

RESEARCH ARTICLE

10.1002/2015SW001303

Key Points:

- Accuracy of the electron fluence forecast by REFM and SNB³GEO is assessed using GOES 13 data
- Prediction efficiency, correlation, and Heidke skill scores indicate higher accuracy of SNB³GEO
- The data correction methodology used by REFM may improve SNB³GEO accuracy even further

Supporting Information:

- Data Set S1

Correspondence to:

M. A. Balikhin,
m.balikhin@sheffield.ac.uk

Citation:

Balikhin, M. A., J. V. Rodriguez, R. J. Boynton, S. N. Walker, H. Aryan, D. G. Sibeck, and S. A. Billings (2016), Comparative analysis of NOAA REFM and SNB³GEO tools for the forecast of the fluxes of high-energy electrons at GEO, *Space Weather*, 14, 22–31, doi:10.1002/2015SW001303.

Received 4 SEP 2015

Accepted 28 DEC 2015

Accepted article online 11 JAN 2016

Published online 28 JAN 2016

©2016. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Comparative analysis of NOAA REFM and SNB³GEO tools for the forecast of the fluxes of high-energy electrons at GEO

M. A. Balikhin¹, J. V. Rodriguez^{2,3}, R. J. Boynton¹, S. N. Walker¹, H. Aryan⁴, D. G. Sibeck⁴, and S. A. Billings¹

¹ACSE, University of Sheffield, Sheffield, UK, ²Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado, USA, ³National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA, ⁴GSFC, NASA, Greenbelt, Maryland, USA

Abstract Reliable forecasts of relativistic electrons at geostationary orbit (GEO) are important for the mitigation of their hazardous effects on spacecraft at GEO. For a number of years the Space Weather Prediction Center at NOAA has provided advanced online forecasts of the fluence of electrons with energy >2 MeV at GEO using the Relativistic Electron Forecast Model (REFM). The REFM forecasts are based on real-time solar wind speed observations at L1. The high reliability of this forecasting tool serves as a benchmark for the assessment of other forecasting tools. Since 2012 the Sheffield SNB³GEO model has been operating online, providing a 24 h ahead forecast of the same fluxes. In addition to solar wind speed, the SNB³GEO forecasts use solar wind density and interplanetary magnetic field B_z observations at L1. The period of joint operation of both of these forecasts has been used to compare their accuracy. Daily averaged measurements of electron fluxes by GOES 13 have been used to estimate the prediction efficiency of both forecasting tools. To assess the reliability of both models to forecast infrequent events of very high fluxes, the Heidke skill score was employed. The results obtained indicate that SNB³GEO provides a more accurate 1 day ahead forecast when compared to REFM. It is shown that the correction methodology utilized by REFM potentially can improve the SNB³GEO forecast.

1. Introduction

The smooth operation of modern society depends upon the ability for the fast exchange of information between various organizations and individuals. Spacecraft operating at geostationary orbit (GEO) form an essential part of our modern technological infrastructure that enables this information exchange. Adverse space weather conditions may result in the interruption of services provided by spacecraft at GEO, thereby leading to a reduction, temporary cessation, or possibly termination of the communication/navigational services used in many aspects of modern society. In spite of the provision of some shielding on board the spacecraft, energetic electrons are still able to penetrate deep inside the spacecraft subsystems and components, potentially resulting in internally induced electrostatic discharges. Internally induced electrostatic discharges can cause anomalies in spacecraft operation and even the failure of vulnerable subsystems [Vampola, 1987; Baker *et al.*, 1987; Wrenn, 1995; Fennell *et al.*, 2000; Bodeau, 2010]. Thus, the reliable forecast of high-energy electron fluxes at GEO can assist in the mitigation of hazardous space weather effects on spacecraft operations there.

In the current paper, the quality of the forecasts of daily averaged fluxes of energetic electrons (>2 MeV) at GEO by the Relativistic Electron Forecast Model (REFM) and SNB³GEO tools is assessed. The forecast of >2 MeV electron fluxes at GEO from three models [REFM Baker *et al.*, 1990, Li, 2004 and FluxPred Ling, 2000; Ling *et al.*, 2010] have been compared by Perry *et al.* [2010] together with persistence and reoccurrence models. [Perry *et al.*, 2010] demonstrated that all of the models had clear solar cycle dependencies and concluded that on average the models perform very similarly. As it was stated in Perry *et al.* [2010] "After evaluating all the models, there was no clear winner; each model did well at different phases of the solar cycle." The present work compares the benchmark REFM with SNB³GEO, which did not operate when Perry *et al.* [2010] was published. The paper is organized as follows. Section 2 reviews the REFM and SNB³GEO tools as well as the approaches

used in their development. Section 3 describes the GOES 13 data and methodology employed in the current paper to compare the quality of the forecasts. Section 4 presents the results of this comparison. Section 5 describes that the correction methodology utilized by REFM has a potential to improve the SNB³GEO forecast.

2. The SNB³GEO and REFM Forecast Tools

The Sheffield SNB³GEO online forecast tool has been operating continually since 2 March 2012. Its results are available at http://sbg.group.shef.ac.uk/sbg2013/UOSSW/2MeV_EF.html. The Sheffield SNB³GEO model was derived using the NARMAX (Nonlinear AutoRegressive Moving Average models with eXogenous input) system identification algorithm [Leontaritis and Billings, 1985a, 1985b].

System identification techniques can automatically derive a model based on data sets that represent the input and output of a system. These are often known as black box methodologies since any knowledge of the processes within the system is not required. It is only necessary to provide data for the external influences on the system (ACE solar wind measurements) and measurements of the output data (GOES 12 electron flux measurements). NARMAX is only one of a number of methodologies that have been developed in systems science to identify unknown dynamical systems. In contrast to some other methodologies such as Neural Networks, NARMAX provides physically interpretable models that can assist in the understanding of the physical processes that underpin the evolution of complex dynamical systems. The NARMAX system identification methodology has been applied to a wide variety of complex systems in many different scientific fields. For example, in biology, Friedrich *et al.* [2009] used the technique to analyze the adaptive changes in the photoreceptors of *Drosophila*. The identification of the Belousov-Zhabotinsky chemical reaction was found from cellular automata NARMAX models [Zhao *et al.*, 2007]. In the field of space physics, NARMAX has been used to model the radiation belts [Balikhin *et al.*, 2011; Boynton *et al.*, 2013, 2015] and geomagnetic storms [Boaghe *et al.*, 2001; Wei *et al.*, 2004; Boynton *et al.*, 2011a]. The methodology has also been used to analyze the magnetospheric system. Boynton *et al.* [2011b] used the Error Reduction Ratio (ERR) methodology to determine a coupling function for the *Dst* index, which was validated by Balikhin *et al.* [2010]. Boynton *et al.* [2013] investigated the solar wind control of the electron fluxes at GEO and determined a relationship between the solar wind parameters and the evolution of the electron flux in various energy ranges. Balikhin *et al.* [2012] compared these results to the energy diffusion equation and determined that local acceleration of electrons was not dominant at GEO.

The NARMAX methodology was first developed by Leontaritis and Billings [1985a, 1985b] and a Forward Regression Orthogonal Least Squares (FROLS) algorithm, which can deduce the model structure and estimate the coefficients, was proposed by Billings *et al.* [1988, 1989]. A comprehensive review of all aspects of NARMAX methodology may be found in Billings [2013]. SNB³GEO uses a Multi-Input Single-Output (MISO) NARMAX model to represent the dynamics of the electron fluxes at GEO. The forecast at time t , $\hat{y}(t)$, resulting from the general MISO NARMAX model can be represented by equation (1).

$$\hat{y}(t) = F[y(t-1), \dots, y(t-n_y), u_1(t-1), \dots, u_1(t-n_{u_1}), \dots, u_m(t-1), \dots, u_m(t-n_{u_m}), \dots, e(t-1), \dots, e(t-n_e)] + e(t) \quad (1)$$

where y , u , and e represent the output, input, and error terms, respectively; m is the number of system inputs; and $n_y, n_{u_1}, \dots, n_{u_m}, n_e$ are the maximum time lags for the output, each of the m inputs, and the error, respectively. In the case of the SNB³GEO model the nonlinear function $F[\cdot]$ is represented in terms of a second-degree polynomial. The error term at time $t-1$ is calculated as the residual of the model, e.g., $e(t-1) = y(t-1) - \hat{y}(t-1)$. The incorporation of error terms e into the NARMAX model intends to account for unobserved inputs and measurement errors.

The NARMAX FROLS algorithm [Billings *et al.*, 1989] is one of the most advanced system identification techniques. The first stage of the algorithm is structure detection. The expansion of equation (1) as a polynomial results in many possible monomials, made up from combinations of the cross-coupled past inputs and past outputs. Most of these monomials will have virtually no effect on the system. Therefore, the monomials with greatest influence on the system need to be extracted. The NARMAX FROLS algorithm detects these monomials by the use of the ERR. This method is able to find and rank the influence of the monomials that make up the model structure. Once the model structure is detected, the next stage is the coefficient estimation. In this stage, the coefficients of each monomial in the model structure are calculated. The final stage is the validation of the model using nonlinear correlation tests [Billings and Voon, 1986; Billings and Zhu, 1995].

The SNB³GEO model was deduced by the FROLS algorithm on data between 11 July 2004 and 11 October 2005 [Boynton *et al.*, 2015]. The input parameters for the model are daily averages of solar wind velocity v , density n , and the percentage of time that the interplanetary magnetic field (IMF) remains southward within each day τ_B , based on real-time measurements from the Advanced Composition Explorer (ACE) spacecraft at L1 [Zwickl *et al.*, 1998]. The output data set were daily averaged > 2 MeV electron fluxes from GOES 12. The maximum lags employed for the FROLS algorithm were 3 days (includes lags of 1, 2, and 3 days), and the nonlinear function, F , was set to be a quadratic polynomial. The resulting NARMAX model was then implemented online and in real time to produce a forecast of > 2 MeV electron fluxes. Since the minimum lag for the inputs in the model is 1 day, it is possible to forecast 1 day into the future. For example, at midnight between days X and $X + 1$, the averaged solar wind parameters for Day X and Day $X - 1$ is calculated using the real-time ACE data that is downloaded from the Space Weather Prediction Center. Therefore, the model is able to predict the daily averaged flux for the day $X + 1$. This prediction is fixed and transferred to the data archive. It is these predictions that we are use to compare SNB³GEO with REFM. At 1 A.M. of the day $X + 1$ the data for solar wind parameters for the time period 00:00 to 01:00 are downloaded. Every hour the model automatically downloads these data (1 min resolution for ACE solar wind and IMF). The mean of the solar wind parameters for the hour 00:00–01:00 h are treated as the mean for the whole day $X + 1$ and are used as model inputs resulting in a preliminary forecast of the daily averaged flux for the day $X + 2$, but this forecast is not fixed. At 02:00, 2 h period of solar wind parameters for Day $X + 1$ is known. It is again treated as the average for the whole day $X + 1$ and an updated preliminary forecast for the day $X + 2$ appear on the model website. This procedure repeats every hour, until midnight from $X + 1$ to $X + 2$. At this moment it is possible to calculate the real value of the averaged solar wind parameters for the day $X + 1$ and forecast for the day $X + 2$ becomes fixed and transferred to the forecast archive. Only the daily averaged fixed forecasts, downloaded from the SNB³GEO archive, are used in this study. The model structure and coefficients are beyond the scope of this paper, but this will be addressed by the following publication.

The REFM, running at the Space Weather Prediction Center (SWPC) since 1996, was developed by C. Smithro based on the linear prediction filter (LPF) method of Baker *et al.* [1990], which predicts radiation belt electron flux in some energy band based on the solar wind speed at some time in the past. In this model, the relationship between some input quantity and the output electron flux is modeled as a convolution of the input quantity with a prediction filter function that is derived from historical comparisons of the input and output data. For the LPF prediction of radiation belt electron fluxes, input quantities such as geomagnetic indices (Kp and AE) or solar wind velocity have commonly been used [Nagai, 1988; Baker *et al.*, 1990].

REFM calculates 1, 2, and 3 day forecasts of the daily > 2 MeV electron fluence at GEO from observations of solar wind velocity at L1, currently from ACE. The prediction filter is derived from 30 days of solar wind velocity. Although the correlation of radiation belt fluxes with solar wind velocity is well established [Paulikas and Blake, 1979], it is unable to account for all observed variability [Blake *et al.*, 1997; Reeves *et al.*, 2011; Li *et al.*, 2011], including short-term (within 1 day) orders-of-magnitude decreases or increases in the flux. In order to improve the short-term predictions, REFM calculates a flux offset based on recent comparisons between forecasts and daily observations. The 1 day ACE forecast uses the most recent forecast and observation while the 2 and 3 day forecasts use comparisons based on the previous 20 days. For more details on REFM, please see <http://www.swpc.noaa.gov/products/relativistic-electron-forecast-model>.

3. Data and Forecast Assessment Methodology

As mentioned above, SNB³GEO has been operating since 2 March 2012 and has generated a continuous set of forecasts from this date. An archive of forecasts is available at http://www.ssg.group.shef.ac.uk/USSW/Archive_EF/All/All_EF.html from which both graphical and tabular formats of the results are available. REFM has been operating since 1996. The most recent REFM 1, 2, and 3 day ahead forecasts are available in text form from <http://services.swpc.noaa.gov/text/relativistic-electron-fluence-tabular.txt>.

For the purposes of the comparison, forecasts from the two models were compared for the period from 2 March 2012 until 1 January 2014. During this period the SNB³GEO model had a continuous record of forecasts. However, the REFM model output file provided by SWPC for this study contains 10 days for which no forecast is available. Thus, the comparison is based on 661 forecasts during this period when there were predictions from both models.

The measurements with which the forecasts are compared come from the > 2 MeV channel of the westward looking Energetic Proton Electron and Alpha Detector (EPEAD) instrument [Hanser, 2011] on the GOES 13 satellite. There are two EPEADs on each GOES, one looking westward and one looking eastward. Since the SWPC forecast office uses the observations from the westward looking EPEAD, we compare the model results with the westward observations. The data for this can be accessed as sets of Daily Particle Data (DPD) files from ftp://ftp.swpc.noaa.gov/pub/indices/old_indices/.

The correlation function and prediction efficiency have been used to assess the accuracy of the 1 day ahead predictions provided by the REFM and SNB³GEO tools. The concept of prediction efficiency is based on the model efficiency [e.g., Nash and Sutcliffe, 1970; Moriasi et al., 2007; Clauer et al., 1981; Nagai, 1988]. The prediction efficiency, which estimates the difference between the forecast (Y) and measured (X) data sets, is defined as

$$PE = 1 - \frac{1}{N} \sum \frac{(X_i - Y_i)^2}{\text{Var}(X)}$$

where N is the number of points in each data set, $\text{Var}(X) = \frac{1}{N} \sum (X_i - \langle X_i \rangle)^2$, and $\langle \dots \rangle$ is the ensemble average of the parameters. It should be noted that as the discrepancy between forecast and measured values increases the value of PE decreases and may be less than zero. In the ideal case of perfect prediction the value of PE should be equal to 1.

The correlation function is also used to provide a comparison between the daily averaged GOES13 measurements of > 2 MeV electron fluxes $F_{2\text{MeV}}$ and the 1 day ahead forecasts by the REFM and SNB³GEO models (F_{REFM} and F_{SNB} , respectively) for both the fluxes and their logarithms. For example, the correlation for the $\log_{10}(\text{flux})$ between the GOES13 measurements and forecast by the SNB³GEO model was defined as

$$C_{\log(\text{SNB})} = \frac{1}{N} \sum_{i=1}^N \frac{(\log_{10}(F_{2\text{MeV}}(i)) - \langle \log_{10}(F_{2\text{MeV}}(i)) \rangle)(\log_{10}(F_{\text{SNB}}(i)) - \langle \log_{10}(F_{\text{SNB}}(i)) \rangle)}{\sqrt{\text{Var}(\log_{10}(F_{2\text{MeV}}))\text{Var}(\log_{10}(F_{\text{SNB}}))}}$$

where $N = 661$ is the number of data points.

The prediction efficiency and correlation are common measures of the performance of a model over a long period. These measures obscure the performance of the model for the relatively infrequent but important space weather cases in which the flux exceeds a large threshold. Bodeau [2010] provides a comprehensive summary of the literature reporting anomalies caused by deep charging. Anomalies have been associated with several quantities related to deep charging such as external flux, external fluence, external current density, and internal fluence. In one study, multiple anomalies observed on a geosynchronous communication satellite were associated with 2 day >2 MeV electron external fluences between 3×10^8 and 1×10^{10} electrons/(cm² sr) [Wrenn, 1995]. This range corresponds approximately to the three daily fluence thresholds selected for the present study. For such infrequent events, the Heidke skill score (HSS) is an appropriate measure [Heidke, 1926; Doswell et al., 1990; Balch, 2008]. The HSS is the ratio of the total number of correct predictions divided by the total number of observations, from both of which has been subtracted the expected number of correct forecasts by chance. Given that w is the number of successful negative predictions, x is the number of successful positive predictions, y is the number of false negatives, and z is the number of false positives, the HSS is given by [Doswell et al., 1990]

$$S = \frac{2(xw - yz)}{y^2 + z^2 + 2xw + (y + z)(x + w)} \quad (2)$$

In the present study, a successful positive prediction is one in which the predicted daily fluence is above some threshold.

4. Results

Table 1 displays the resulting values of prediction efficiency PE and correlation calculated for the fluxes and their logarithms using the whole data set of forecasts/measurements. With the exception of the prediction efficiency for fluxes from REFM, all other parameters point to a very similar accuracy for the forecasts by the two models, with a marginally ($\approx 5\text{--}10\%$) better accuracy in favor of SNB³GEO. The prediction efficiency

Table 1. A Comparison of the Prediction Efficiencies and Correlations Obtained by Comparing the Forecasts of the > 2 MeV Electron Flux and \log_{10} (Flux) From the REFM and SNB³GEO Models With Measurements From the GOES 13 Satellite

Model	PE Flux	Correlation Flux	PE \log_{10} Flux	Correlation \log_{10} Flux
REFM	-1.31	0.73	0.70	0.85
SNB ³ GEO	0.63	0.82	0.77	0.89

for fluxes from the REFM model has a negative value -1.31, indicating it to be substantially worse than the forecasts by SNB³GEO which has a PE = 0.6313.

The large difference between the PEs for $F_{2\text{MeV}}$ and $\log_{10}(F_{2\text{MeV}})$ requires some consideration of which is a better measure of model performance. The scatter plots for the two models are similar (Figure 1), with the somewhat greater scatter in the REFM results leading to slightly larger correlation values for SNB³GEO. The large differences in PE for $F_{2\text{MeV}}$, especially the large negative PE for REFM, are dominated by the residuals due to the greatest fluences ($> 10^9 \text{ cm}^{-2} \text{ sr}^{-1} \text{ d}^{-1}$), which are greater for REFM than for SNB³GEO. In contrast, the contributions to the numerator of the PE for $\log_{10}(F_{2\text{MeV}})$ are similar across the entire dynamic range. Since the radiation belt electron flux is approximately lognormally distributed and therefore the logarithm of flux is approximately normally distributed, the variance of $\log_{10}(F_{2\text{MeV}})$ is a more meaningful statistical measure than the variance of $F_{2\text{MeV}}$. Therefore, the PE for $\log_{10}(F_{2\text{MeV}})$ for both REFM and SNB³GEO (and for any radiation belt prediction model) is the more meaningful of the two. This comparison period is short with respect to the solar cycle due to the fact that the SNB³GEO model archive started to operate just recently (since 2012). As such, it would be of interest to show how the PE and the correlation for the SNB³GEO would relate over a longer period. The only possibility is to calculate a “past-cast” for the longer period. The model output has been calculated for such a longer period (01.01.1998–01.03.2012) with exclusion of training period (11.07.2004–11.10.2005). The measurements of the output were obtained from the Space Weather Prediction Center. Depending upon the year, measurements by GOES 9, 8, 11, 12, and 13 have been used. The resulting PE and the correlation obtained for this “past-cast” period are PE = 0.6739 and C = 0.8245, correspondingly, for fluxes, while PE = 0.8187 and C = 0.9081 for logarithms of the fluxes. These values are very close to those in the Table 1. The value of the prediction efficiency (PE = 0.8187) is similar but seems slightly higher than the mean of the values from Figure 2 of Perry *et al.* [2010] obtained for three forecast models that were investigated by Perry *et al.* [2010].

Perhaps the most important requirement for the forecast of fluxes of relativistic electrons is to obtain accurate estimates for periods when the fluxes are high. The Heidke skill scores for the prediction of daily fluences greater than 10^8 , $10^{8.5}$, and 10^9 by REFM and SNB³GEO are summarized in Tables 2 and 3. These tables show that during the comparison period there were 129 days with fluences $> 10^8$, 44 days with fluences $> 10^{8.5}$, and only 7 days with fluences $> 10^9 \text{ cm}^{-2} \text{ sr}^{-1} \text{ d}^{-1}$. At all three thresholds, SNB³GEO had better HSS. At the greatest

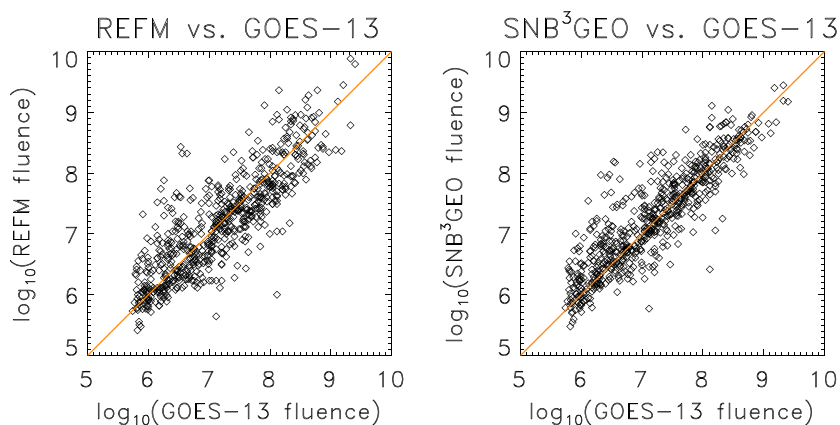


Figure 1. Scatterplots of (left) REFM and (right) SNB³GEO 1 day predictions versus GOES 13 observations for the period of interest (2 March 2012 to 31 December 2013). The diagonal is the line of perfect correlation. The lower cutoff in the observations corresponds to an instrument flux background of 10 electrons/(cm² sr s).

Table 2. Contingency Tables and Heidke Skill Scores for the REFM Predictions

Fluence ($\text{cm}^{-2} \text{sr}^{-1} \text{d}^{-1}$)	$>10^8$		$>10^{8.5}$		$>10^9$	
REFM HSS	0.666		0.482		0.437	
Observation	Yes	No	Yes	No	Yes	No
Forecast						
Yes	$x = 86$	$z = 22$	$x = 23$	$z = 22$	$x = 4$	$z = 7$
No	$y = 43$	$w = 510$	$y = 21$	$w = 595$	$y = 3$	$w = 647$

threshold, REFM had seven false positives and three false negatives, while SNB³GEO had only two false positives and three false negatives. For a threshold of $10^{8.5}$, SNB³GEO had fewer false positives and negatives than REFM, while at the lowest threshold, it had many fewer false negatives though more false positives.

We estimate the sensitivity of the skill score to a single event by decreasing or increasing x , y , or z by 1 and accordingly decreasing or increasing one of the other variables by 1. This results in 12 unique perturbed combinations giving S , from which we calculate the RMS difference with respect to the actual S . For the threshold of 10^8 , this RMS difference was 0.66% for REFM and 0.53% for SNB³GEO. For the threshold of $10^{8.5}$, this RMS difference was 2.4% for REFM and 1.6% for SNB³GEO. For the threshold of 10^9 , this RMS difference was 13% for REFM and 12% for SNB³GEO. Clearly, the Heidke skill score is sensitive to a single event when there are very few events above a given threshold.

5. Discussion

The parameters of the forecast accuracy presented in Tables 1–3 show that the accuracy of the forecast is very similar for the NOAA and SNB³GEO models with SNB³GEO performing slightly (5%–10%) better. One of the currently unresolved problems in the forecast of high-energy electron fluxes at GEO with an advance time in excess of the time required for the solar wind to propagate from L1 to the magnetopause is the inability to predict dropouts caused by magnetopause shadowing [Shprits *et al.*, 2006; Loto'aniu *et al.*, 2010; Turner *et al.*, 2012, 2014]. This problem can only be solved by the accurate forecast of solar wind parameters at L1. Without such a forecast at L1 the reliability of any model to forecast the dropouts of fluxes caused by the earthward motion of the magnetopause will be very low. Moreover, in the case of the SNB³GEO model a significant dropout due to a large displacement of the magnetopause can affect more than one forecast value. The REFM model methodology includes a correction factor for the subsequent forecast based on recent comparison between forecasts and daily observations as it was mentioned above. In general, the incorporation of such a correction procedure into a NARMAX-type model should not lead to a significant improvement in the performance of the model since, according to the NARMAX methodology, the error terms (e) of the model should account for the effects of inputs to the dynamical system that are not included within the model. In the development of SNB³GEO, it was found that the incorporation of error terms had little effect on the forecast quality. However, it is worth checking if some improvement in the accuracy of the SNB³GEO model with respect to magnetopause shadowing can be achieved by incorporating a methodology similar to that used by REFM, but in a simplified version since the time scale of magnetopause displacements is less than 1 day. This correction approach has been adopted for the forecasts of the SNB³GEO model in the following way.

Table 3. Contingency Tables and Heidke Skill Scores for the SNB³GEO Predictions

Fluence ($\text{cm}^{-2} \text{sr}^{-1} \text{d}^{-1}$)	$>10^8$		$>10^{8.5}$		$>10^9$	
SNB ³ GEO HSS	0.738		0.634		0.612	
Observation	Yes	No	Yes	No	Yes	No
Forecast						
Yes	$x = 106$	$z = 33$	$x = 31$	$z = 19$	$x = 4$	$z = 2$
No	$y = 23$	$w = 499$	$y = 13$	$w = 598$	$y = 3$	$w = 652$

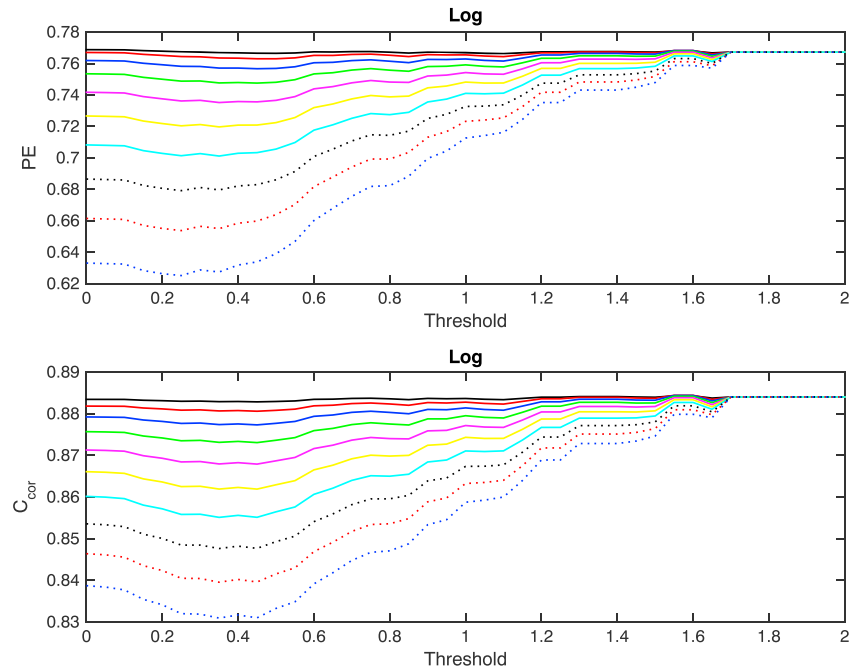


Figure 2. Variation of the (top) prediction efficiency and (bottom) correlation as functions of the threshold for periods when the SNB³GEO forecasts overshoot the measured GOES 13 values. The various lines correspond to the use of different correction factors (10% = black, 20% = red, 30% = blue, 40% = green, 50% = magenta, 60% = yellow, 70% = cyan, 80% = dotted black, 90% = dotted red, and 100% = dotted blue).

If at day i the value of $\log_{10}(F_{\text{SNB}}(i))$ is greater than the value measured by GOES 13 and the difference between them D exceeds some threshold T such that $D = \log_{10}(F_{2\text{MeV}}(i)) - \log_{10}(F_{\text{SNB}}(i)) \geq T$, the forecast for the day $i + 1$ was calculated as

$$\log_{10}(F_{\text{SNB}}(i + 1)) - D \cdot C/100\%, \quad (3)$$

where $\log_{10}(F_{\text{SNB}}(i + 1))$ is the actual forecast by the SNB³GEO model for day $i + 1$ and $0 \leq C \leq 100\%$ is a correction factor. Figures 2 and 3 display the resulting PE and values of the correlation for the range of threshold values $0.1 \leq T \leq 2$. The X axis corresponds to the values of the threshold, while the Y axis represents the PE (Figure 2, top) and the correlation value (Figure 2, bottom). In both cases, the various lines correspond to the use of different correction factors (10% = black, 20% = red, 30% = blue, 40% = green, 50% = magenta, 60% = yellow, 70% = cyan, 80% = dotted black, 90% = dotted red, and 100% = dotted blue). Since the calculations have been performed using only the SNB³GEO model, the 10 data points that correspond to the days when the REFM forecast was not available were not excluded in these calculations.

It can be seen from Figures 2 and 3 that for threshold values greater than 1.7 (i.e., for thresholds: 1.75, 1.80, 1.85, 1.90, 1.95, and 2.00), the values of PE and the correlations do not depend upon the correction factor and threshold. This occurs because D is always below 1.7, implying that if the threshold level is set to a value that exceeds 1.7 the correction procedure will not be triggered, and so the values for the PE and correlation correspond to the performance of the SNB³GEO model without any correction. It is obvious from this figure that there is no significant increase in the values of the PE or correlation for any combination of the threshold and the correction factor. It is evident from these figures that the black and red lines, which correspond to the 10% and 20% correction factors, do not exhibit significant variations. All lines corresponding to higher correction factors indicate a significant reduction in the values of PE and correlation for lower thresholds. However, as the threshold value increases, the significance of the reduction diminishes and for values of the threshold above 1.7 all lines converge to the values of PE and correlation function without any correction. Overall Figure 2 illustrates the reliable performance of the model and shows that any additional correction will not lead to any significant improvement in the statistical characteristics of the forecast reliability. Figure 3 is the same as Figure 2 but with significantly increased Y axis scale to emphasize the evolution of the dynamics of the PE and correlation values that correspond to the correction factors 10% (black), 20% (red), and 30% (blue).

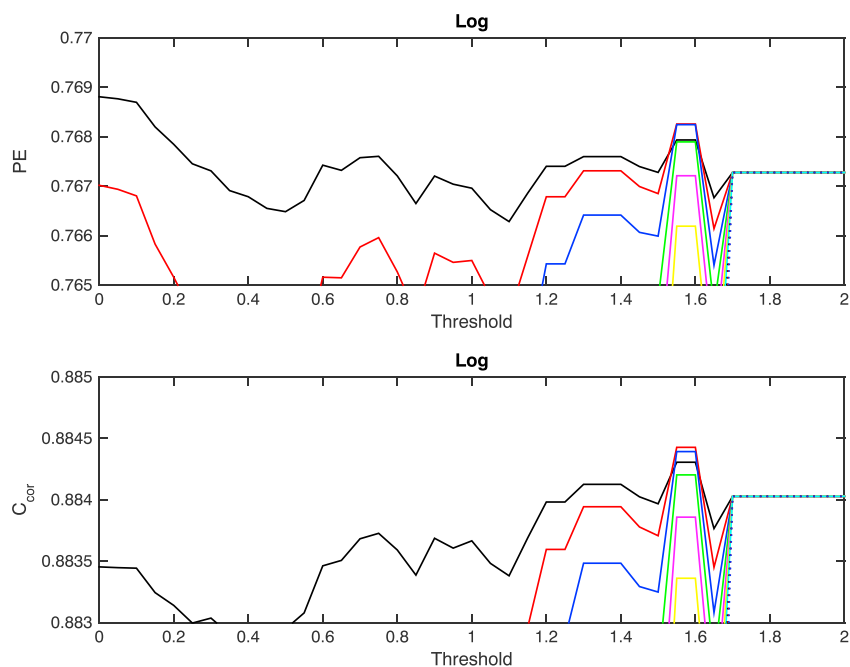


Figure 3. Same as Figure 2 but with a higher resolution for the Y axis to emphasize small changes observed in the prediction efficiency and correlation.

It can be seen that for thresholds $T < 0.2$ and a correction factor of 10% the resulting prediction efficiency is slightly higher in comparison to the value without any correction but the difference is statistically negligible ($\approx 0.15\%$). For the same correction factor (10%) and range of thresholds $0.2 < T < 1.5$ the correlation function is insignificantly lower in comparison to the output of the SNB³GEO model without any correction. For threshold values $T > 1.55$ and correction factors below 40%, the values of the PE and the correlation function exceed the values of the uncorrected output of the SNB³GEO model. Statistically, the change introduced by the correction procedure is negligible and, in the first instance, it appears that the correction procedure is not able to improve the performance of SNB³GEO. However, let us consider the very minor improvement that the correction procedure exhibits for values of the threshold around 1.55. If the difference between the logarithms of the forecast and real measurements $\log_{10}(F_{2\text{MeV}}(i) - \log_{10}(F_{\text{SNB}}(i))$ exceeds the threshold $T = 1.55$, the predicted flux should be at least $10^{1.55} \approx 34$ times higher than that measured. For the whole period used in the comparison there were only 3 days with such a high discrepancy between the predicted and measured daily fluxes. While for the whole comparison data set the improvement in statistical parameters is negligible, it should be emphasized that this is a result of the correction applied to only 3 days from the 671 day period. Therefore, the correction that corresponds to days that follow the significant overshoot by the model may be very valuable. It is important to improve the forecast accuracy, even in cases of a very few points but with significant mismatch with the data. To quantify these improvements, the mean-squared errors (MSE) were calculated for these three points using the SNB³GEO model with a zero, 10%, 20%, 30%, and 40% correction factors. Without correction MSE is equal to 0.77. The values of MSE with correction procedure implemented were 0.66, 0.61, 0.61, and 0.67 for correction factors 10%, 20%, 30%, and 40%, respectively. Therefore, for a correction factor of 20% the procedure leads to a reduction of the MSE by about 21%.

While the improvement in MSE occurs, the set of only three events cannot be used to justify the conclusion that the correction procedure has the positive effect on the forecast. Therefore, the same calculations have been done for the past-cast period (1 January 1998 to 01 March 2012) with exclusion of training period (11 July 2004 to 11 October 2005). All together, there were 36 events for which the threshold 1.55 has been reached. Without correction MSE is equal to 0.75. The values of MSE with correction procedure implemented were 0.66, 0.65, 0.71, and 0.85 for correction factors 10%, 20%, 30%, and 40%, respectively. Therefore, for a correction factor of 20% the procedure for longer intervals lead to a reduction of the MSE by about 13%. It is worth noting that while the set of the 36 events is definitely more representative than the set of only three events, it is still too small to support the definite statement that the correction procedure improves the forecasts.

However, it points out that this correction procedure has a potential to make the forecast more reliable on the days after significant forecast overshoots. The effect of applying a similar procedure on the “undershoots” of the forecast also has been investigated and it was shown that while it may improve MSE for a few events during the forecast period (2 March 2012 to 31 December 2013), the effect of the correction during the longer past-cast period was adverse.

6. Conclusions

A comparison of the accuracy of the 1 day ahead forecasts of daily averaged fluxes of energetic electrons with energies in excess of 2 MeV from the Sheffield SNB³GEO and NOAA REFM models has been carried out for the time interval from 2 March 2012 to 1 January 2014. GOES 13 data were used to identify the accuracy of the forecast for both models. The concepts of prediction efficiency, correlation function, and Heidke skill scores have been exploited for the quantitative assessment of the reliability of the two forecasts. The main conclusion of the study is that both the prediction efficiency and the values of the correlation function, as well as the Heidke skill scores for extreme fluxes, point to the higher accuracy of the SNB³GEO forecast. In addition, the effect of the methodology used for forecast correction in the REFM model on the SNB³GEO prediction accuracy has been investigated. It is shown that while the incorporation of this methodology into SNB³GEO model results in a negligible increase in the overall prediction efficiency and the correlation, it may reduce errors related to the small number of data points that follow the significant overshoot for the SNB³GEO forecast relative to the GOES 13 measurements. The very low number of such data points explains why the subsequent reduction of errors leads to a negligible increase of the statistical forecast parameters (prediction efficiency, correlation, and Heidke skill score). However, the incorporation of the REFM methodology into the SNB³GEO model has the potential to reduce the effects of SNB³GEO significant overshoots and therefore also may improve the forecast quality.

Acknowledgments

M.A.B. would like to acknowledge financial support from ISSI. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 637302. M.A.B., S.A.B., and R.J.B. would like to acknowledge financial support from EPSRC UK, Platform grant EP/H00453X/1. H.A. would like to acknowledge financial support from NASA Postdoctoral Program at Goddard Space Flight Center. DGS acknowledges that portions of the work at NASA/GSFC were funded by the Van Allen Probes mission. As mentioned in the text, forecasts from the SNB³GEO model are available from http://www.ssg.group.shef.ac.uk/USSW/Archive_EF/All/All_EF.html and the GOES 13 Daily Particle Data (DPD) files from ftp://ftp.swpc.noaa.gov/pub/indices/old_indices/. We thank H. Singer and SWPC for providing the REFM model outputs. The REFM outputs are provided as supporting information.

References

- Baker, D., R. Belian, P. Higbie, R. Klebesadel, and J. Blake (1987), Deep dielectric charging effects due to high-energy electrons in Earth's outer magnetosphere, *J. Electrostat.*, *20*(1), 3–19.
- 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*(A9), 15,133–15,140, doi:10.1029/JA095iA09p15133.
- Balch, C. C. (2008), Updated verification of the Space Weather Prediction Center's solar energetic particle prediction model, *Space Weather*, *6*, S01001, doi:10.1029/2007SW000337.
- Balikhin, M. A., R. J. Boynton, S. A. Billings, M. Gedalin, N. Ganushkina, D. Coca, and H. Wei (2010), Data based quest for solar wind-magnetosphere coupling function, *Geophys. Res. Lett.*, *37*, L24107, doi:10.1029/2010GL045733.
- Balikhin, M. A., R. J. Boynton, S. N. Walker, J. E. Borovsky, S. A. Billings, and H. L. Wei (2011), Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit, *Geophys. Res. Lett.*, *38*, L18105, doi:10.1029/2011GL048980.
- Balikhin, M. A., M. Gedalin, G. D. Reeves, R. J. Boynton, and S. A. Billings (2012), Time scaling of the electron flux increase at GEO: The local energy diffusion model vs observations, *J. Geophys. Res.*, *117*, A10208, doi:10.1029/2012JA018114.
- Billings, S. A. (2013), *Nonlinear Systems Identification: NARMAX, Methods in the Time, Frequency, and Spatio-Temporal Domains*, Wiley-Blackwell, Chichester, U. K.
- Billings, S. A., and W. S. F. Voon (1986), Correlation based model validity tests for non-linear models, *Int. J. Control*, *44*(1), 235–244.
- Billings, S. A., and Q. M. Zhu (1995), Model validation tests for multivariable nonlinear models including neural networks, *Int. J. Control*, *62*(2), 749–766, doi:10.1080/00207179508921566.
- Billings, S. A., M. Korenberg, and S. Chen (1988), Identification of non-linear output affine systems using an orthogonal least-squares algorithm, *Int. J. Syst. Sci.*, *19*, 1559–1568.
- Billings, S. A., S. Chen, and M. Korenberg (1989), Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator, *Int. J. Control*, *49*(6), 2157–2189.
- Blake, J. B., D. N. Baker, N. Turner, K. W. Ogilvie, and R. P. Lepping (1997), Correlation of changes in the outer-zone relativistic-electron population with upstream solar wind and magnetic field measurements, *Geophys. Res. Lett.*, *24*(8), 927–929, doi:10.1029/97GL00859.
- Boaghe, O., M. Balikhin, S. A. Billings, and H. Alleyne (2001), Identification of nonlinear processes in the magnetospheric dynamics and forecasting of Dst index, *J. Geophys. Res.*, *106*, 30,047–30,066, doi:10.1029/2000JA900162.
- Bodeau, M. (2010), High energy electron climatology that supports deep charging risk assessment in GEO, in *48th AIAA Aerospace Sciences Meeting*, pp. 2010–1608, AIAA, Orlando, Fla.
- Boynton, R. J., M. A. Balikhin, S. A. Billings, A. S. Sharma, and O. A. Amariutei (2011a), Data derived NARMAX Dst model, *Ann. Geophys.*, *29*(6), 965–971, doi:10.5194/angeo-29-965-2011.
- Boynton, R. J., M. A. Balikhin, S. A. Billings, H. L. Wei, and N. Ganushkina (2011b), Using the NARMAX OLS-ERR algorithm to obtain the most influential coupling functions that affect the evolution of the magnetosphere, *J. Geophys. Res.*, *116*, A05218, doi:10.1029/2010JA015505.
- Boynton, R. J., M. A. Balikhin, S. A. Billings, G. D. Reeves, N. Ganushkina, M. Gedalin, O. A. Amariutei, J. E. Borovsky, and S. N. Walker (2013), The analysis of electron fluxes at geosynchronous orbit employing a NARMAX approach, *J. Geophys. Res. Space Physics*, *118*, 1500–1513, doi:10.1002/jgra.50192.
- Boynton, R. J., M. A. Balikhin, and S. A. Billings (2015), Online NARMAX model for electron fluxes at GEO, *Ann. Geophys.*, *33*, 405–411, doi:10.5194/angeo-33-405-2015.

- Clauer, C. R., R. L. McPherron, C. Searls, and M. G. Kivelson (1981), Solar wind control of auroral zone geomagnetic activity, *Geophys. Res. Lett.*, *8*(8), 915–918, doi:10.1029/GL008i008p00915.
- Doswell, C. A. I., R. Davies-Jones, and D. L. Keller (1990), On summary measures of skill in rare event forecasting based on contingency tables, *Weather Forecasting*, *5*, 576–585.
- Fennell, J. F., H. C. Koons, M. W. Chen, and J. B. Blake (2000), Internal charging: A preliminary environmental specification for satellites, *IEEE Trans. Plasma Sci.*, *28*(6), 2029–2036.
- Friedrich, U., D. Coca, S. A. Billings, and M. Juusola (2009), Data modelling for analysis of adaptive changes in fly photoreceptors, in *Neural Information Processing, 16th International Conference on Neural Information Processing (ICONIP 2009)*, vol. 5863, edited by C. S. Leung, M. Lee, and J. H. Chan, pp. 34–48, Springer, Berlin.
- Hanser, F. A. (2011), EPS/HEPAD calibration and data handbook, *Tech. Rep. GOESN-ENG-048D*, Assurance Technol. Corp., Carlisle, Mass.
- Heidke, P. (1926), Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst, *Geogr. Ann.*, *8*, 301–349.
- Leontaritis, I., and S. A. Billings (1985a), Input-output parametric models for nonlinear systems. Part I: Deterministic nonlinear systems, *Int. J. Control*, *41*, 309–328.
- Leontaritis, I., and S. A. Billings (1985b), Input-output parametric models for nonlinear systems. Part II: Stochastic nonlinear systems, *Int. J. Control*, *41*, 329–344.
- 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., M. Temerin, D. N. Baker, and G. D. Reeves (2011), Behavior of MeV electrons at geosynchronous orbit during last two solar cycles, *J. Geophys. Res.*, *116*, A11207, doi:10.1029/2011JA016934.
- Ling, A. G. (2000), FLUXPRED: A software tool for electron flux prediction at geosynchronous orbit, *Rep. AFRL-VS-TR-1597*, Air Force Res. Lab., Hanscom AFB, Mass.
- 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*, *8*, S09003, doi:10.1029/2010SW000576.
- Loto'aniu, T. M., H. J. Singer, C. L. Waters, V. Angelopoulos, I. R. Mann, S. R. Elkington, and J. W. Bonnell (2010), Relativistic electron loss due to ultralow frequency waves and enhanced outward radial diffusion, *J. Geophys. Res.*, *115*, A12245, doi:10.1029/2010JA015755.
- Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. Am. Soc. Agric. Biol. Eng.*, *50*(3), 885–900.
- Nagai, T. (1988), "Space weather forecast": Prediction of relativistic electron intensity at synchronous orbit, *Geophys. Res. Lett.*, *15*(5), 425–428, doi:10.1029/GL015i005p00425.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.
- Paulikas, G. A., and J. B. Blake (1979), Effects of the solar wind on magnetospheric dynamics: Energetic electrons at the synchronous orbit, in *Quantitative Modeling of Magnetospheric Processes*, *Geophys. Monogr. Ser.*, vol. 21, edited by W. P. Olson, pp. 180–202, AGU, Washington, D. C.
- Perry, K. L., G. P. Ginet, A. G. Ling, and R. V. Hilmer (2010), Comparing geosynchronous relativistic electron prediction models, *Space Weather*, *8*, S12002, doi:10.1029/2010SW000581.
- Reeves, G. D., S. K. Morley, R. H. W. Friedel, M. G. Henderson, T. E. Cayton, G. Cunningham, J. B. Blake, R. A. Christensen, and D. Thomsen (2011), On the relationship between relativistic electron flux and solar wind velocity: Paulikas and Blake revisited, *J. Geophys. Res.*, *116*, A02213, doi:10.1029/2010JA015735.
- Shprits, Y. Y., R. M. Thorne, R. Friedel, G. D. Reeves, J. Fennell, D. N. Baker, and S. G. Kanekal (2006), Outward radial diffusion driven by losses at magnetopause, *J. Geophys. Res.*, *111*, A11214, doi:10.1029/2006JA011657.
- Turner, D. L., Y. Shprits, M. Hartinger, and V. Angelopoulos (2012), Explaining sudden losses of outer radiation belt electrons during geomagnetic storms, *Nat. Phys.*, *8*, 208–212.
- Turner, D. L., et al. (2014), On the cause and extent of outer radiation belt losses during the 30 September 2012 dropout event, *J. Geophys. Res. Space Physics*, *119*, 1530–1540, doi:10.1002/2013JA019446.
- Vampola, A. L. (1987), Thick dielectric charging on high-altitude spacecraft, *J. Electrostat.*, *20*(1), 21–30.
- Wei, H. L., S. A. Billings, and M. Balikhin (2004), Prediction of the Dst index using multiresolution wavelet models, *J. Geophys. Res.*, *109*, A07212, doi:10.1029/2003JA010332.
- Wrenn, G. L. (1995), Conclusive evidence for internal dielectric charging anomalies on geosynchronous communications spacecraft, *J. Spacecr. Rockets*, *32*(3), 514–520.
- Zhao, Y., S. A. Billings, and A. F. Routh (2007), Identification of the Belousov-Zhabotinskii reaction using cellular automata models, *Int. J. Bifurcat. Chaos*, *17*(5), 1687–1701, doi:10.1142/S0218127407017999.
- Zwicky, R. D., et al. (1998), The NOAA Real-Time Solar-Wind (RTSW) system using ACE data, *Space Sci. Rev.*, *86*(1–4), 633–648.