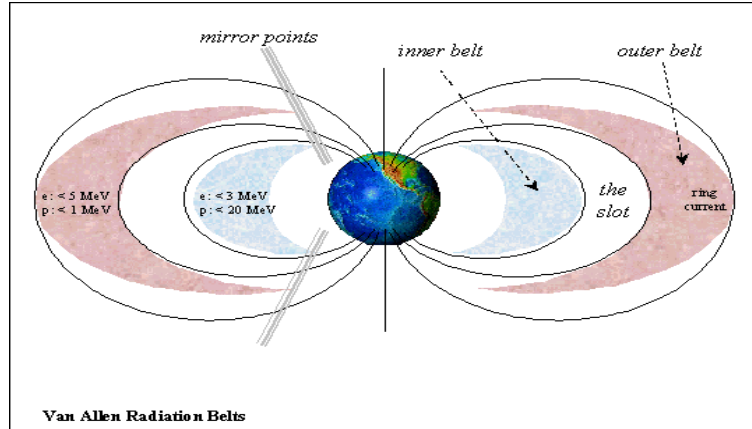


**Figure 2.** Magnetic field line configuration during a southward pointing  $B_z$  (IMF). Excerpted from [Crown et al., 2000].

Geomagnetic activity is therefore highly dependent on the IMF orientation. One measure of geomagnetic activity is a quantity known as the Kp index. The official planetary Kp-index is computed through a weighted average from a worldwide geomagnetic observatory network. The Kp-index is a quantification of the earth's horizontal magnetic field disturbance that ranges semi-logarithmically from 0-9, with Kp = 1 indicating calm weather and Kp  $\geq$  5 indicating a geomagnetic storm. This index is computed at regular 3-hour intervals where the maximum positive and negative deviations are summed to get the total maximum fluctuation, and this is the value known as  $\sum Kp$ . These geomagnetic storms are triggered from solar events such as coronal mass ejections (CMEs), high-speed streams from coronal holes, and solar flares.

Existing in the magnetospheric cavity are the inner and outer radiation belts, as illustrated

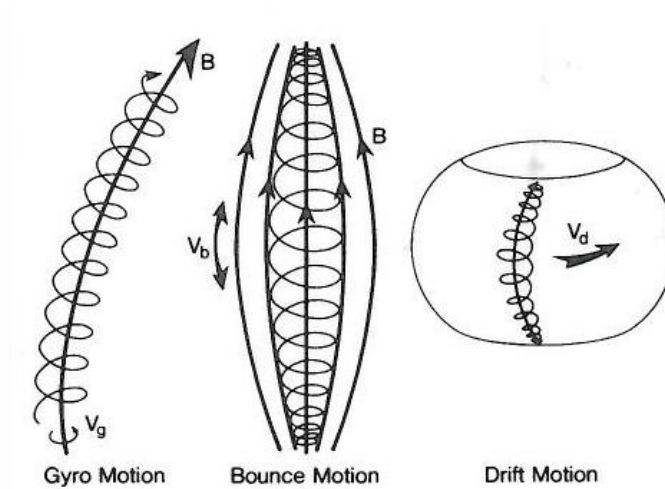
in Figure 3.



**Figure 3.** Illustration of the inner and outer radiation belts. Excerpted from [Crown *et al.*, 2000].

The inner radiation belt extends from  $\sim 100$ - $10,000$  km above the earth's surface. This is commonly denoted as extending from 0.01 to 1.5 earth radii, or from  $L \sim 0.01$  to  $L \sim 1.5$ . The inner belt is a relatively stable region that consists primarily of high energy protons, with particles sometimes being trapped for hundreds of years. The outer radiation belt extends from  $\sim L = 3$  to  $L = 7$ , and consists primarily of high-energy electrons. Converse to the inner belt, the particle population is highly dynamic, with the average electron lifetime on the order of 1-3 days.

Charged particles found on these magnetic field lines conserve what are known as the three adiabatic invariants. The first is associated with a particle gyrating around a field line, the second is associated with a particle bouncing back and forth from pole to pole, and the third is associated with the slow drift of the bouncing, gyrating particle around the earth. An illustration of these is shown in Figure 4.



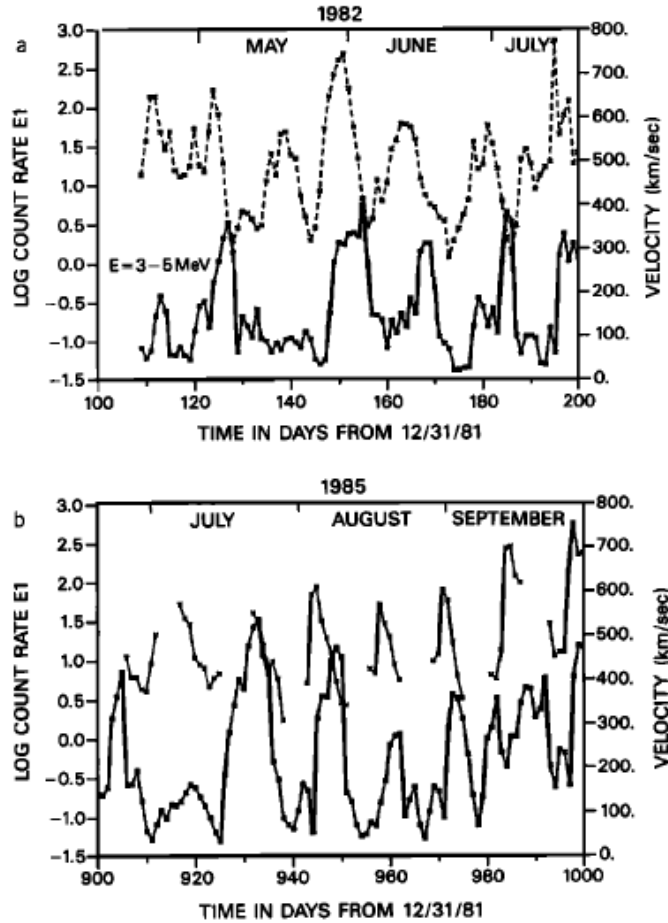
**Figure 4.** Illustration of the three adiabatic invariants. Excerpted from [Kivelson *et al.*, 1995].

Waves resonant with any of these invariants can transfer energy to particles traveling along the magnetic field lines, where a number of effects can occur. If either the first or second invariant is violated, the period of the particle bouncing back and forth will be changed, and in some cases provide it enough energy to have its “bounce point” (commonly denoted the mirror point) within the atmosphere, and thus be lost due to frictional effects. This occurs in the slot region between the radiation belts, which is why there is an absence of particles present in this location. Ultra low frequency (ULF) waves can be resonant with the particle drift period for a narrow range of particle energies that are located outside the outer radiation belt, causing the drift velocity  $V_d$  to increase. This then has the effect of migrating the particle inward to conserve the third adiabatic invariant, and thus the process of radial diffusion occurs. It is this mechanism by which the outer radiation belts are assumed populated. A physical analogy would be to consider the magnetosphere as a bell and the solar wind as a hammer: when the bell is struck, sound waves emanate from it; conversely, when the magnetosphere is “struck,” ULF waves are emitted from it [Engebretson *et al.*, 1998], energizing the particles further out and causing them to diffuse inward. The flux of these high energy electrons ( $> 2$  MeV)

migrating inward is the quantity that is sought to be predicted, specifically at GEO ( $L = 6.6$ ).

### **1.3 Input Parameters for Forecasting Models**

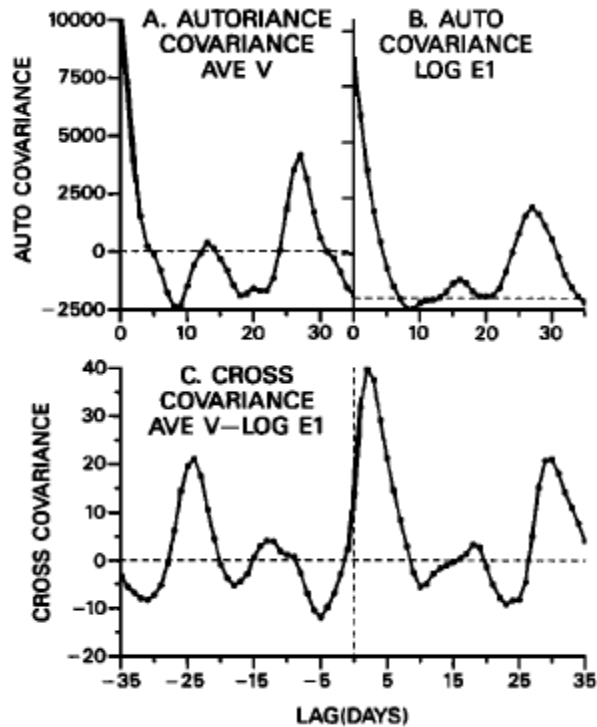
There are many possible drivers for a geosynchronous electron flux forecasting model, and the task of reducing these to a manageable set can be quite daunting. It is apparent from the preceding discussion that both the solar wind and geomagnetic fluctuations are correlated with the electron flux at GEO. This relationship is described by *Baker et al.* [1990]. Figure 5a shows the data between April and July of 1982 for solar wind speed and 3- to 5-MeV electron count rate at GEO, as measured by the Advanced Composition Explorer (ACE) and a Geostationary Operational Environment Satellite (GOES), respectively. Details of these satellites will be discussed in Chapter 2. Figure 5b shows the same for the period between July and October of 1985.



**Figure 5.** (a) Solar wind speed (top curve) and 3- to 5-MeV electron count rates (bottom curve) for April through July 1982, with the left axis corresponding to the bottom curve and the right axis corresponding to the top curve. (b) Similar to 5a for the period July to October 1985. Excerpted from [Baker *et al.*, 1990].

It can be seen that the 3-5 MeV electron counting rates tend to increase on the declining side of the solar wind speed peaks. The quality of the data in part *b* is much poorer, but the correlation between electron flux and solar wind speed can still be seen. The electron count rate E1 is shown as a daily average to remove diurnal variations and plotted as the logarithm. This suggests that fluctuations in solar wind velocity can be correlated with fluctuations in electron flux after a certain amount of time lag. Baker *et al.* [1990] did their correlation analysis as follows. The data was detrended so that both quantities have a mean of zero, and just the fluctuations are being compared. This was done by

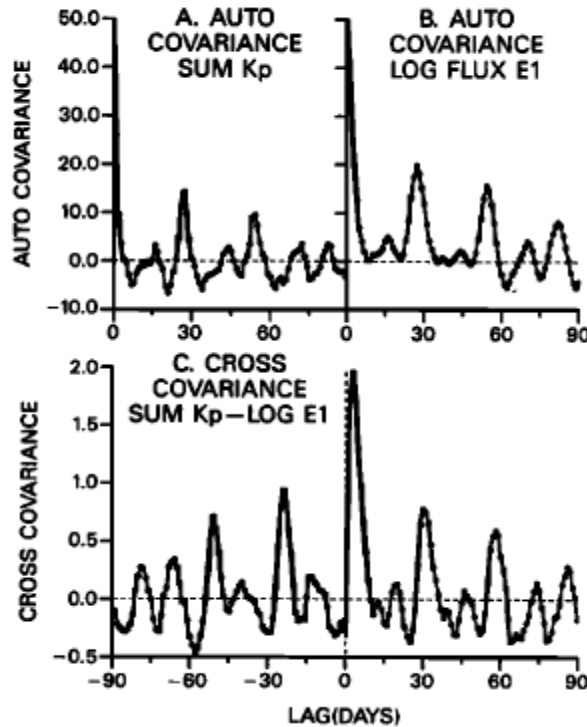
constructing 50-day running means of the input and output time series, fitting 5<sup>th</sup>-order polynomials to the sequence of means, and removing this polynomial from the raw data. After detrending the data, the autocovariance and cross covariance of the solar wind speed  $V$  (the input time series) and the electron count rate  $E1$  (the output time series) was determined, with the results shown in Figure 6.



**Figure 6.** (a) The autocovariance of available solar wind speed daily averages (1982-1985). (b) The autocovariance of the log of the 3- to 5-MeV electron counting rates (1982-1985). (c) The cross covariance of the parameters analyzed in Figures 6a and 6b. Excerpted from [Baker *et al.*, 1990].

The solar wind velocity (Figure 6a) shows a strong autocovariance peak for 27 days lag and a weaker peak for 13 days lag. The same relationship is shown for the logarithmic electron flux (Figure 6b). These are related to the period of a solar rotation being 27 days. The cross covariance of log  $E1$  and solar wind speed shows a strong correlation with a 2-3 day lag, with additional peaks at  $\pm 27$  days.

The same analysis done by *Baker et al.* [1990] using summed  $K_p$  fluctuations as the input time series and multi-MeV electron flux fluctuations as the output time series is an extension of earlier work done by *Nagai* [1988], and can be seen in Figure 7.



**Figure 7.** (a) The autocovariance of the  $\sum K_p$  data between April 1982 and 1985. (b) The autocovariance of the log of the 3- to 5-MeV electron counting rates (1982-1985). (c) The cross covariance of  $\sum K_p$  and log E1 (1982-1985). Excerpted from [*Baker et al.*, 1990].

As can be seen, the cross covariance of log E1 and summed  $K_p$  shows a strong correlation with a 1-2 day lag, with additional peaks at  $\pm 27$ , 54, and 81 days. It is for this reason that the ensemble model uses both solar wind velocity and summed  $K_p$  data as inputs to produce an electron flux forecast at GEO a maximum of 72 hours in advance.



# Chapter 2

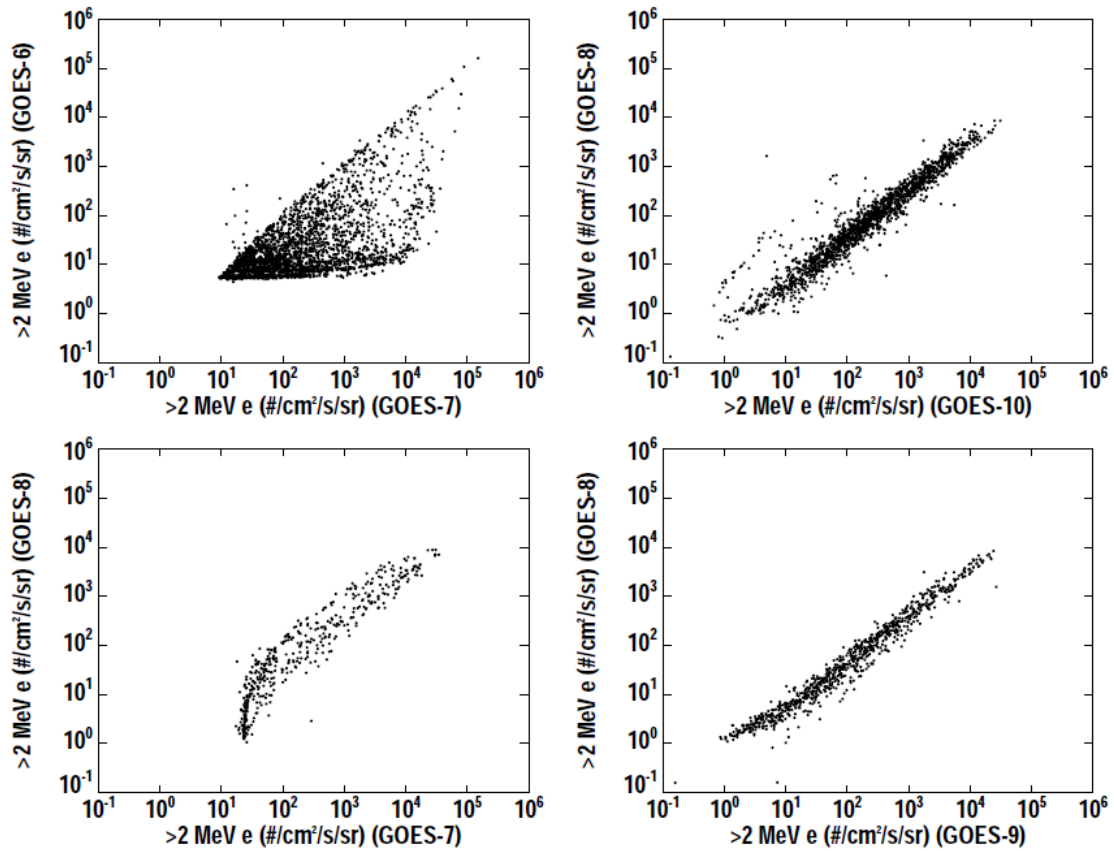
## Methods and Models

### 2.1 Data Preprocessing

The  $>2$  MeV electron flux used as both input and validation data to the ensemble model is measured at GEO by the Geostationary Operational Environmental Satellite (GOES) family of satellites. These satellites perform a number of functions such as weather forecasting, storm tracking, and meteorology research, and have been in continuous use since the launch of GOES-1 on October 16, 1975. From the beginning, this series of spacecraft has carried a Space Environment Monitor (SEM) package. The SEM contains four instruments: the energetic particles sensor (EPS), the high energy proton and alpha detector (HEPAD), the X-ray sensor (XRS), and a magnetometer. The EPS unit contains three channels for measuring electrons (E1, E2, and E3) that correspond to three different energy levels ( $\geq 0.6$  MeV,  $\geq 2$  MeV, and  $\geq 0.6$  MeV, respectively). Each channel is a set of two  $1500\text{-}\mu\text{m}$ ,  $25\text{-mm}^2$ , silicon surface detectors that each have an independent field of view with a different thickness moderator in front of them corresponding to the desired energy threshold. The silicon surface detectors output charge pulses to a preamplifier, which converts them to voltage pulses before being sent to a signal analyzer unit/data

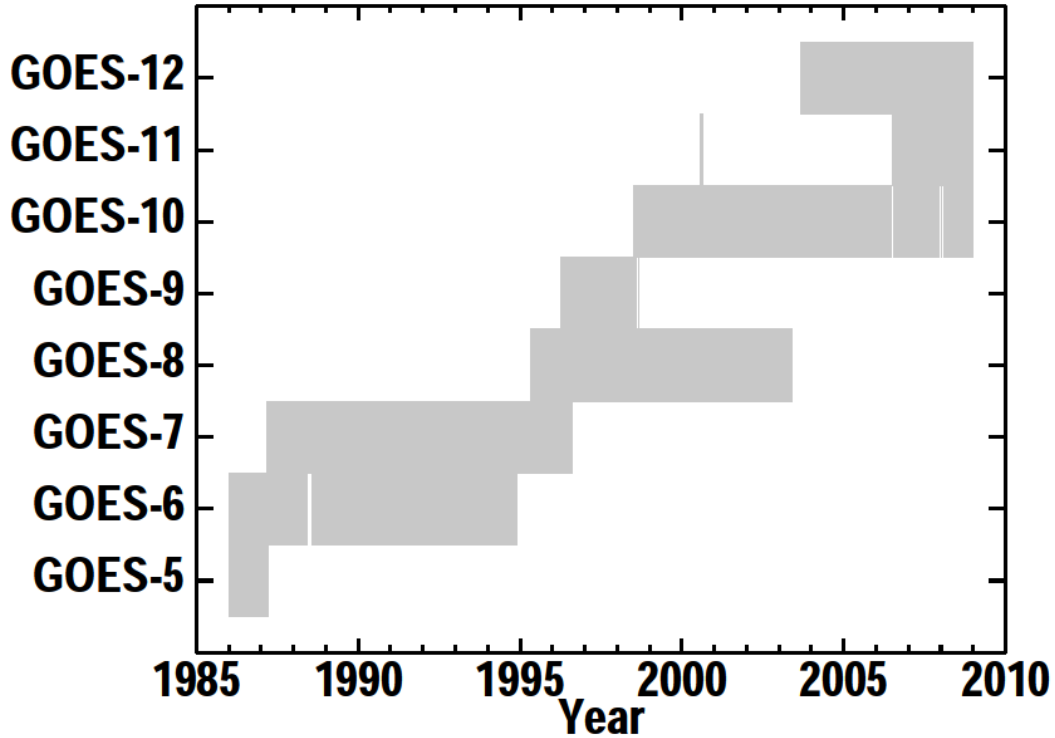
processing unit (SAU/DPU) [Space Systems-Loral, 1994]. From GOES-8 onward, the SEM instruments all showed a strong linear correlation with each other, as shown in

Figure 8.



**Figure 8.** Daily averages of the  $>2$  MeV electron flux for selected GOES satellite pairs. Excerpted from [Ling et al., 2010].

GOES-8 and GOES-7 are arguably correlated, but this relationship is much more ambiguous than in the later satellite packages; GOES-7 and GOES-6 show too large of a spread to determine a meaningful correlation. It is for this reason that only data from GOES-8 onward is used, with the objective being to have at least a solar cycle's worth of data, or  $\sim 11$  years. The reasons for this time interval selection will be detailed in Section 2.5. The availability of all GOES data from GOES-5 onward is shown in Figure 9.



**Figure 9.** The times for which  $>2$  MeV electron flux data is available for each GOES satellite. Excerpted from [Ling *et al.*, 2010].

The  $>2$  MeV electron flux data used in this study is from 1 July 1995 to 31 December 2008, coincident with GOES-8 onward. A desired aspect of the data is that the variation of small values should not be less significant than the variation of large values when considered locally, so the logarithm of the electron flux is used for all computations. The  $>2$  MeV logarithmic electron flux data was combined from each operational satellite at the time to get an average for geosynchronous orbit rather than the value at one particular orbit location. This was accomplished by normalizing all satellite data to GOES-8 through normalization parameters developed by Ling *et al.* [2010] and reproduced below:

**Table 1.** GOES >2 MeV electron flux normalization parameters. Excerpted from [Ling *et al.*, 2010].

$y^a$	$x^b$	$m^c$	$b^d$	$\sigma^{2e}$
GOES-8	GOES-9	0.935	-0.224	0.0353
GOES-8	GOES-10	0.963	-0.350	0.0646
GOES-8	GOES-11	0.943	0.048	0.00613
GOES-8 <sup>f</sup>	GOES-12	1.03	-0.260	
GOES-10 <sup>f</sup>	GOES-12	1.07	0.0937	0.0476
GOES-8 <sup>g</sup>	GOES-12	1.05	-0.720	
GOES-10 <sup>g</sup>	GOES-12	1.09	-0.384	0.0693

<sup>a</sup>The satellite corresponding to the ordinate data

<sup>b</sup>The satellite corresponding to the abscissa data

<sup>c</sup>The slope resulting from the fit of  $\log(y) = m \log(x) + b$

<sup>d</sup>The y-intercept of the fit

<sup>e</sup>The average squared deviation of the log of the data from the log of the normalized data.

<sup>f</sup>Normalization from 1 May 1995 to 30 June 2006.

<sup>g</sup>Normalization from 1 July 2006 to 31 December 2008.

The  $\sigma^2$  value is defined by the author as the average squared deviation of the log of the data from the log of the normalized data. Note that since GOES-12 was not operational at the same time as GOES-8, the GOES-12 data is first normalized to GOES-10 and the result is then normalized to GOES-8. Once normalized, the mean of the logarithmic flux values is then used as the electron flux for that given time. GOES electron flux data is recorded in 5-minute intervals, so there are 288 of these mean logarithmic flux values for any given day. These were averaged to provide one mean logarithmic flux value for the day, and this is the electron flux that was input to each of the models. It is known that proton contamination can occur for >2 MeV electron flux data during times of high proton flux, so when the average >30 MeV proton flux of all relevant GOES satellites exceeded 50 pfu, this data was treated as missing.

Solar wind velocity data is gathered from the Advanced Composition Explorer (ACE), located at the L1 Lagrange point between the sun and earth. Lagrangian points are a set of five points between two orbital bodies where gravitational effects exactly cancel out such that an object in one of these points can remain stationary. This data is available real-time in minute increments to provide advance warning of geomagnetic storms.

These 1-minute values were daily-averaged to get a single solar wind velocity approximation for each particular day coincident with the electron flux dataset. Also, the 3-hour indices for the summed Kp data were summed for the day to get a total measure of daily fluctuation.

## 2.2 REF M Model

The Relativistic Electron Flux Model (REFM) uses a linear prediction filtering technique successfully described in [Baker *et al.*, 1990] to couple the solar wind velocity with the electron flux at geosynchronous orbit. This method assumes that a filter  $g(\tau)$  can describe the relationship between an input time series  $I(t)$  and an output time series  $O(t)$  as

$$O(t) = \int_0^{\infty} g(\tau)I(t - \tau)d\tau,$$

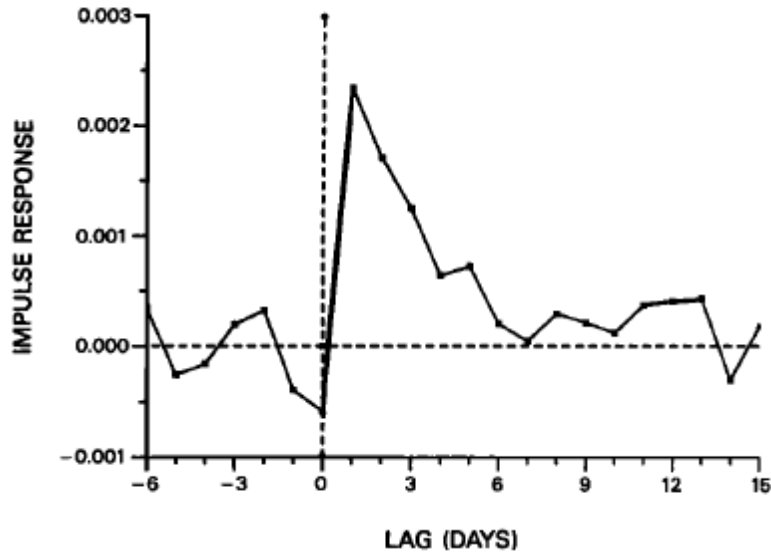
with  $t$  being the observation time and  $T$  being the time lag. The system is assumed to be linear and time stationary so that the output  $O(t)$  can be described by the input  $I(t)$  as it responds to the impulse response function  $g(\tau)$ . One can Fourier transform into the frequency domain to solve

$$G(\omega) = [O(\omega)]/[I(\omega)],$$

and then transform back into the time domain to get the corresponding  $g(\tau)$ . Essentially, the impulse response can be thought of as the array of elements  $a$  that is the solution of

$$R_{in} a = C_{IO},$$

where  $R_{in}$  is the input time series autocovariance matrix, and  $C_{IO}$  are the cross covariance elements between the input and output time series. The impulse response function for logarithmic electron flux E1 due to solar wind velocity  $V$  is shown in Figure 10.



**Figure 10.** The linear prediction filter between  $V$  and the log of the 3- to 5-MeV electron count rates at geosynchronous orbit. Excerpted from [Baker *et al.*, 1990].

As can be seen, the linear prediction filter exhibits a large negative value at zero lag and a large positive value at about 1 day, falling rapidly by day 3. This is the basis for the 1-, 2-, and 3-day forecasts provided by REFM. The coefficients for this linear predictive filter method along with past daily-averaged solar wind velocity are used in REFM to make a 72-hour prediction. Both coefficients  $c_i$  and daily-averaged solar wind velocity  $v_i$  are used for the last 30 days to get the raw forecast as in

$$rawfc = \sum_{i=1}^{30} c_i v_i$$

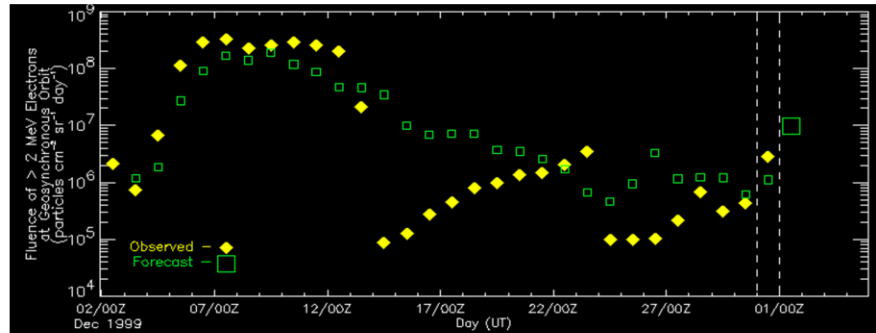
Once the raw forecast has been made, a ratio between the previous day's forecast and

observed value is used to modify the forecast such as

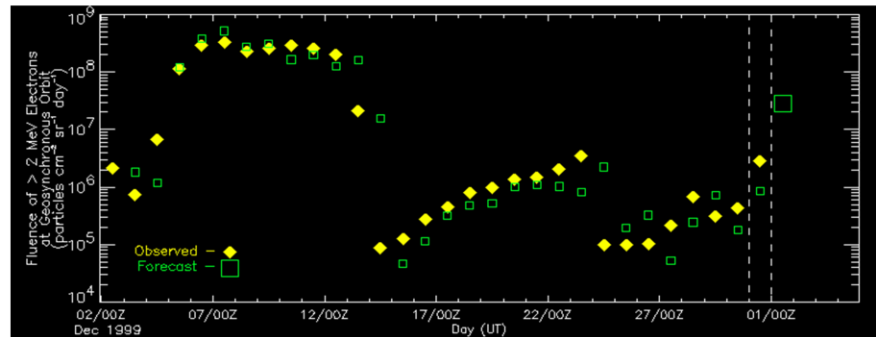
$$modfc_i = \frac{rawfc_i * obs_{i-1}}{rawfc_{i-1}}$$

This is done to catch any sudden electron events that the solar wind parameter did not characterize or to account for physical changes in the magnetosphere. The difference between including this correction factor and not can be seen in Figure 11.

The solar wind directly drives the electron fluence much of the time, and a forecast based on the linear prediction filter often works well. But as seen in the first example, the linear prediction alone cannot account for sudden changes in the magnetosphere's response.



To improve the forecast's performance statistics, we have included a correction factor. As shown, this allows the 1 Day forecast to quickly recover from sudden changes in the magnetosphere.

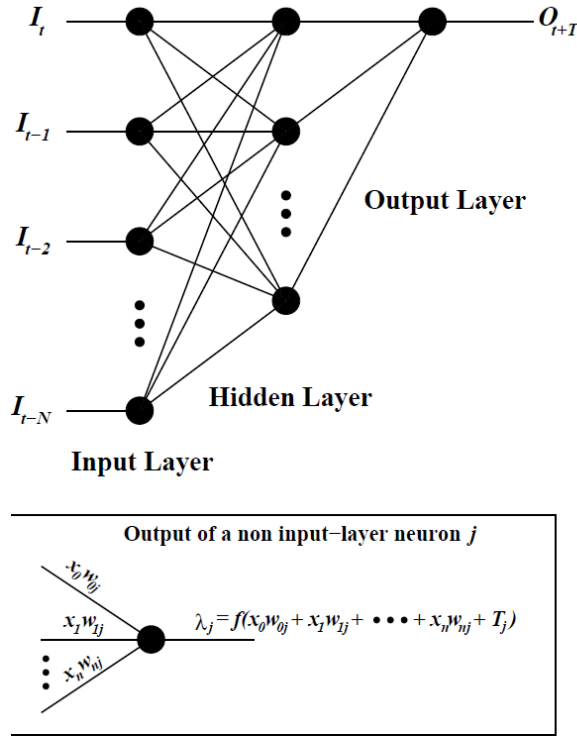


**Figure 11.** Demonstration of the flux correction factor in the REFM 1 day forecast model output. Excerpted from [Onsager et al.]

REFM deems electron flux data bad if the >30 MeV proton flux exceeded 50 pfu, the logarithm of electron fluence is greater than 10, the solar wind speed is greater than 1200 m/s or less than 200 m/s, or there is missing data. Since time dependence is integral to the analysis being performed, persistence was substituted for these values, or the last known daily-averaged electron flux, to get a coherent time series spanning the 11-year testing period.

## 2.3 Fluxpred Model

The Fluxpred model developed by *Ling et al.*[2010] is a multi-layer feed-forward neural network (MLFN) with one hidden layer, shown schematically in Figure 12.



**Figure 12.** Schematic of a multi-layer feed-forward neural network having 1 hidden layer. Neurons are shown as filled black circles and solid lines indicate the left to right flow of data. Ellipses indicate omitted neurons for clarity of presentation. Excerpted from [*Ling et al.* 2010].

The output  $O_{t+T}$  is produced from a time series of  $N$  inputs ( $I_{t-\tau}$ , where  $\tau$  is the time lag) after being fed through  $N$  input neurons and one hidden layer. With the exception of the input neurons, the output  $\lambda_j$  of each neuron  $j$  is given by

$$\lambda_j = f(\alpha_j),$$

where  $\alpha_j$  is the net input to a neuron  $j$  given by

$$\alpha_j = \sum_{i=0}^{N-1} x_i w_{ij} + T_j,$$



and  $x_i$  is the  $i$ th input to neuron  $j$ ,  $w_{ij}$  is the weight of input  $i$  to neuron  $j$ , and  $T_j$  is a threshold for neuron  $j$ . This net input  $\alpha_j$  is evaluated in an activation function

$$f_j(\alpha_j) = \frac{1}{1 + e^{-\alpha_j}}$$

Each  $w_{ij}$  and  $T_j$  are determined through network training. This is accomplished by first initializing the weights through a genetic optimization algorithm [Whitley, 1994]. Each member of the initial population is a binary string containing a set of weights which are randomly generated. The members are then evaluated by using their weights on each of the corresponding variables and calculating the error associated with the answer; in this case the observed electron flux value. Each string is then assigned a “fitness” value based on this error, with the desire that smaller values of error will map to larger values of fitness. This mapping function is defined as

$$f(v) = e^{kv},$$

where  $v$  is the error and  $k$  is chosen to be -20, which maps an error of 0.001 to a fitness value of 0.98, 0.01 to 0.82, and so on and so forth. The fitness for each string is normalized with respect to the average fitness of the population, i.e  $f_i/f_{ave}$ . The selection process of the strings from the current population to the intermediate population is done by a process called remainder stochastic sampling. For each string where the normalized fitness is greater than one, the integer portion determines how many copies of that string are placed directly into the intermediate population. The fractional portion of this ratio is then used to determine the probability that an additional copy of the string will be placed in the intermediate population. For example, a normalized fitness value of 1.81 places one copy of the string into the intermediate population and then gets a 0.81 chance of placing a second copy. For any normalized fitness value below one, the corresponding

string has this value's chance of being inserted into the intermediate population. This is done by initially filling the population with the integer strings till all strings only have fractional values, and then one of these is randomly chosen and its fractional value is compared to a random number between 0 and 1. If its fractional fitness value is greater than this random probability, it is chosen and inserted into the intermediate population after removal from the fractional fitness population. After the intermediate population is produced, recombination is used to create the next population. Two parents are randomly chosen from the intermediate population to reproduce with a probability of 0.80. A random probability is generated like above and compared to see if 0.80 is greater than it. If so, then a crossover point is randomly determined and the parents reproduce. For example, consider a binary string 10010110 and a binary string xxyxyyyx (where x and y are 1 and 0 but represented like this for clarity). Say the parents pass the probability test and the crossover point turns out to be in the middle of these, so that these parents would produce offspring of 1001yyyx and xxyx0110. After this recombination occurs, a 0.0001 probability of mutation can occur, in which a random bit in the child is flipped. This process of evaluation, selection, recombination and mutation constitutes one generation of the genetic algorithm. Fluxpred does this for three generations to initialize weights before the training begins, with a population size of 50 for each generation. After initialization, the weights  $w_{ij}$  and thresholds  $T_j$  were optimized through minimizing the mean squared error E,

$$E = \frac{1}{N_{TS}} \sum_{n=0}^{N_{TS}-1} E_n,$$

with  $N_{TS}$  being the number of training samples and  $E_n$  the error for a single training sample,

$$E_n = (f_j' - f_j)^2,$$

where  $f_j$  is the actual (model) activation function of the output neuron  $j$  and  $f_j'$  is the correct (data) activation of the output neuron. A conjugate gradient method was used to minimize  $E$ , with the partial derivative of  $E$  being with respect to each network weight  $w_{ij}$  as in

$$\frac{\partial E}{\partial w_{ij}} = -\lambda_j \frac{\partial f_j}{\partial \alpha_j} (\lambda_j' - \lambda_j),$$

where  $\lambda_j'$  is the desired output of neuron  $j$  and  $\lambda_j$  is the actual output of neuron  $j$ . A cumulative set of partial derivatives over the training set results in a weight gradient vector. Steps in  $w_{ij}$ -space are then taken in the direction of the vector until  $E$  is minimized.

For a MLFN comprised of  $N_L$  layers, the number of parameters needed to define the network through weights and thresholds is given by

$$N_{WT} = \sum_{i=0}^{N_L-1} (N_i + 1) N_{i+1},$$

with  $N_i$  being the number of neurons in layer  $i$ . The number of degrees of freedom this network has is then defined by

$$v = N_{TS} - N_{WT},$$

leading to the requirement that  $v > 0$  for a unique solution of minimizing  $E$  to exist. This condition was applied to an earlier neural network model for electron flux at geosynchronous orbit developed by *Koons and Gorney* [1991] (KG), which was the starting point for Fluxpred. The KG model had 10 input neurons, 6 hidden-layer neurons, and one output neuron, resulting in a network with 73 weights. The training period for the KG model was 62 days of data, of which 10 consecutive days were of summed Kp index for input to the network. This implied a training sample size of 52, meaning the

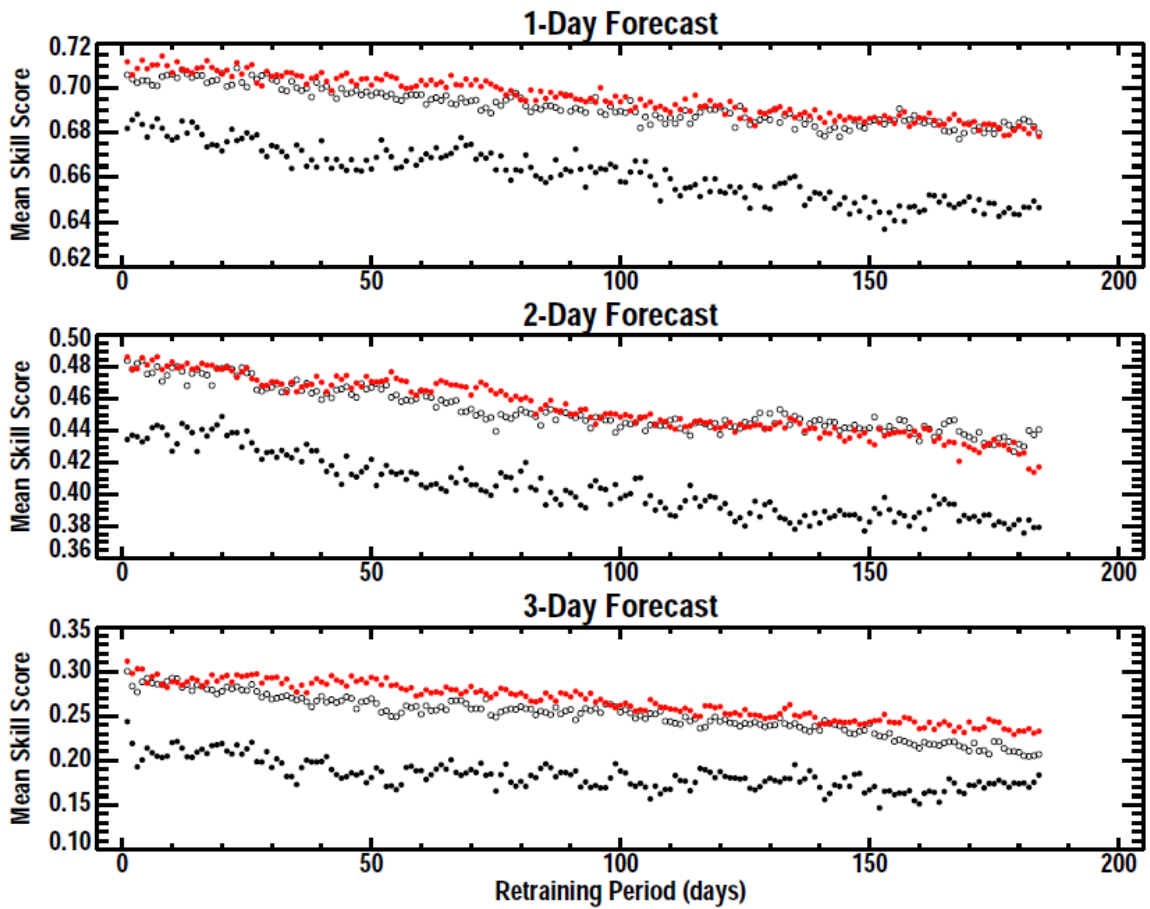
number of free parameters (weights and thresholds) exceeded the number of data points.

To decrease the possibility of over fitting in Fluxpred, all networks satisfied the condition

$$N_{TS} > 2N_{WT}.$$

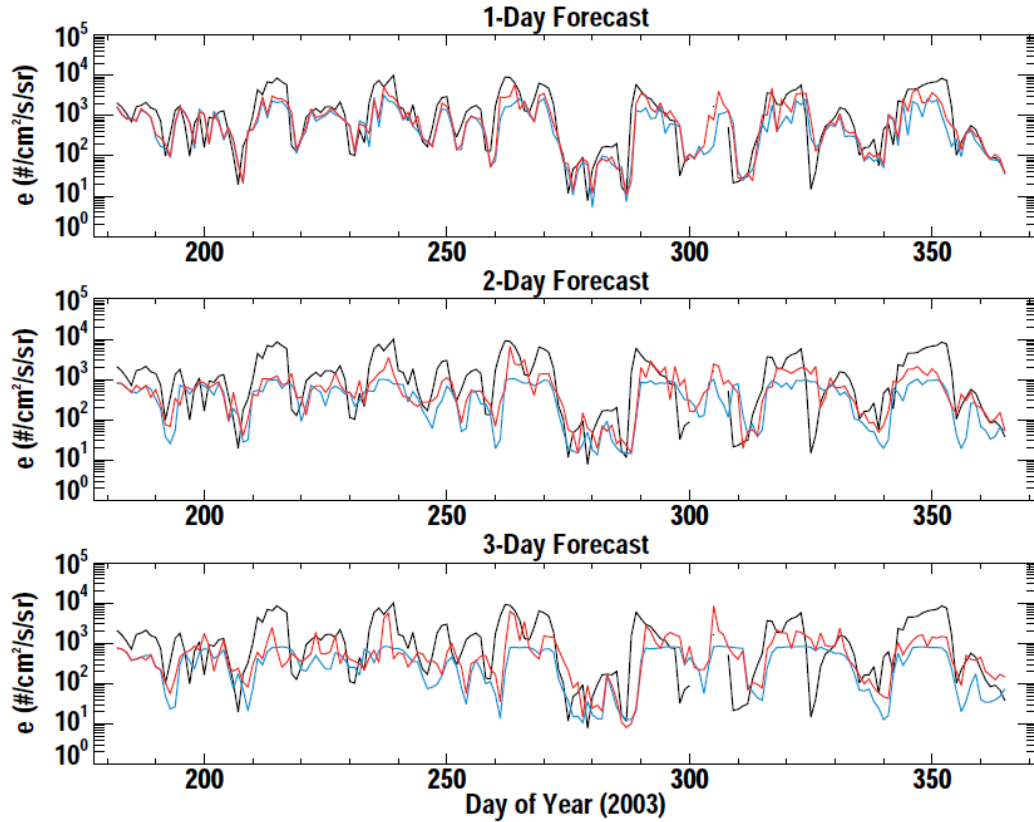
Each Fluxpred network was evaluated on its ability to reproduce the electron flux data used to train it and its ability to reproduce electron flux data not in the training set. It was found that summed Kp data by itself was inadequate to even now-cast electron flux, and certainly not to forecast it. The network was then modified to include current and historical electron flux along with summed Kp data, with an optimized network resulting in 10 electron flux input neurons corresponding to the last 10 days of daily-averaged electron flux, 7 summed Kp index input neurons corresponding to the last 7 days of daily summed Kp, and 2 hidden layer neurons.

Once the optimal network structure is found, the training period duration and the time between retraining need to be determined. Given the seasonal variations of the data [Cliver *et al.* 2000], it is important to choose a training set size of at least one year to minimize any data asymmetry. Networks of 1, 1.5, and 2 years of training data are examined with a 6-month retraining period, and the results are shown in Figure 13.



**Figure 13.** The mean skill scores for 1-day (top), 2-day (middle), and 3-day (bottom) electron flux forecasting as a function of the retraining period for 6-month test intervals over the period 1998-2008. The black filled circles denote a 1-year training set size, the black open circles an 18-month training set size, and the red filled circles a 2-year training set size. Excerpted from [Ling *et al.* 2010].

As can be seen, the 2-year training set size is the best on average. Retraining every 6 months versus every day was investigated, with the results shown in Figure 14.



**Figure 14.** Electron Flux as a function of time for 1-day (top), 2-day (middle), and 3-day (bottom) forecasts for the test interval 1 July 2003-31 December 2003. The black lines indicate the actual data, and the blue (red) lines indicate a retraining period ending 6 months (1 day) prior to the test period. The training set size for all networks is 2 years. Excerpted from [Ling *et al.*, 2010].

The data indicates daily retraining is superior to every 6 months, so the model is optimized with a 2 year training period to predict 1, 2, and 3 days into the future, with daily retraining.

## 2.4 Li Model

A model developed by Xinlin Li [Li, 2004] uses solar wind data to model the radial diffusion rate and in turn predict MeV electron fluxes at geosynchronous orbit. The model uses the standard radial diffusion equation proposed by [Schulz and Lanzerotti, 1974] and shown below:

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

In this equation,  $D_{LL}$  is the diffusion coefficient,  $\tau$  is the average electron life time, and  $f$  is the electron phase space density. This is related to differential flux  $j$  by

$$f = \frac{j}{p^2}$$

where  $p$  is the relativistic electron momentum.  $L$  is an adiabatic invariant related to the third, but by approximating the earth's magnetic field as a dipole,  $L$  is also the distance in earth radii at the equator. Both  $D_{LL}$  and  $\tau$  are functions of  $L$ , given by

$$D_{LL} = D_0 (L/6.6)^{10}$$

$$\tau = \tau_0 (6.6/L)^9$$

Average electron lifetime  $\tau_0$  is found to be 2.86 days for a 1.8-3.5 MeV electron, and  $D_0$  is a function of solar wind and interplanetary magnetic field (IMF) given by

$$D_0 = C(v/v_0)^{\gamma_1} [1 + ((v_x b_z + |v_x b_z|)/\alpha)^2]^{\gamma_2} \left[ \left( \frac{\Delta v^2}{\Delta t} \right)^2 / \beta \right]^{\gamma_3}$$

In the first term,  $v$  is the solar wind velocity and  $v_0$  is the nominal value of solar wind velocity (425 km/s). The second term is a function of the solar wind electric field y-component that varies only when  $b_z$  (the IMF z-component in GSM coordinates) is negative, since  $v_x$  is always negative. The third term deals with solar wind velocity fluctuations window-averaged over 1.5 hours for 10 minute data resolution, and  $\beta$  is the average of the fluctuations over 1995-1996, having a value of 9000 km<sup>4</sup>/min<sup>6</sup>. The constants were optimized by author for the period of 1995-1999 to minimize  $\chi^2$ , given by

$$\chi^2 = \frac{1}{N} \sum_i^N [\log_{10}(j_i) - \log_{10}(j'_i)]^2$$

where  $j_i$  is the modeled result and  $j'_i$  is the true (data) result. The parameters for 1.8-3.5 MeV electrons are  $C = 0.021/\text{day}$ ,  $\alpha = 68.0 \text{ km/s nT}$ ,  $\gamma_1 = 1.98$ ,  $\gamma_2 = 0.160$ , and  $\gamma_3 = 0.032$ . The inner boundary for the diffusion process is set at  $L = 4.5$  and the outer boundary is set at  $L = 11$ , chosen for its association with the last closed drift orbit. The radial diffusion equation is solved by setting  $(f_{outer} / f_{inner}) = 10^4$ , resulting in a radial gradient and thus diffusion. This value was chosen due to its approximate agreement with data [Li *et al.*, 1997]. The time for electrons to diffuse inward to geosynchronous orbit from  $L = 11$  is then found to be 1-2 days, in close agreement with observations [Li, 2004]. On day 0, the model steps through a day of solar wind data, solving the diffusion equation at each time step of 10 minutes. This results in unit-less electron flux values between  $L = 4.5$  and  $L = 11$ . The results at  $L = 6.6$  (geosynchronous orbit) are daily-averaged to get a unit-less electron flux value. Measured electron flux value from GOES is then averaged over the same time period and the model finds a factor to scale the result with as

$$0.0016 + scale * model = measured$$

Forecasting is done by averaging the diffusion coefficient over the last 2 hours and then using this value to run the diffusion equation for the next 2 days. The results are daily-averaged and scaled by the aforementioned factor, resulting in 1- and 2-day forecasts.

## 2.5 Ensemble Model

Owing to the 2 year training period that was deemed optimal for Fluxpred's neural network, the data used for testing the ensemble now begins at 1 July 1997, with the end



as previously stated at 15 August 2008. The total number of data points is 4064, which equates to 11.13 years. A solar cycle is ~11 years, during which the sun’s activity peaks to maximum and then diminishes to a minimum. Recurrence, or the value of electron flux 27 days into the past, is a much better estimate than persistence during solar minimum, where active regions on the sun can persist for multiple rotations. Persistence, however, is a much better estimate during solar maximum, where many active regions are coming and going that change the configuration. Also, CMEs may change the structures built up in the solar wind rather quickly, leaving the best “guess” at the electron flux as the last known value. It is for this reason that a data set of ~11 years is optimum for analysis, since the relevant physics of both persistence and recurrence are represented equally throughout the cycle, and one phenomenon is not biased over the other.

On any given day, a model will make predictions for the next three days (two for Li).

This is represented as:

**Table 2:** Prediction designations for individual models.

Model	Day +1	Day +2	Day +3
REFM	fcr1	fcr2	fcr3
FLUXPRED	fcf1	fcf2	fcf3
Li	fcl1	fcl2	

The goal of the ensemble model is to find a forecast of the form

$$f' = \alpha * fcr1 + \beta * fcf1 + \gamma * fcl1,$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are constants to be determined by a multivariate least squares regression over some time interval in the past. The multivariate least squares regression proceeds as follows. A fitting matrix (or regressor)  $B$  is formed, where for the 1-day forecast

$$B = \begin{bmatrix} fcr1_1 & fcf1_1 & fcl1_1 \\ fcr1_2 & fcf1_2 & fcl1_2 \\ \vdots & \vdots & \vdots \\ fcr1_n & fcf1_n & fcl1_n \end{bmatrix}.$$

A data vector  $d$  is also formed,

$$d = \begin{bmatrix} obs1 \\ obs2 \\ \vdots \\ obsn \end{bmatrix},$$

where the observed values are the actual daily-averaged electron flux values on the date the forecast was made for. The scale factors  $x$  are then found from

$$x = (B' B)^{-1} B' d,$$

where  $B' B$  is the covariance matrix. The result of this minimizes  $d - Bx$ , or minimizes the difference between the observed and forecast values for all vectors.

For analysis purposes, it is more convenient to line up the data according to date forecast for rather than date forecast on, as shown in Table 3:

**Table 3:** Data alignment for statistical calculations.

Day	REFM fc1	REFM fc2	REFM fc3	FPRED fc1	FPRED fc2	FPRED fc3	Li fc1	Li fc2	Li fc3
day+ 0									
day+ 1	fcr1_0			fcf1_0			fcl1_0		
day+ 2	fcr1_1	fcr2_0		fcf1_1	fcf2_0		fcl1_1	fcl2_0	
day+ 3	fcr1_2	fcr2_1	fcr3_0	fcf1_2	fcf2_1	fcf3_0	fcl1_2	fcl2_1	fcl3_0
day+ 4	fcr1_3	fcr2_2	fcr3_1	fccf1_3	fcf2_2	fcf3_1	fcl1_3	fcl2_2	fcl3_1

Where, for example, fcr1\_0 indicates an REFM forecast made on day 0 for 1 day into the future, fcf3\_1 indicates a Fluxpred forecast made on day 1 for 3 days into the future, etc.

Arranging the data in this manner then allows comparisons of accuracy between the

relevant 1-, 2-, and 3-day forecasts for any particular date. If day+0 is considered the starting day of the data set, then relevant statistics cannot be computed until day+3, when it is possible to check the 3-day forecast. Thus, the first 3 days of the dataset are not used.

Model performance is evaluated by two different methods. The first is the prediction efficiency, defined as

$$\text{Prediction Efficiency} = 1 - \frac{MSE}{\sigma^2},$$

where  $MSE$  is the mean squared error defined as

$$MSE = \frac{\sum_{i=1}^N (P_i - T_i)^2}{N-1},$$

where  $P_i$  is the log of the predicted value for the  $i$ th element of the time series,  $T_i$  is the log of the observed value for the same time element as given by the data, the sum is over the number of days,  $N$ , in the series, and  $\sigma^2$  is the variance of the data about the mean.

The prediction efficiency is then a measure of how much better the model does when compared to just forecasting the mean  $\mu$  for each day. A value of one would be considered perfect prediction efficiency, whereas a value of zero indicates that the model is just as good as the mean at forecasting the electron flux. Prediction efficiency can go negative, indicating that the mean would have been a better forecast than the result from the model. The second performance parameter is the relative error of the forecast, defined as

$$\text{Relative Error} = \frac{\sum_{i=1}^N |P_i - T_i|}{\sum_{i=1}^N |T_i|}$$

The absolute value of the numerator and denominator are taken separately because both can be both positive and negative. Daily-averaged flux is related to fluence through

$$\langle flux \rangle = \frac{fluence}{\#sec/day},$$

And daily-averaged logarithmic flux is related to logarithmic fluence through

$$\log_{10} \langle flux \rangle = \log_{10} \left( \frac{fluence}{86400} \right) = \log_{10} fluence - \log_{10} 86400.$$

Therefore, if the logarithmic fluence is less than  $\log_{10} 86400$ , the daily-averaged logarithmic flux will be negative. Absolute value is used because the primary concern is how far off of the true value the forecast was. Summations are used in both the numerator and denominator to smooth out local effects. If the relative error was defined by a single summation of each of the individual relative errors, then it could be artificially influenced by our chosen range, since zero is a valid value for daily-averaged logarithmic flux  $T_i$ .

The accuracy of our measurements should be discussed in the context of significant figures. GOES design and manufacturing is done by NASA, and once in orbit all satellite operations are conducted by the National Oceanic and Atmospheric Administration (NOAA). According to NOAA operators, GOES-1 to GOES-15 had no requirements for either absolute or relative accuracy, and thus the errors of the instruments are unknown. Starting with GOES-16, set to launch in 2015, there is a requirement for flux measurement accuracy to within 25%. This being the case, the significant digits will be carried out to the number given in the input data, which is three. Three significant digits was also the number recommended by NOAA operators when quoting values derived from GOES electron flux data.

# Chapter 3

## Results

### 3.1 Historical Ensemble Model

An analysis over the full 11-year period is first done for validation, and the variance is with respect to the mean logarithmic electron flux for the 11-year period, given by

$$\mu = \frac{\sum_{i=1}^N \log_{10} f_i}{N},$$

where  $f_i$  is the electron flux for a particular day  $i$ ,  $N=4064$  for the present 11-year analysis, and  $\mu$  is found to be 2.078. The prediction efficiency comparison with respect to  $\mu$  is shown in Table 4, and the relative error comparison is shown in Table 5:

**Table 4:** Prediction efficiency comparison for the 11-year dataset.

Day	REFM(PE)	FP(PE)	LI(PE)	ENSEMBLE(PE)
1	0.702	0.762	0.726	0.786
2	0.312	0.597	0.475	0.642
3	0.200	0.464	0	0.501

**Table 5:** Relative error comparison for 11-year dataset.

Day	REFM(ERROR)	FP(ERROR)	LI(ERROR)	ENSEMBLE(ERROR)
1	0.183	0.161	0.164	0.151
2	0.297	0.220	0.239	0.206
3	0.319	0.258	0	0.249

As can be seen, the ensemble model has both a higher prediction efficiency and a lower

### Chapter 3. Results

relative error than any one of the other three models, with FLUXPRED a close second. Since the Li model only makes predictions up to 2 days in advance, the 3 day ensemble may benefit from having a third regressor like the previous days. Recurrence is inserted on day 3 to fill this parameter, with the results shown in Tables 6 and 7:

**Table 6:** Prediction efficiency comparison for 11-year dataset with recurrence added on day 3.

Day	REFM(PE)	FP(PE)	LI(PE)	ENSEMBLE(PE)
1	0.702	0.762	0.726	0.786
2	0.312	0.597	0.475	0.642
3	0.200	0.464	-0.0623	0.523

**Table 7:** Relative error comparison for 11-year dataset with recurrence added on day 3.

Day	REFM(ERROR)	FP(ERROR)	LI(ERROR)	ENSEMBLE(ERROR)
1	0.183	0.161	0.164	0.151
2	0.297	0.220	0.239	0.206
3	0.319	0.258	0.363	0.245

The prediction efficiency is negative for recurrence, which suggests that the 11-year solar mean would be better at predicting the electron flux rather than recurrence on forecast day 3. This, however, is for analysis over the full 11-year period, and it will be shown in the following section why recurrence is a better parameter. The overall effects of adding recurrence on forecast day 3 are beneficial to the ensemble model, as its prediction efficiency increases by ~2.2% and its error goes down by ~0.4%.

It is of interest to examine the weights that the multivariate regression assigns to the individual models to see the effects of adding recurrence for forecast day 3, as this is a measure of how much the ensemble “trusts” each model. The non-normalized weights are shown both before and after the forecast day 3 recurrence addition in Tables 9 and 10.

**Table 8.** Weights for each of the models as determined by the multivariate regression.

Day	REFM	FP	LI
1	0.211	0.560	0.217
2	0.216	0.517	0.263
3	0.258	0.736	

**Table 9.** Weights for each of the models after recurrence is added on forecast day 3 as determined by the multivariate regression.

Day	REFM	FP	LI
1	0.211	0.560	0.217
2	0.216	0.517	0.263
3	0.209	0.627	0.165

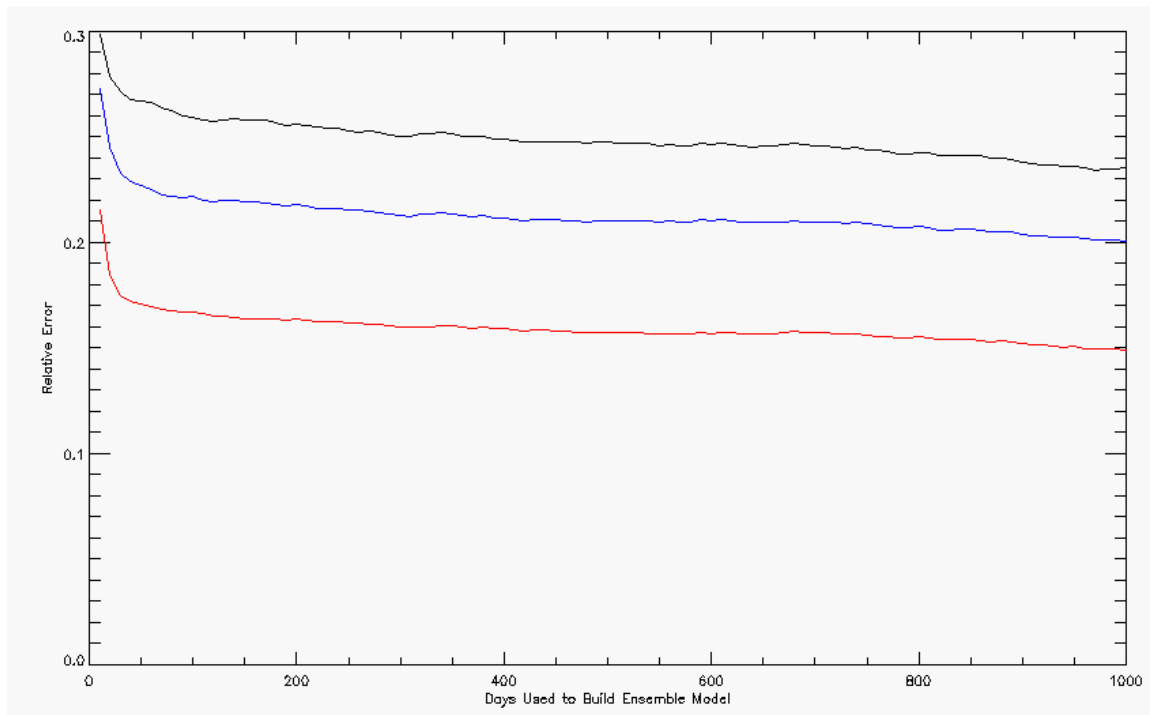
As expected, Fluxpred is weighted the heaviest for all three forecast days. The REFM and Li models are equally trusted on forecast day 1, but then the Li model becomes much more influential on forecast day 2. With the addition of recurrence for forecast day 3, Both REFM and Fluxpred lose some credibility from the ensemble’s perspective, and thus recurrence is added as a forecast day 3 model for the final ensemble form.

## 3.2 Predictive Ensemble Model

The previous section’s results were predictive in the sense that the 1-, 2-, and 3-day forecasts from each model were compared to the actual values, but the multivariate regression was done over 11 years, thus providing the best result possible for the dataset since the coefficients generated were optimized for the entire period. In reality, the forecaster will need to decide how many days of historical data to use to determine the regression coefficients, so an algorithm was developed to optimize this. It was experimentally determined that a model using the last 10 days of historical data to calculate its coefficients is the minimum length of time for a coherent prediction, with

### Chapter 3. Results

anything less prone to undesirable results. Because the data set being used to find the optimum time is limited to ~11 years, it was important to not choose too large of an upper limit for historical data in building the model, as a certain portion of the dominant physics present in either solar maximum or solar minimum will not be given an equal representation. For example, if 1000 days of historical data is used to build the model, then the model can only test itself on the remaining 3064 points in the time series, which may be unfairly skewed towards effects dominated by either solar minimum or solar maximum. The upper limit is initially chosen as 1000 days to characterize the relationship and ensure that there is not a global optimization minimum that is missed later on. A plot of error versus number of days to build the model is shown in Figure 15.



**Figure 15.** Relative error versus number of days used to build the ensemble, with day 1 (red), day 2 (blue), and day 3 (black) forecasts. Each relative error is an average of the error for 3000 models of  $x$ -day historical data.

This optimization uses a sliding model to compute an average more indicative of an ensemble model characterized by those days. For instance, the 10-day model will take



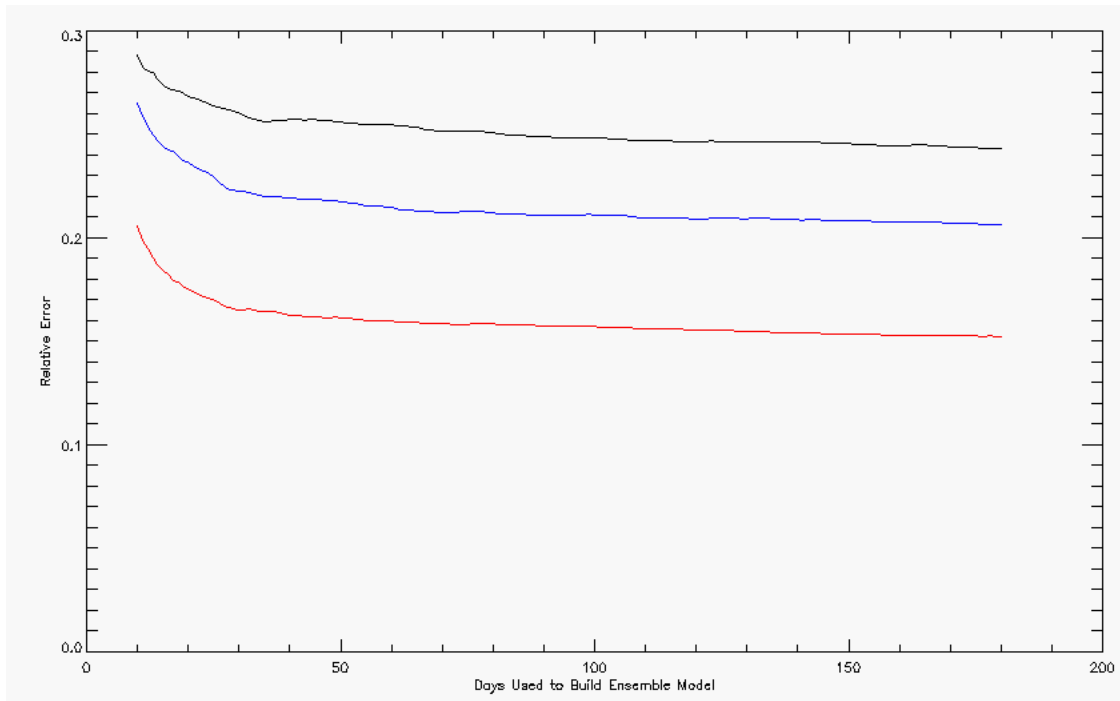
### *Chapter 3. Results*

the data from days 1-10, calculate coefficients, and forecast 1, 2 and 3 days out on day 11. It will then build another 10-day model from days 2-12 and forecast, then 3-13 and forecast, and so and so forth. These 1-, 2-, and 3-day predictions are then compared to the actual electron flux values to calculate relative error. Prediction efficiency is not used in this case since the data for the mean would only be built from the last 10 days of electron flux, so when a 1000-day model is built the average will be much closer to the real value. This then makes a comparison between the different ensembles impossible since the prediction efficiency is entirely relative to the electron flux mean.

Another parameter that is held constant between each of the ensemble models is the number of times it shifts. As previously discussed, a 1000-day ensemble model has 3064 days to shift through, whereas a 10-day ensemble model would have 4054 days available to shift through. This leads to another parameter that will vary between the different ensemble models, so the number of times each ensemble model shifts needs to be set constant regardless of how many days of historical data make up that model. This is determined by the largest model built, in this case 1000 days, so that 3000 shifts are used to characterize each model in Figure 15. This then means that for an  $x$ -day historical model the relative error shown is an average of 3000 models all comprised of  $x$  days of historical data, so that solar cycle variations do not influence the result (although they do in this case, for the unfair physics representation mentioned earlier). Unfair physics representation can be eliminated by choosing a smaller upper limit to build the ensemble models, so that ~11 years worth of data will be available to shift through. Figure 15 shows the majority of the error decrease occurs before 200 days, with the relative error gradually declining as the coefficients account for more of the 11-year data. The

### Chapter 3. Results

optimization is then recomputed for an upper limit of 180 days, with 3800 shifts allowed to characterize an  $x$ -day historical data model. The results are shown in Figure 16.

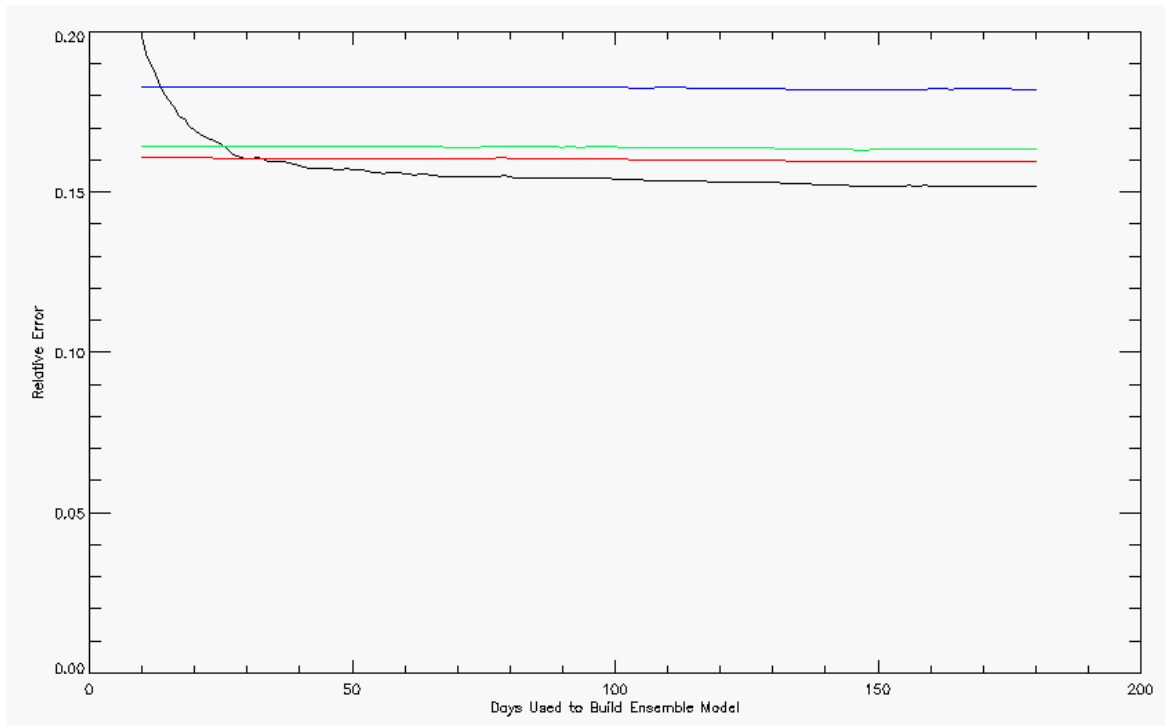


**Figure 16.** Relative error versus number of days used to build the ensemble, with day 1 (red), day 2 (blue), and day 3 (black) forecasts. Each relative error is an average of the error for 3800 models of  $x$ -day historical data.

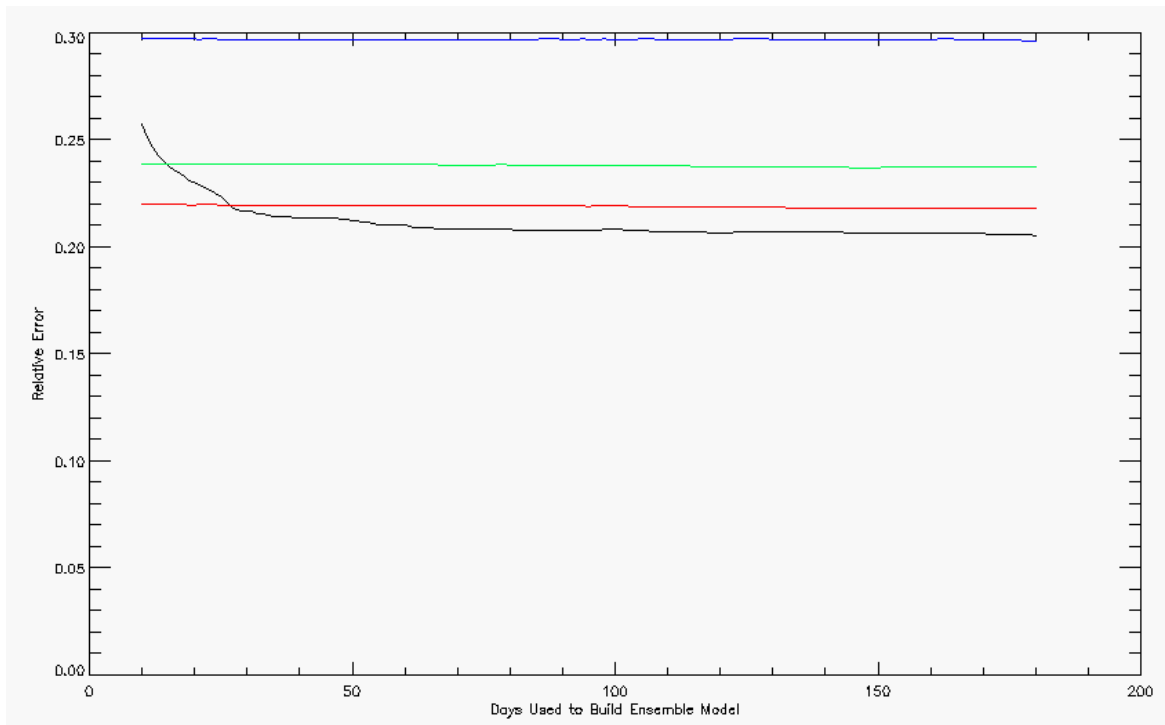
The error for forecast day 1 and forecast day 2 decreases rapidly leading up to a model comprised of 27 days of historical data, with forecast day 3 lagging slightly behind. This is as expected since the solar rotation is 27 days, so the majority of the error can be accounted for by having a historical model comprised at least of this, especially during solar minimum when recurrence dominates. Three solar rotations are chosen to include any additional error decrease, so the predictive ensemble model is built using 81 days of historical data.

As a comparison, the performance of each model in reference to the above analysis is shown in Figures 17, 18, and 19 for forecast days 1, 2, and 3, respectively.

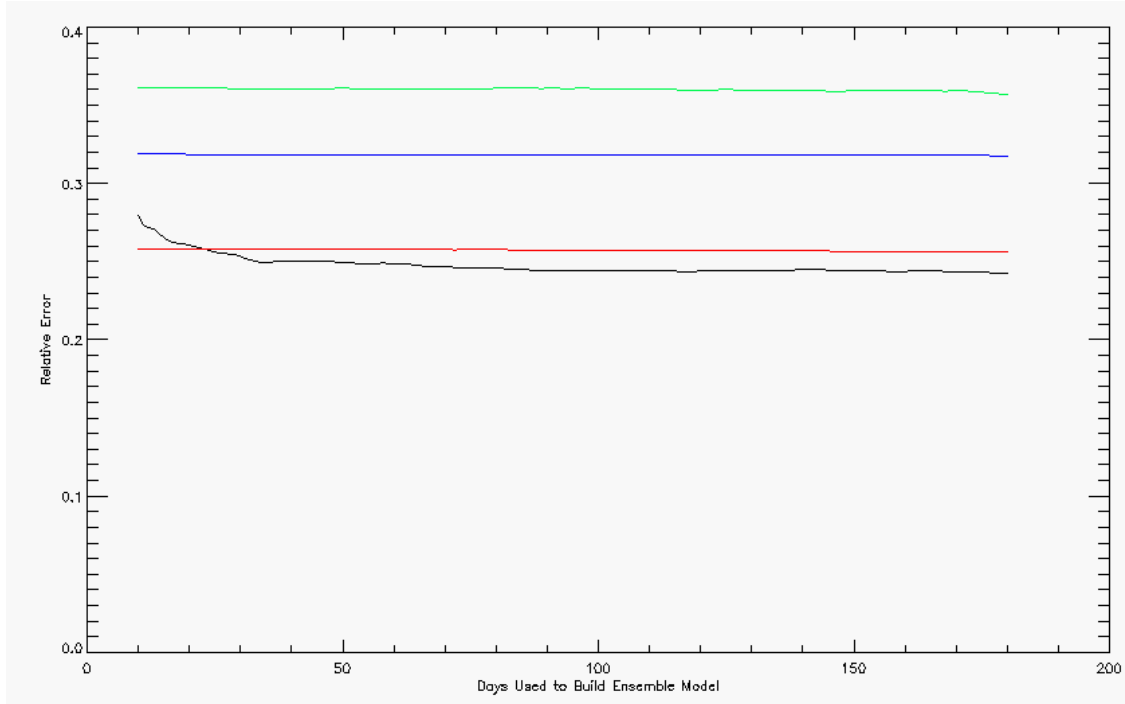
Chapter 3. Results



**Figure 17.** Relative error versus number of days used to build the model for forecast day 1, with ensemble (black), FLUXPRED (red), Li (green), and REFM (blue).



**Figure 18.** Relative error versus number of days used to build the model for forecast day 2, with ensemble (black), FLUXPRED (red), Li (green), and REFM (blue).



**Figure 19.** Relative error versus number of days used to build the model for forecast day 3, with ensemble (black), FLUXPRED (red), recurrence (green), and REFM (blue).

It should be noted that the functions for each of these models are approximately linear since the forecast for any particular model that isn't the ensemble is not dependent on the days used to build the model like the ensemble is. Essentially then, it is the characteristic error of the model. Fluxpred is the closest to the ensemble on all days, with error differences comparable to the values shown in Tables 6 and 7. A sliding average of the 81-day historical data ensemble model is done using 3800 shifts, with the results shown in Table 10:

**Table 10.** Relative error comparison for 81-day historical data ensemble model with recurrence added in Li's day 3 column.

Day	REFM(ERROR)	FP(ERROR)	LI(ERROR)	ENSEMBLE(ERROR)
1	0.183	0.161	0.164	0.155
2	0.297	0.219	0.238	0.208
3	0.318	0.258	0.361	0.246

### *Chapter 3. Results*

The results are similar to the relationship between the models shown earlier, with Fluxpred very close in accuracy to the ensemble. The ensemble has lower relative error by ~0.6%, ~1.1%, and ~1.2% for forecast days 1, 2, and 3, respectively. From the forecaster's perspective, this may not warrant the use of the ensemble model over Fluxpred due to the added complexity of running two additional models. However, there is a definite advantage to using the ensemble over Fluxpred in terms of a stochastic (probability-based) forecast, as shown in Section 3.3.

## **3.3 Stochastic Ensemble Model**

The Air Force Weather Agency (AFWA) mandates that if the 3-day fluence (flux\*time) either exceeds or is expected to exceed 1 billion particles that a warning should be issued for deep dielectric charging. A stochastic ensemble forecasting model was created by analyzing the error over the past 81-day historical dataset used to build the model and applying each of the 81 errors to the current forecast to get a distribution of forecasts. This was done for 1-, 2-, and 3-day forecasts, with a separate 81-day error set associated with each respected day. These three distributions are then used to form a probability of the 3-day fluence exceeding 1 billion particles. All permutations need to be considered for this probability, as any given forecast can assume any one of the errors it had in the last 81 days. This is done by adding the first error from the 1-day forecast error distribution to the current 1-day forecast value, adding the first error from the 2-day forecast error distribution to the current 2-day forecast value, and adding each of the 3-day forecast errors from the last 81 days to the current 3-day forecast value and checking whether each 3-day fluence would cause the fluence to exceed 1 billion particles. Next,

### Chapter 3. Results

the same first error from the 1-day forecast error distribution is kept for the current 1-day forecast value, the second error from the 2-day forecast error distribution is added to the current 2-day forecast value, and each of the errors from the 3-day forecast error distribution is again added to the current 3-day forecast value to check whether the fluence exceeds 1 billion particles. This continues on until all permutations are considered; thus, for an 81 day model,  $81^3 = 531441$  chances are evaluated that the fluence will exceed 1 billion particles over the next three days. The number of times the 3-day fluence exceeded the threshold is divided by the total number of chances (531441) it had to exceed the threshold to get the probability that the 3-day fluence will exceed 1 billion particles. These are calculated each day for the 11-year data set and compared to the days that actually exceeded the 1 billion particle count threshold. This comparison is done using a Brier Score (BS), which measures the accuracy of a set of probabilities, and is given by

$$BS = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2,$$

where  $f_i$  is the forecast probability and  $o_i$  is the actual outcome (either a 1 or 0) for exceeding the threshold [Brier, 1950]. This is done for the 1, 2, and 3-day forecast mode discussed above for both the ensemble and Fluxpred to compare their performance. For the 11-year period, the ensemble was found to have a BS of 0.0321, whereas Fluxpred was found to have a BS of 0.0327. Thus, the ensemble did ~2% better at predicting probabilities on average, since a lower score is better.

The 3-day fluence addition is done as a sliding sum, so on any given day using a 3-day forecasting model three separate fluence checks can be performed, as shown in Table 11.

**Table 11.** Fluence forecasting that could be performed on any given day.

Day	1st fluence	2nd fluence	3rd fluence	For day
0	fc1_0	fc2_0	fc3_0	3
1	fc1_1	fc2_1	fc3_1	4
	fc1_0	fc1_1	fc2_1	3
	observed_0	fc1_0	fc1_1	2
2	fc1_2	fc2_2	fc3_2	5
	fc1_1	fc1_2	fc2_2	4
	observed_1	fc1_1	fc1_2	3
3	fc1_3	fc2_3	fc3_3	6
	fc1_2	fc1_3	fc2_3	5
	observed_2	fc1_2	fc1_3	4

If day 0 was the first day that the model used, it could only make a 1-, 2-, and 3-day forecast, and thus only one 3-day total fluence forecast. On day 1, the same can be done as previously mentioned, but with two additional fluence forecasts. One can be made for 2 days out using the forecast made yesterday for the current day, the forecast made today for tomorrow, and the forecast made today for the day after tomorrow. Additionally, a 3-day fluence forecast can be made for tomorrow using the actual fluence from yesterday, the forecast made yesterday for the current day, and the forecast made for tomorrow. In essence, the steady state operation of the model allows 3-day fluence checks for 1, 2, and 3 days into the future.

It is clear that the 1-day forecast is the most important, since it is the closest to the time of incident. Also, it is used twice along with observed data to forecast tomorrow's 3-day fluence, as shown in Table 11. Consider then a stochastic model built entirely of 1-day forecasts, where

$$3 \text{ day fluence} = fc1_{i-2} + fc1_{i-1} + fc1_i,$$

where the subscript indicates the  $i$ th day forecasted on. The BS for this mode is then

### Chapter 3. Results

calculated for the 11-year period and the ensemble was found to have a BS of 0.0185, whereas Fluxpred was found to have a BS of 0.0210. Thus, the ensemble did ~12% better at predicting probabilities on average than its best member (Fluxpred). This analysis can be done for both a model built entirely of 2-day forecasts and a model built entirely of 3-day forecasts (although in actuality a forecaster would not use these models) to characterize how accurate the probability predictions for 2-day and 3-day forecasts are. The full results are shown in Table 12, with the current day chosen to be day 3 both as an example and in reference to the above Table 11. The first fluence member set is for 3 days into the future (day 6) as was first discussed, whereas the last three fluence members are models composed entirely of 1-, 2-, or 3-day forecasts to predict whether the fluence will exceed the threshold tomorrow (day 4). Keep in mind the current day being day 3 is only chosen for clarity to see the originating day for the forecasts, and the actual analysis is done over the full 11 years.

**Table 12.** Brier Score results for both the ensemble and FLUXPRED.

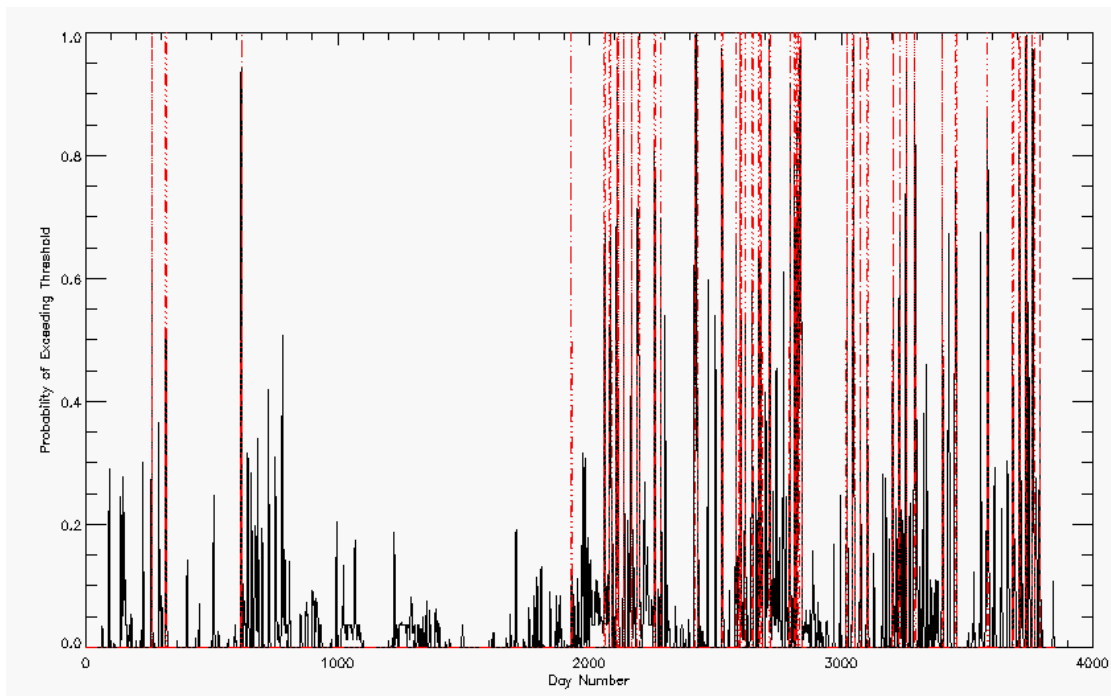
Fluence Members	Ensemble BS	FLUXPRED BS	Percent better
fc1_3,fc2_3,fc3_3	0.0321	0.0327	1.89
fc1_1,fc1_2,fc1_3	0.0185	0.0210	11.9
fc2_0,fc2_1,fc2_2	0.0272	0.0305	11.0
fc3_-1,fc3_0,fc3_1	0.0340	0.0341	0.154

For forecasting both day 1 and day 2 probabilities the stochastic ensemble did considerably better than FLUXPRED, but was barely better for day 3. This is probably due to the fact that recurrence is used as a model on day 3 since the Li model does not forecast this, and is a rather inaccurate model to use for an entire solar cycle. Thus, the set of physics encompassing radial diffusion is not represented in day 3, and probability information based on these physics is also lost.



### Chapter 3. Results

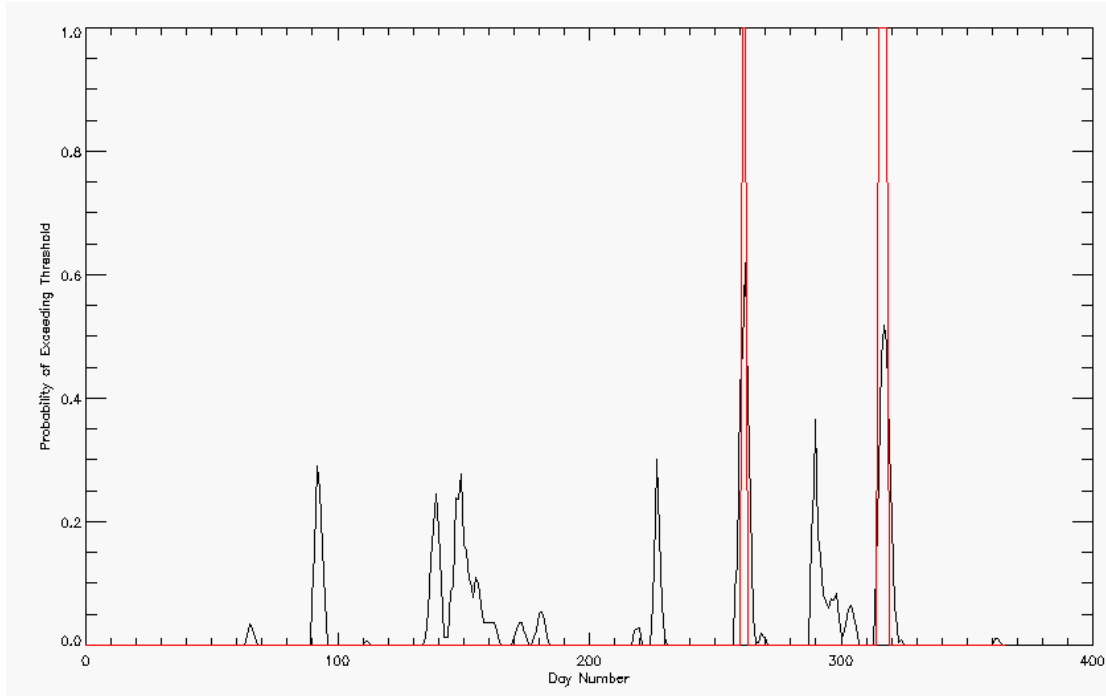
The performance of the stochastic 1-day forecast ensemble model should be investigated further as its probabilities are the most important for predicting whether or not the 3-day fluence will exceed 1 billion particles. This is done by comparing the probability plot to past occurrences where the 3-day fluence did exceed 1 billion particles and looking for a correlation. This occurred for 159 days in the 11-year period of study. Figure 20 shows the entire 11-year probability set, with a red vertical bar around the days that the fluence exceeded the threshold.



**Figure 20.** Probability of exceeding threshold versus day in the 11-year period, with a red bar around the days that threshold was in fact exceeded.

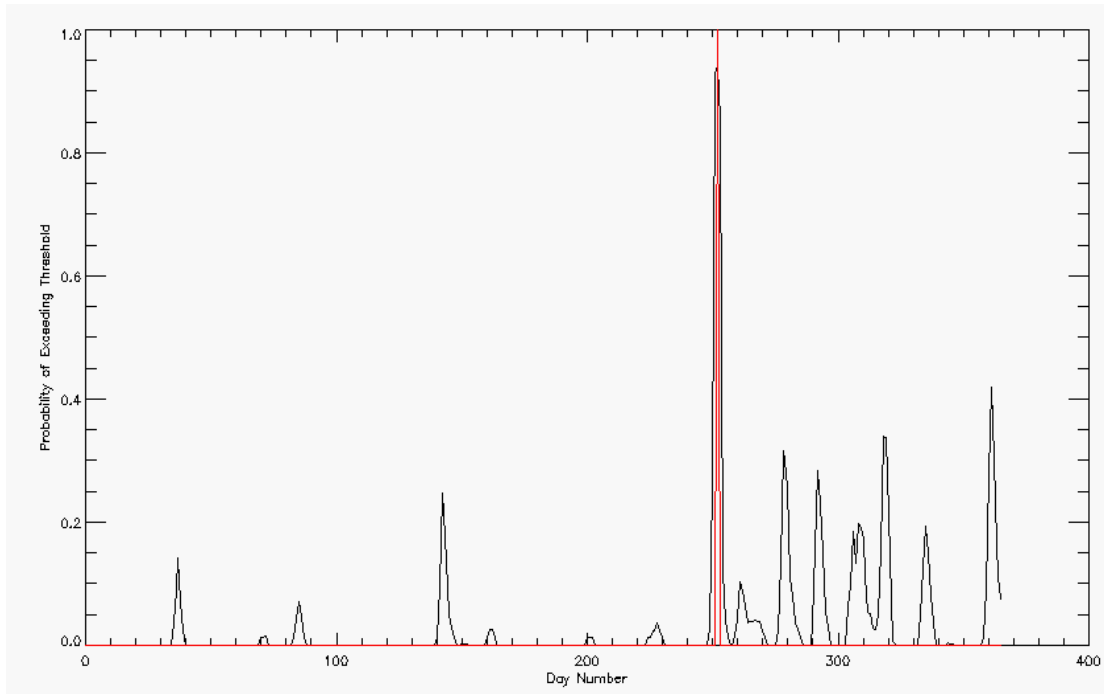
Since it is rather hard to see how well the probabilities are correlated to the events, the above is graphed on a year-by-year basis so that the results can be analyzed in further detail. The ensemble model dataset starts in July of 2007, but 81 days are needed to build the model, and 81 additional days of ensemble forecast are needed to make an error distribution, so the stochastic dataset begins the following January. Figure 21 shows the

first year of the dataset, which is January 1998 – January 1999.



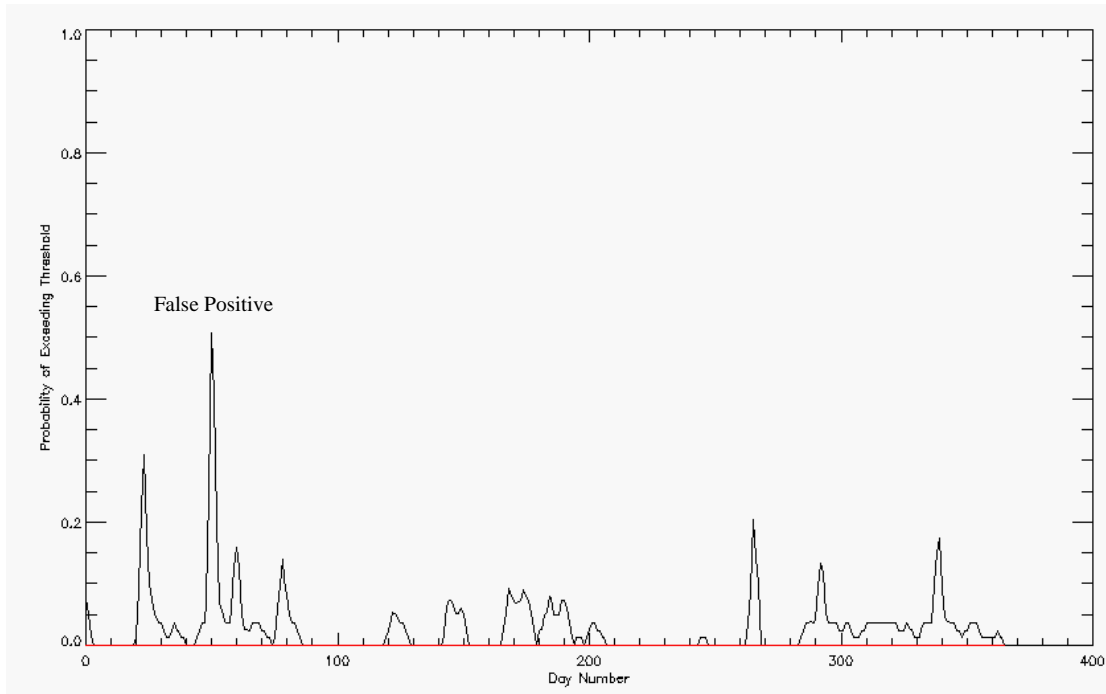
**Figure 21.** Probability of exceeding threshold versus day for January 1998-January 1999, with a red bar around the days that threshold was in fact exceeded.

The two highest probability peaks corresponded to the two actual events, but some sort of quantitative comparison should be done to evaluate false positives/false negatives and thus the performance of this model. Technically speaking, false positives and false negatives only apply to a deterministic model and do not have meaning in a stochastic model, but if a probability threshold is selected as reference then they can be evaluated accordingly. Choosing anything over 50% as a positive, and anything under 50% as a negative, it can be seen that Figure 21 contains neither false positives nor false negatives, as every instance where the probability was over 50% corresponded with an actual event. Year 2 of the 11-year period is shown in Figure 22.

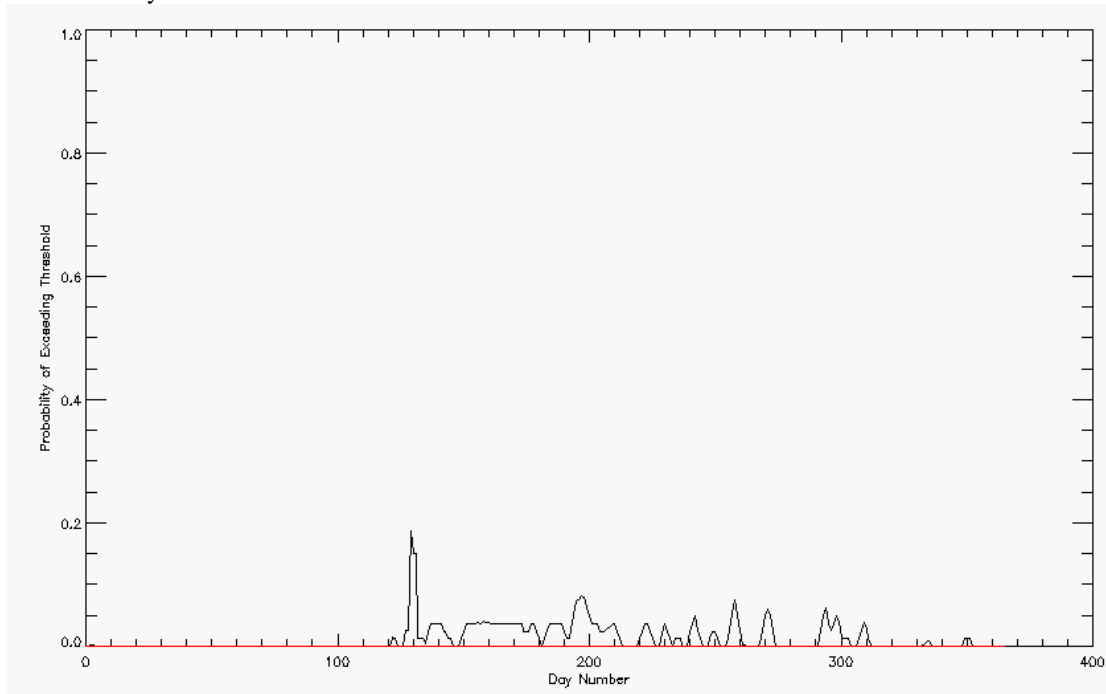


**Figure 22.** Probability of exceeding threshold versus day for January 1999-January 2000, with a red bar around the days that threshold was in fact exceeded.

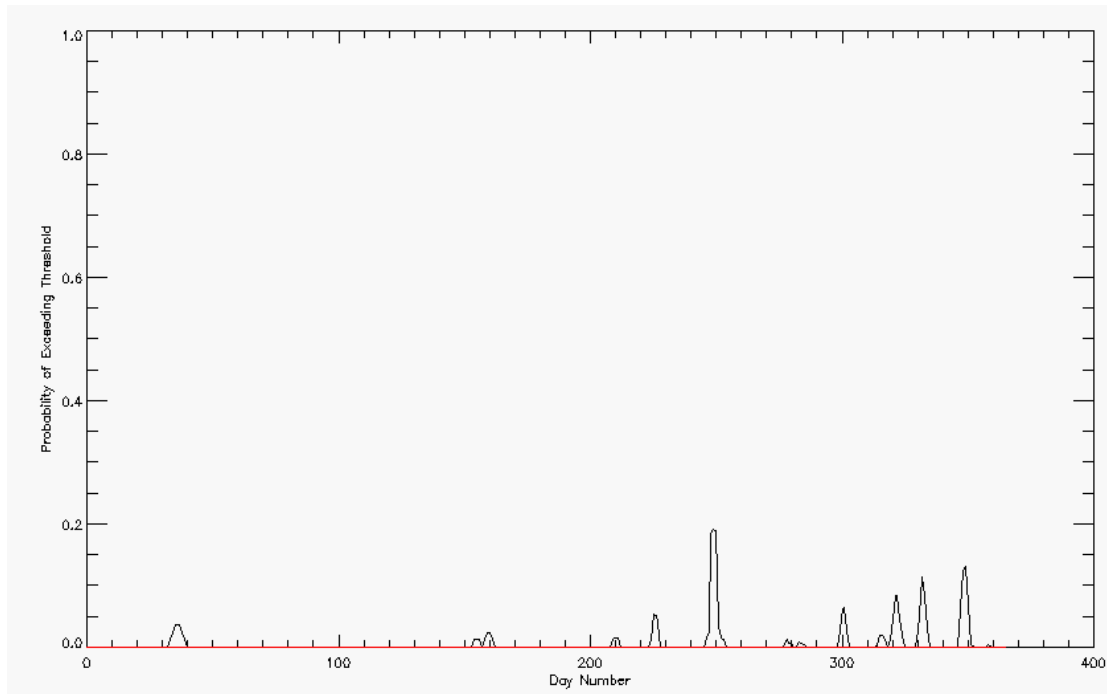
Again, there are no false positives or false negatives for the year. Figures 23, 24, and 25, corresponding to the period between January 2000 and January 2003, all had no events occurring but are shown below for instances of false positives.



**Figure 23.** Probability of exceeding threshold versus day for January 2000-January 2001, with a red bar around the days that threshold was in fact exceeded.

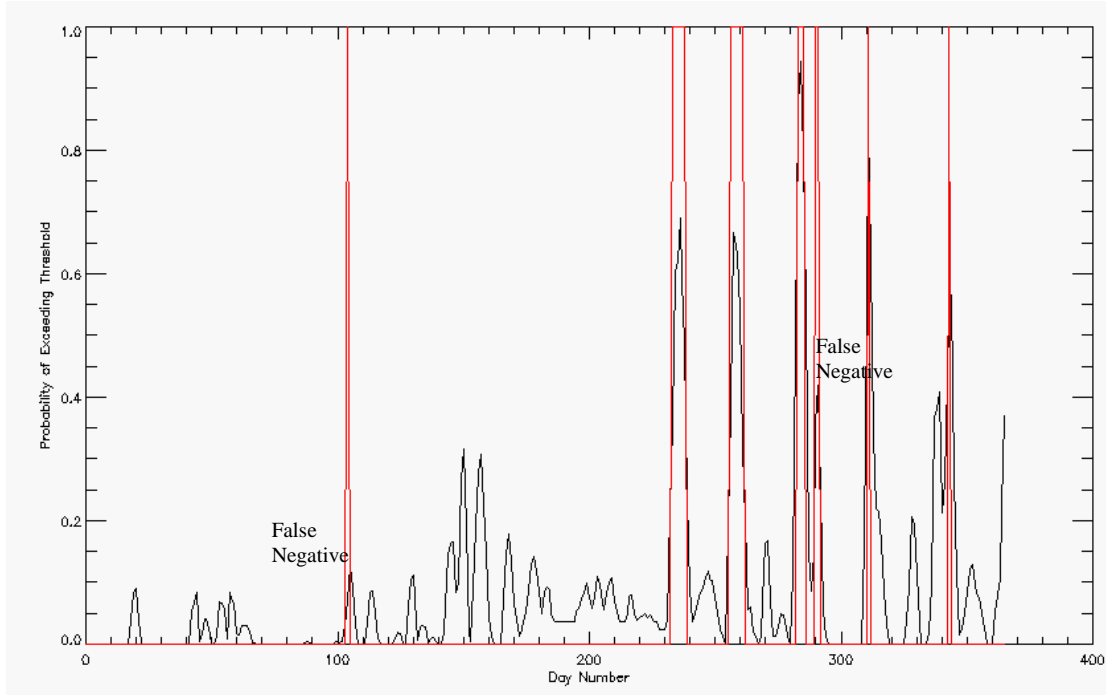


**Figure 24.** Probability of exceeding threshold versus day for January 2001-January 2002, with a red bar around the days that threshold was in fact exceeded.



**Figure 25.** Probability of exceeding threshold versus day for January 2002-January 2003, with a red bar around the days that threshold was in fact exceeded.

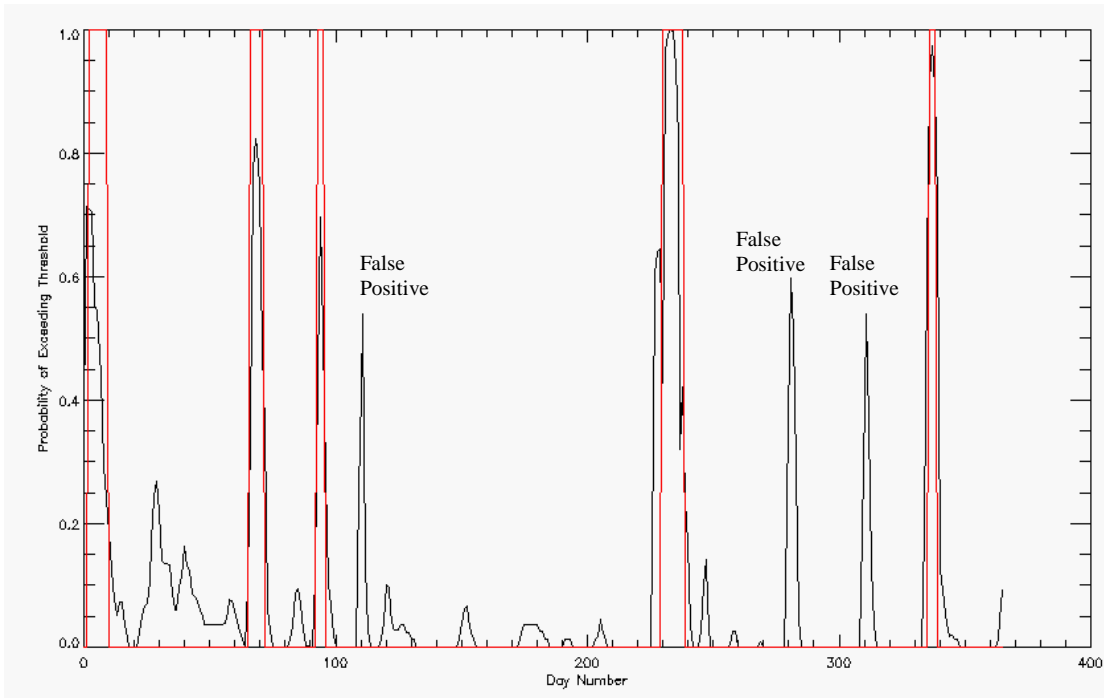
Figure 23 shows one false positive, whereas figures 24 and 25 contain none. Electron events start picking up as solar maximum declines to solar minimum, as is seen for Figure 26.



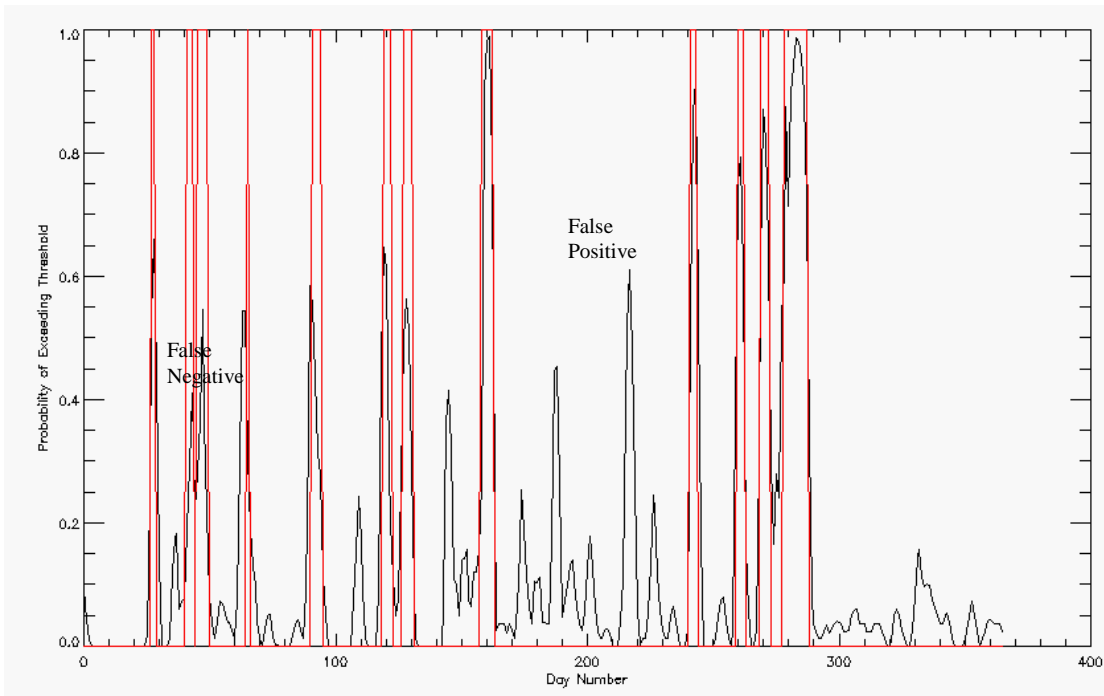
**Figure 26.** Probability of exceeding threshold versus day for January 2003-January 2004, with a red bar around the days that threshold was in fact exceeded.

The first instance of a false negative can be seen in Figure 26, occurring near day 100. A second one occurs farther on a few days after a correctly predicted event. There are no false positives for this year, as every probability over 50% corresponded to an event.

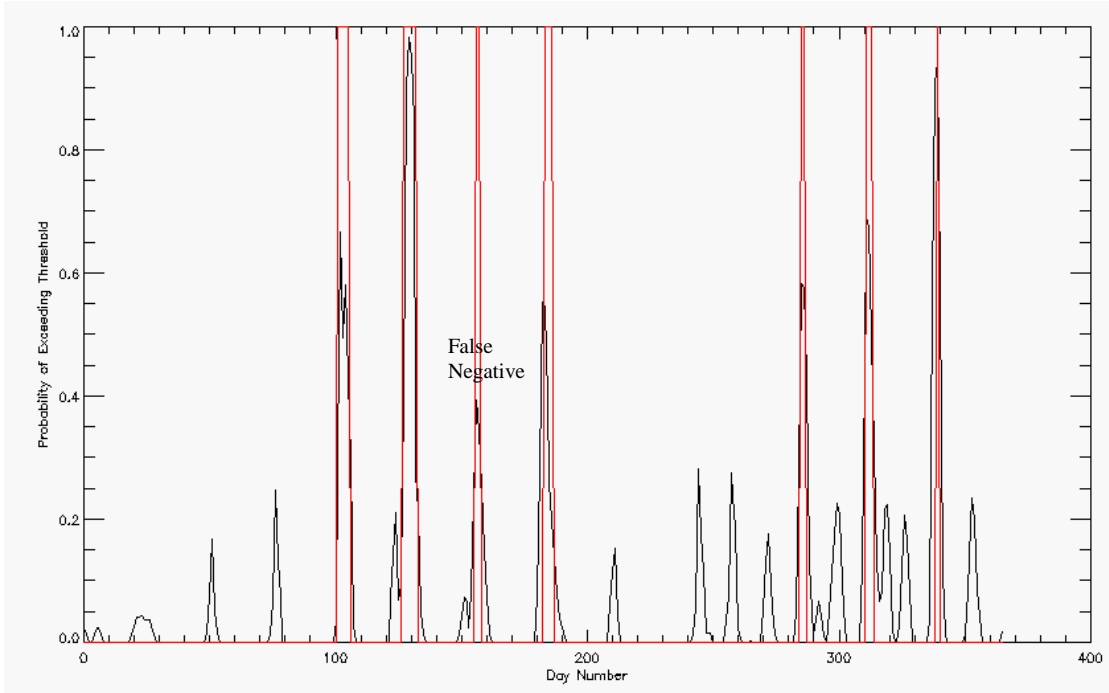
Figures 27-31 show the probabilities for the remaining years.



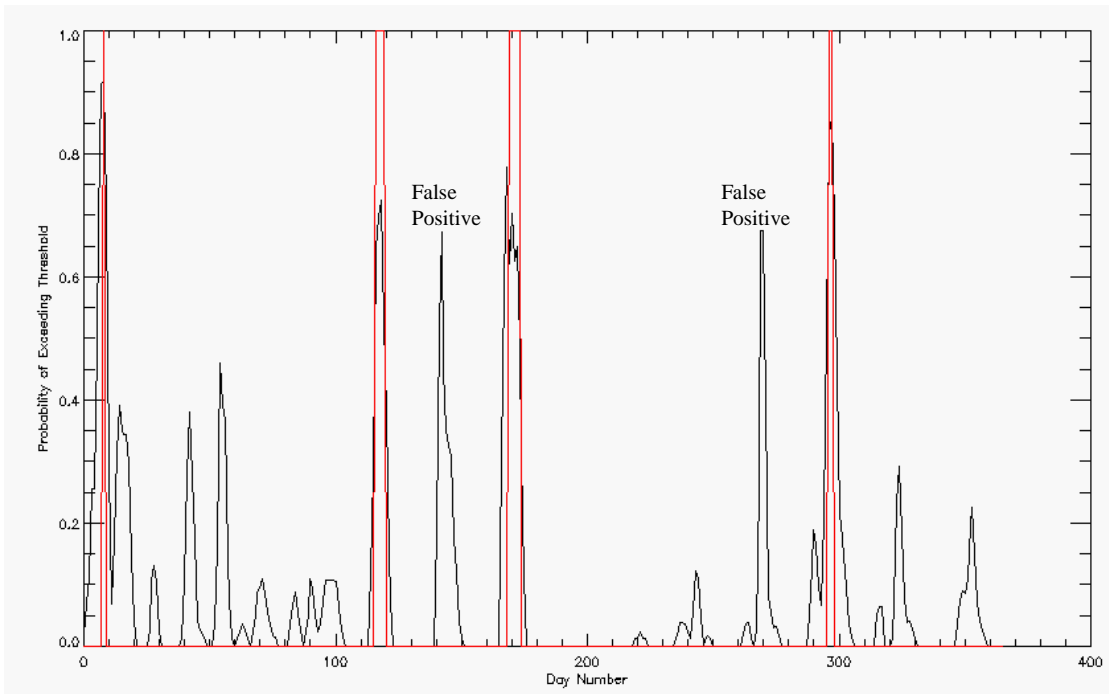
**Figure 27.** Probability of exceeding threshold versus day for January 2004-January 2005, with a red bar around the days that threshold was in fact exceeded.



**Figure 28.** Probability of exceeding threshold versus day for January 2005-January 2006, with a red bar around the days that threshold was in fact exceeded.

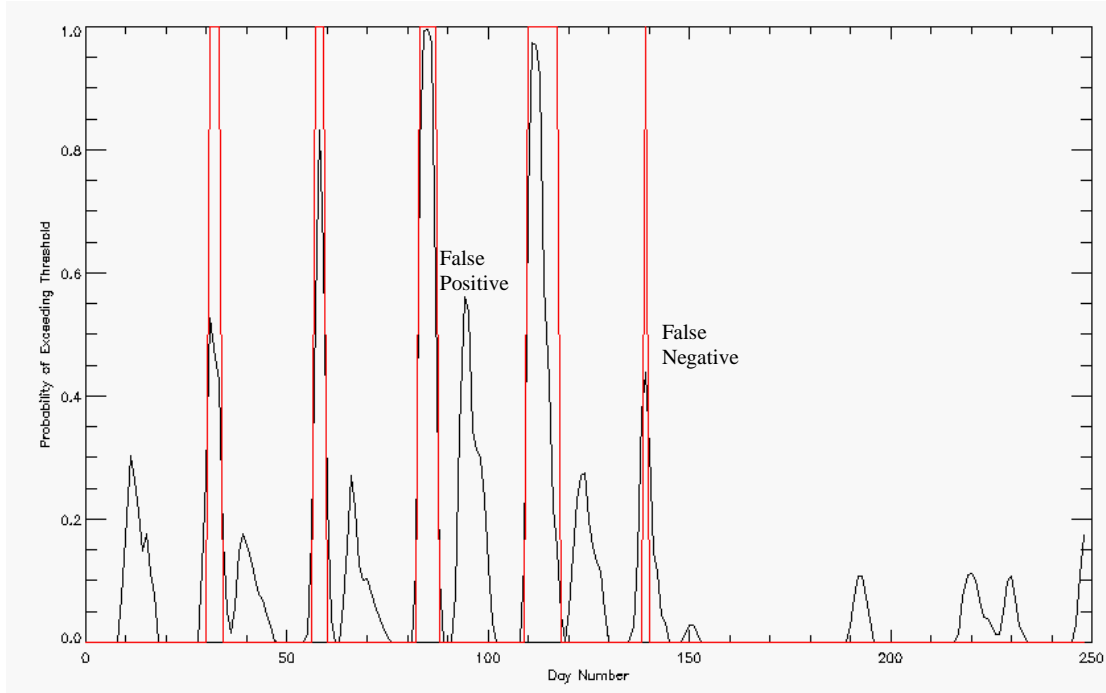


**Figure 29.** Probability of exceeding threshold versus day for January 2006-January 2007, with a red bar around the days that threshold was in fact exceeded.



**Figure 30.** Probability of exceeding threshold versus day for January 2007-January 2008, with a red bar around the days that threshold was in fact exceeded.





**Figure 31.** Probability of exceeding threshold versus day for January 2008-August 2008, with a red bar around the days that threshold was in fact exceeded.

It is clear that the false negative cases are the most important parameter, as a forecaster would not want to miss an event. The first false negative occurred for one day, the second for two days, the third for three days, the fourth for two days, and the fifth for one day, giving a total of ten days of false negatives out of 159 days of occurrences. 149 out of 159 days correct yields a probability of 93.7% that the stochastic ensemble will predict an event 1 day in advance, given that prediction of an event corresponds to a probability over 50%.

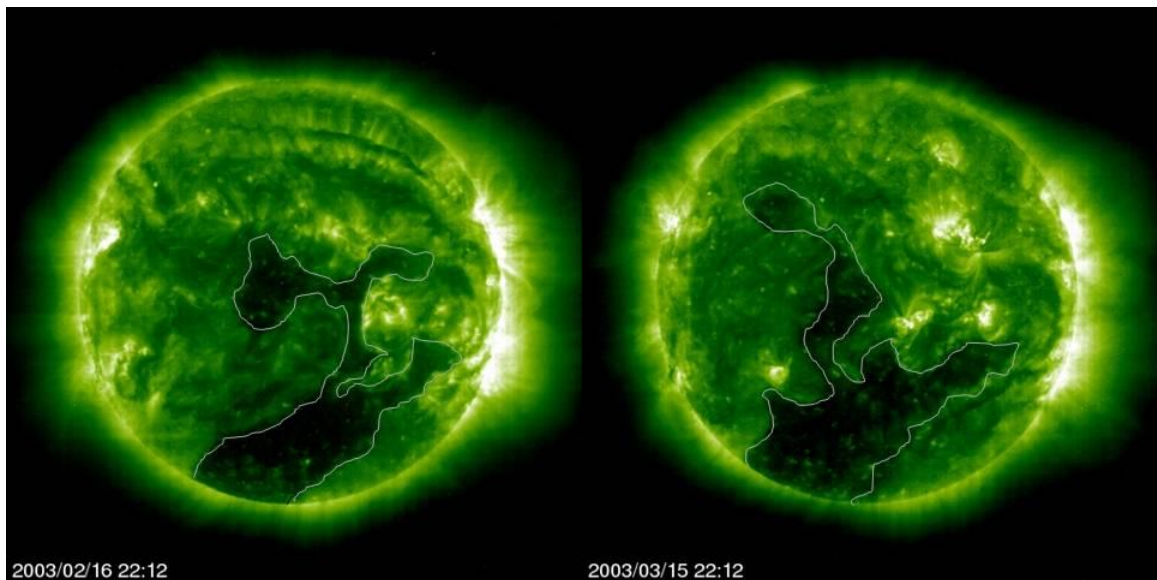
With the exception of the first false negative (shown slightly after day 100 in Figure 26), all of the false negatives had a probability of at least 40%. It is then reasonable to say that if the forecaster defined false positives and false negatives differently, i.e. with respect to a 40% probability, then 158 out of 159 days would be correctly predicted 1 day in advance, corresponding to a probability of 99.4%. This, however, will also increase

### Chapter 3. Results

the number of false positives issued, and analysis of the above images shows four additional ones. This is certainly acceptable from a forecaster's perspective and well worth the trade-off in correctly guessing four valid events.

The first false negative event bears closer scrutiny to determine why the stochastic model missed it. The 3-day fluence for this event exceeded the threshold on 22 March 2003.

Figure 32 shows NASA SOHO Extreme Ultraviolet Imaging Telescope (EIT) data for the time period just before the event and the time period 27 days prior.

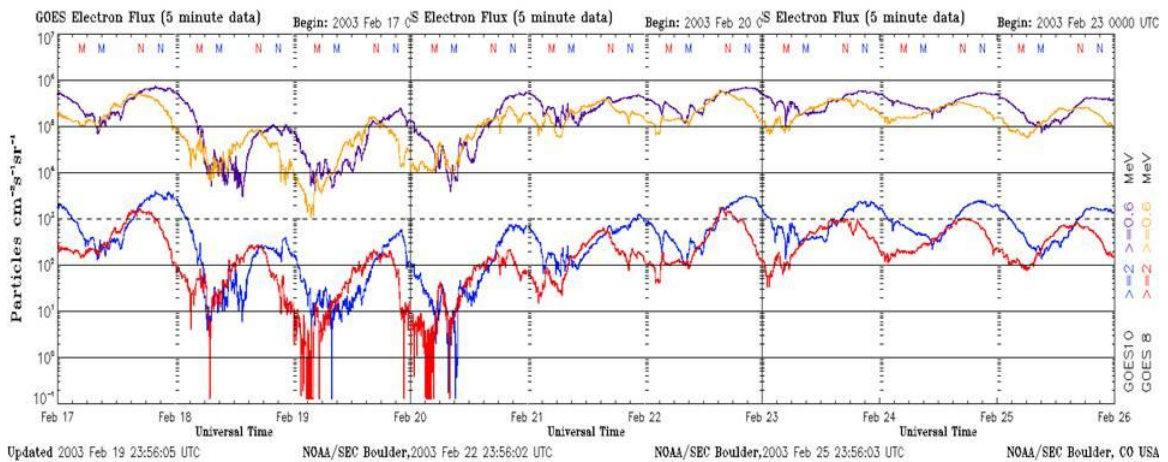


**Figure 32.** NASA SOHO EITS 195 images, courtesy of <http://sohowww.nascom.nasa.gov/data/>.

The first thing to note is the differences in the coronal holes between the two solar rotations. In February the coronal hole is fairly narrow overall, and only produced minimal geomagnetic storming at active levels (summed  $K_p = 4$ ), with  $>2$  MeV electron flux levels exceeding  $1e03$  pfu for seven days post storming. The coronal hole next rotation is wider and produces significant geomagnetic storming for six days beginning on 16 March, 2003. This storming reached major geomagnetic storming (summed  $K_p = 6$ ) with severe geomagnetic storming (summed  $K_p = 8$ ) at high latitudes for the first two

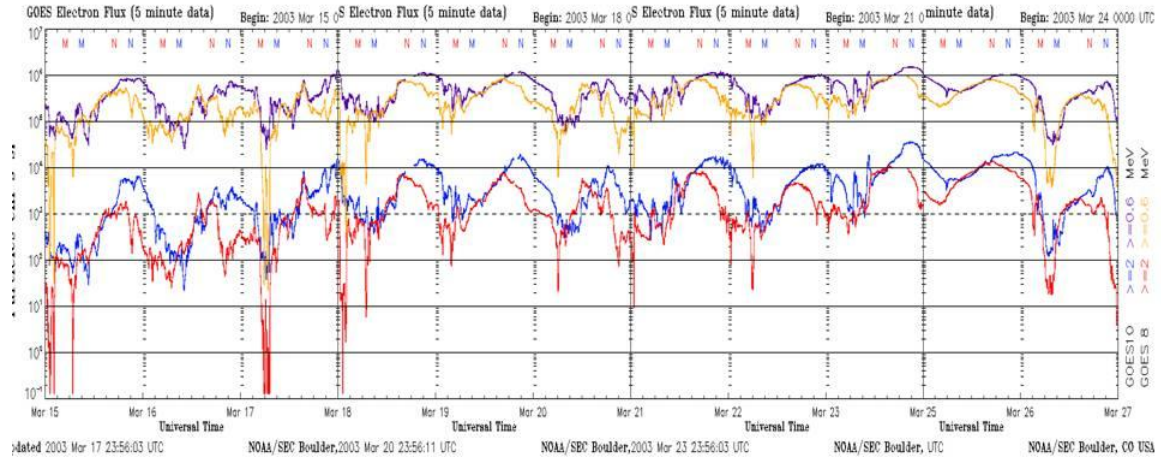
### Chapter 3. Results

days of the stream. Another contributing factor to the increased storming levels was a coronal mass ejection that arrived on 20 March, 2003, which further pushed already unsettled geomagnetic storming up to what is considered minor storming levels (summed  $K_p = 5$ ) for a day and a half. The electron flux during and after this storming period became elevated beginning on 16 March, 2003, and remained near or at  $1e4$  pfu, which sustained is enough to surpass the 1 billion particle fluence limit (i.e.,  $3*86400*1e4$  is greater than 1 billion). Electron flux data for both the event and one month prior are shown in Figures 33 and 34, respectively, with the red and blue lines corresponding to the  $>2$  MeV electron flux.



**Figure 33.** NOAA SWPC GOES electron flux data 1 month prior to the event, with the red and blue lines corresponding to  $>2$  MeV electron flux as measured from GOES 8 and GOES 10, respectively.

### Chapter 3. Results



**Figure 34.** NOAA SWPC GOES electron flux data for the event, with the red and blue lines corresponding to  $>2$  MeV electron flux as measured from GOES 8 and GOES 10, respectively.

The models cannot accurately account for CMEs, and the probability calculation is based on the last 81 days of error. This event matured mid rotation, and thus the model had very little information to be able to detect this event. There is a probability spike at this point which is due to the elevated background flux levels leading up to the event, and the combination of this plus the arrival of a CME two days prior was enough to push the electron count over the threshold. This situation is assumedly the exception to the rule, as it does not occur anywhere else in the 11-year period.

# Chapter 4

## Conclusions

As mankind becomes more dependent on satellites, energetic particles leading to satellite failure through radiation damage become increasingly important. Geosynchronous orbit has a radiation environment that is of primary concern due to the large number of commercial and military satellites operating there. Thus, forecasting energetic electron fluxes that can lead to satellite failure is a valuable tool in protecting space assets.

The ensemble model presented here, comprised of three separate models, is found to have both lower relative error and significantly better prediction efficiency than any one of its individual members. Additionally, a stochastic version of this model is presented to forecast the probability that an electron event threshold will be crossed. Out of 159 days that the threshold was in fact exceeded during the 11-year period of study, the stochastic model missed only one of these (issuing a false negative in relation to a 40% probability threshold). This missed event was correlated with elevated background flux levels and the arrival of a CME that contributed enough to push the 3-day electron fluence over the threshold.

Ensemble forecasting is an important tool to have in a system whose physics are not yet fully-known. Ensemble forecasting is in its infancy in the space weather field, as very

#### *Chapter 4. Conclusions*

little research has been done in this respect, so the true potential has yet to be explored. Additionally, the majority of space weather forecasting models are deterministic, so a stochastic ensemble model is ground-breaking research in this field. There are currently no known ensemble models, either stochastic or deterministic, in operational space weather forecasting; this is the first of its kind.

# References

- Baker, D. N., R. L. McPherron, T. E. Cayton, and R. W. Klebesadel (1990), Linear prediction filter analysis of relativistic electron properties at  $6.6 R_E$ , *J. Geophys. Res.*, *95*, 15,133.
- Brier, G. W. (1950), Verification of Forecasts Expressed in Terms of Probability, *Monthly Weather Review*, *78*: 1-3.
- Crown, M., and B. Davis (2000), Forecasting the Space Environment, presented as official course curricula at the Air Force Research Laboratory.
- Engebretson, M., et al. (1998), The dependence of high-latitude Pc5 wave power on solar wind velocity and on the phase of high-speed solar wind streams, *J. Geophys. Res.*, *103*, 26, 271.
- Kivelson, M. G., and C. T. Russell (1995), *Introduction to Space Physics*, Cambridge University Press, New York, New York.
- Koons, H. C., and D. J. Gorney (1991), A Neural Network Model of the Relativistic Electron Flux at Geosynchronous Orbit, *J. Geophys. Res.*, *96*, A4, 5549.
- Li, X. (2004), Variations of 0.7-6.0 MeV electrons at geosynchronous orbit as a function of solar wind, *Space Weather* *2*, S03006, doi:10.1029/2003SW000017.
- Li, X., D. N. Baker, M. Temerin, D. Larson, R. P. Lin, G. D. Reeves, J. B. Blake, M. Looper, R. Selesnick, and R. A. Mewaldt (1997), Source of relativistic electrons in the magnetosphere: Present knowledge and remaining questions, presented at the 8<sup>th</sup> Scientific Assembly of IAGA, Int. Assoc. of Geomagn. And Aeron., Uppsala, Sweden.
- Ling, A. G., G. P. Ginet, R. V. Hilmer, and K. L. Perry (2010), A Neural Network Based Geosynchronous Relativistic Electron Flux Forecasting Model, *Space Weather*, submitted.
- Nagai, T. (1988), "Space Weather Forecast": Prediction of Relativistic Electron Intensity at Synchronous Orbit, *Geophys. Res. Lett.* *15*, 425.
- Onsager, T., C. Smithtro, Relativistic Electron Flux Model.
- Schulz, M., and L. Lanzerotti (1974), *Particle Diffusion in the Radiation Belts*, Springer-Verlag, New York.
- Shea, M. A., and D. F. Smart (1998), Space Weather: The effects on operation in space, *Adv. Space res.* *22*, 1,29.
- Space Systems-Loral (1994), *GOES I-M Databook*, Space Systems-Loral, Palo Alto, California.
- Whitely, D. (1994), A Genetic Algorithm Tutorial, *Statistics and Computing* (4):65-85.