White Rose
university consortium
Universities of Leeds, Sheffield & York

Annales
Geophysicae

Open Access

Application of nonlinear autoregressive moving average exogenous input models to geospace: advances in understanding and space weather forecasts

R. J. Boynton[1], **M. A. Balikhin**[1], **S. A. Billings**[1], **and O. A. Amariutei**[2]

[1]ACSE, University of Sheffield, Sheffield, UK
[2]Finnish Meteorological Institute, Helsinki, Finland

Correspondence to: R. J. Boynton (r.boynton@sheffield.ac.uk)

Abstract. The nonlinear autoregressive moving average with exogenous inputs (NARMAX) system identification technique is applied to various aspects of the magnetospheres dynamics. It is shown, from an example system, how the inputs to a system can be found from the error reduction ratio (ERR) analysis, a key concept of the NARMAX approach. The application of the NARMAX approach to the Dst (disturbance storm time) index and the electron fluxes at geostationary Earth orbit (GEO) are reviewed, revealing new insight into the physics of the system. The review of studies into the Dst index illustrate how the NARMAX approach is able to find a coupling function for the Dst index from data, which was then analytically justified from first principles. While the review of the electron flux demonstrates how NARMAX is able to reveal new insight into the physics of the acceleration and loss processes within the radiation belt.

Keywords. Magnetospheric Physics (Solar wind–magnetosphere interactions)

1 Introduction

The standard approach to the study of physical systems is to build a mathematical model of the processes involved from first principles and then conjugate these models into dynamical equations that govern how the physical object will evolve over time. However, with our present level of knowledge, there are many complex systems that we are not able to deduce a model from first principles. For example, the human brain and other biological systems are many years away from being understood in a manner in which a model can be derived from first principles. For such systems, there may be many possible external influences but only one or two that actually control how it will evolve over time, i.e, the number of degrees of freedom is not known. However, it is known they evolve under some external influences, these can be considered as the inputs to the system. Measurements of the how the system responds to these inputs can also be assumed to represent the state, which can be considered the output of the system. From the input–output data, system identification techniques can be employed to automatically determine dynamical equations that govern the evolution of the complex physical system.

The methods of system identification require the mapping of the inputs to the output, which can be achieved by using a number of different approaches. One of the most well known techniques is neural networks (NN) (McCulloch and Pitts, 1943). A neural network consists of multiple interconnected mathematical neurons, forming a network. There are many different topologies that the network can take, the most popular and most implemented network is the multi-layer perceptron (Rumelhart and MacClelland, 1986). It is a feed-forward network, starting from an input layer, through one or more hidden layers containing the neurons, each with activation function, connected by weights and ending at the output. This makes it very difficult to understand how the inputs are coupled within the network. Herein lies the major problem of NN: they are not physically interpretable. The nonlinear autoregressive moving average with exogenous inputs (NARMAX) technique (Leontaritis and Billings, 1985a, b) is a similar technique to NN but more useful, in that the algorithm can return a physically interpretable polynomial. The NARMAX model can be represented by the equation:

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

Here, the output at time t can be represented as a function, F, of the previous values of inputs $u(t)$, output $y(t)$ and noise $e(t)$, where n_y, n_{u_1}, ..., n_{u_m}, n_e are the maximum time lags of the output, the m inputs of the system and the noise respectively. The function F can be set to a polynomial with a specified degree of nonlinearity, where the monomials will be the cross-coupled lagged inputs, outputs and noise. As the number of inputs, lags and degree of nonlinearity increase, the number of possible monomials will increase drastically. However, most of these monomials will have no physical meaning for the system, so an algorithm needs to search these cross-coupled combinations for the terms with the most significance. This is the first stage of the NARMAX methodology, called model structure detection, and is achieved by the orthogonal least squares–error reduction ratio (OLS–ERR) algorithm. The second stage of estimating the coefficients for each of the terms identified by structure detection is also encompassed by this algorithm, while the final stage validates the model by exploiting both dynamic and statistical approaches (Billings and Voon, 1986; Billings and Zhu, 1989).

In Sect. 2, a brief description of the NARMAX algorithm is given, along with the definition of the ERR. Section 3 employs an example system to show that the ERR is able to find the inputs of the system, while the correlation function, which is often used in the search for inputs, cannot. In Sect. 4 the studies of the Dst index, using the NARMAX approach, are reviewed, while Sect. 5 reviews the NARMAX studies of the electron flux.

2 The NARMAX algorithm

In the case of a polynomial basis, $F[\cdot]$ represents a linear-in-the-parameters polynomial model. The terms of this polynomial model are comprised of all the possible cross-coupled combinations of the components to the predetermined power. Thus, Eq. (1) becomes

$$y = \sum_{i=1}^{M} p_i \theta_i + e, \qquad (2)$$

where y is the output time series vector, θ_i is the coefficient of the ith time series monomial vector p_i and M is the total number of monomials. The OLS–ERR utilises the Gram-Schmidt procedure so that each of the of the monomial time vectors, p_i, are made orthogonal to each other. So, orthogonalising Eq. (2) results in

$$y = \sum_{i=1}^{M} w_i g_i + e, \qquad (3)$$

where w_i is the ith orthogonalised monomial time series vector and g_i is the coefficient. By orthogonalising the monomials, the multiplication between different orthogonalised monomials, w_i, will result in zero, e.g. $w_i^T w_j = 0$, where $i \neq j$. This allows for the separation of each monomial's contribution to the explained output variance. Multiplying Eq. (3) by y^T leads to

$$y^T y = \sum_{i=1}^{M} w_i^T g_i \sum_{j=1}^{M} w_j^T g_j + \sum_{i=1}^{M} w_i^T g_i e$$
$$+ e^T \sum_{j=1}^{M} w_j^T g_j + e^T e, \qquad (4)$$

where $w_k^T e = 0$ and $e^T w_k = 0$ assuming all stochastic processes are ergodic, and the noise of the system is zero mean and uncorrelated with the monomials; $e^T e$ is the variance of the noise, σ_e^2; and all $w_i^T w_j = 0$ for $i \neq j$. This yields

$$y^T y = \sum_{i=1}^{M} w_i^T w_i g_i^2 + \sigma_e^2, \qquad (5)$$

where each $w_i^T w_i g_i^2$ represents the monomial's contribution to the outputs dependent variable variance. Thus, the ERR for the ith monomial is defined as

$$\text{ERR}_i = \frac{w_i^T w_i g_i^2}{y^T y} \qquad (6)$$

and represents the percentage of total output dependent variable variance attributed to each monomial. Therefore, each of the many monomials can be quantified and the monomials with the highest ERR are selected for the model structure, concluding the first stage of the NARMAX methodology. The coefficient, θ, for each of the selected monomials can then be calculated from the orthogonalised monomials by employing a least squares method, completing the second stage of the methodology and resulting in the model.

The NARMAX methodology is highly versatile and is currently employed in many different fields, ranging from biological systems to financial systems. Therefore, the NARMAX is a very powerful technique and ideal for scientific fields such as space physics since it is possible to, in some sense, reverse engineer the results to gain physical understanding about the system and the processes involved.

In the field of space physics, the magnetosphere is a highly complex system, with many processes taking place on spatial scales from metres to tens of kilometres. In many cases, it is not known what parameters influence a certain state of the magnetosphere, out of the many possible parameters that act upon it. To solve this problem, the structure detection

stage of the NARMAX algorithm, where the ERR analysis is applied, can be used to search through many combinations of many different parameters to find the terms that have the most significance on the system. The term with the higher ERR accounts for a larger amount of the output variance and is therefore a more appropriate term.

In the past, the correlation function has been employed to find the combination of solar wind parameters that most influence certain aspects of the magnetosphere (Newell et al., 2007). However, applying the correlation function to a nonlinear system, such as the terrestrial magnetosphere, may lead to ambiguous results. (Boynton et al., 2011b) illustrated a simple example of this, using a simple quadratic equation were the output y is equal to the square of the of a zero mean input x, $y = x^2$. even though x is the input the correlation between between y and x will be zero. Therefore, the linear correlation function should not be applied to nonlinear systems.

3   The ERR analysis

An artificial system was created to show that the ERR is able to identify the inputs. This system is represented by

$$y(t) = -0.25u(t) + 0.2w(t-2) - 0.3q(t-1)$$
$$+ 0.07pr^2(t-3) - 0.04q^3(t-2) + e(t), \tag{7}$$

where the output y at time t is a function of the inputs p, q, r, u and w, and the noise e. Here, e was a zero mean signal to simulate the noise. However, in the real case of obtaining the model structure there will be many possible inputs, so, more inputs, s, v and x, were included in the search, which like the other inputs were just random signals. It must be noted that each of the inputs and noise signal all had 1000 data points. Also, since the degree of nonlinearity or the maximum lags of the system are not known either, these were both set to be four. Therefore, the algorithm would search through four lags, plus the current time (t, $t-1$, ..., $t-4$), and every combination of the inputs to the power of four, resulting in a total of 135 750 terms to search. Table 1 shows the terms with the five highest ERR. The ERR analysis has found all the model's terms, linear and nonlinear, from Eq. (7), with $r^2 p(t-3)$ accounting for the most output variance.

On the other hand, if the correlation function is employed to find the model structure, the results will be misleading. To demonstrate this fact, the 135 750 terms that the ERR searched through were correlated with the output. Table 2 shows the terms with the five highest correlations with the output. The $q(t-1)$ term is involved in all five of the terms, which on its own accounted for the second highest ERR. However, according to the correlation function, the lags of v, which is not even included Eq. (7), also have a large influence on the output. The correlation function does not even recognise any of the other terms included in Eq. (7) and, therefore, it is highly unreliable.

Table 1. ERR test results for the example system.

Term	ERR (%)
$r(t-3)r(t-3)p(t-3)$	39.0
$q(t-1)$	35.8
$u(t)$	12.2
$w(t-2)$	10.3
$q(t-2)q(t-2)q(t-2)$	2.58

This example demonstrates the power of using the NARMAX ERR data analysis technique over more simple techniques such as the correlation function. The ERR identified all the terms in Eq. (7), while the correlation function could only obtain one of the terms in Eq. (7) out of the terms with the highest five correlations. This emphasises that for nonlinear systems, only methods that are designed to account for nonlinearities should be applied, otherwise the results can be misleading.

4   The Dst index

The Dst (disturbance storm time) index is widely employed for studying the disturbances associated with geomagnetic storms and many attempts at modelling the dynamics of the Dst index have been made. The magnetosphere system, including the Dst index, is known to be a low dimensional system (Sharma, 1995; Valdivia et al., 1996; Klimas et al., 1996) and evolve under the influence of the solar wind. However, the question "*what combination of solar wind parameters control the evolution of the Dst index?*" still has no definitive answer, despite the quest for a solar wind–magnetosphere coupling function being the subject of many studies. One of the first attempts to model the Dst index was by Burton et al. (1975), where they used two inputs, the solar wind velocity V multiplied by the southward IMF (interplanetary magnetic field) B_s ($B_s = 0$ for $B_z \geq 0$ and $B_s = -B_z$ for $B_z < 0$) and the square root of the solar wind dynamic pressure p. The aim of Perreault and Akasofu (1978) was to find a solar wind–magnetosphere coupling function by estimating the interplanetary flux in terms of the Poynting flux, VB^2. An important observation in this study was that the they found evidence for small geomagnetic activity even when the IMF was orientated northward. As such, to account for the importance of the IMF orientation, instead of employing a rectifier that allows only negative values of B_z (Burton et al., 1975), they used a function of the IMF clock angle $\sin^4(\theta/2)$ where $\theta = \tan^{-1}(B_y/B_z)$. Therefore, the resulting coupling function was $VB^2 \sin^4(\theta/2)$. Kan and Lee (1979) justified the clock angle function analytically by deriving the power delivered by the solar wind from the field line reconnection geometry. There are many other coupling functions that have been derived, using different methods for obtaining a coupling function, such as correlation (Newell et al., 2007) or

Table 2. Correlation test results for the example system.

Term	Correlation (%)
$v(t-4)q(t-1)$	60.5
$v(t-3)q(t-1)$	60.4
$q(t-1)$	60.4
$v(t-1)q(t-1)$	60.3
$v(t-2)q(t-1)$	60.2

Table 3. Solar wind-Dst index coupling functions assembled by the ERR algorithm.

Coupling Function	ERR (%)
$p^{1/2}V^{4/3}B_T\sin^6(\theta/2)(t-1)$	5.46
$p^{1/2}V^2 B_T\sin^6(\theta/2)(t-1)$	3.18
$n^{1/6}V^2 B_T\sin^4(\theta/2)(t-1)$	3.15
$D_{st}(t-2)$	2.96
$p^{1/2}V B_T\sin^6(\theta/2)(t-1)$	2.77

trial and error (Temerin and Li, 2006), which can be found in the study by Boynton et al. (2011b).

The physical interpretability of the NARMAX algorithm has been used in the past to study the Dst index. A NARMAX model was derived using an input of VB_s in the study by Boaghe et al. (2001). Then by mapping this model into the frequency domain to produce a generalised frequency response function, the dominant nonlinear characteristics were studied, revealing the existence of energy storage processes that involve multi-wave coupling. A similar study was performed by Balikhin et al. (2001), which focused on the processes of energy loading for the Dst index. They concluded that there was no evidence for models that assume a time delay storage of energy. However, these studies never used the NARMAX algorithm to combine solar wind parameters into a solar wind coupling function and instead used VB_s as the sole input. Boynton et al. (2011b) employed the NARMAX ERR algorithm ability to search through and assess many combinations of solar wind parameters to obtain the most appropriate solar wind–Dst index coupling function.

4.1 NARMAX ERR derived solar wind–Dst coupling function

The aim of the study by Boynton et al. (2011b) was to derive a solar wind–Dst index coupling function that could be used as an input to model the Dst index. To do this, they utilised the structure detection stage of the NARMAX algorithm to combine solar wind parameters and find the most appropriate function with the highest ERR. As with the example from Sect. 3, there are many possible solar wind parameters that can influence the Dst index. Therefore, Boynton et al. (2011b) used a wide range of solar wind parameters as inputs. These inputs ranged from basic parameters, such as V, p, density n, IMF components B_x, B_y, B_z and the tangential IMF $B_T = \sqrt{B_y^2 + B_z^2}$, to nonlinear functions of the parameters, like $V^{4/3}$, $p^{1/2}$, $n_{1/6}$, B_s, $\sin^4(\theta/2)$ and $\sin^6(\theta/2)$. Due to the large number of parameters, four ERR analysis tests were carried out to narrow down what parameters had the most control over the Dst index. Table 3 displays Table 4 from Boynton et al. (2011b), where they used a fourth degree of nonlinearity, 5 time lags and inputs: V, $V^{4/3}$, $p^{1/2}$, $n^{1/6}$, B_s, B_T, $\sin^4(\theta/2)$ and $\sin^6(\theta/2)$.

The results from the table show that the coupling function should consist of density (given that $p = \frac{1}{2}nV^2$), velocity, tangential IMF and clock angle function, since these parameters appear in four of the top five functions with the highest ERR. Therefore, according to the results of Boynton et al. (2011b), the most appropriate coupling functions should be of the form:

$$n^\alpha V^\beta B_T^\gamma \sin^\delta\left(\frac{\theta}{2}\right). \tag{8}$$

From their results, they concluded that α should have a value between 1/6 and 1/2, γ should be equal to 1 and δ equal to 6. The value for β is the most inconclusive but should be in the range of 2–3.

In this study, Boynton et al. (2011b) analysed a number of clock angle functions. These included the purely southward component from B_s; $\sin^4(\theta/2)$, which was pioneered by Perreault and Akasofu (1978) and justified by Kan and Lee (1979); and $\sin^6(\theta/2)$. These functions are very similar and only significantly differ when the clock angle is directed east or west. One of the most interesting results of this study was that the $\sin^6(\theta/2)$ function was continuously selected by the algorithm as the most appropriate function for explaining the dependent variable variance of the Dst index, throughout each of the ERR analysis tests.

4.2 Analytical explanation for the coupling function

Since the results of NARMAX can be reverse engineered to gain physical understanding about the system, the coupling function by Boynton et al. (2011b) should be related to the interaction between the solar wind and the magnetosphere. One of the main conclusions of Boynton et al. (2011b) was that $\sin^6(\theta/2)$ was the most appropriate function for the IMF clock angle. The $\sin^6(\theta/2)$ IMF clock angle function goes against what is seen in most studies, where either $\sin^4(\theta/2)$ or the southward component were employed. Burton et al. (1975) empirically deduced the southward component of the IMF from scatter plots of the dawn to dusk component of the electric field against the ring current injection rate. They found that positive electric fields had a linear relationship with injection rate, which correspond to a southward IMF. While for negative dawn–dusk electric fields, which was analogous to a northward IMF, the injection rate was close

to zero. Therefore, they concluded that the southward component was the function of the clock angle. For the $\sin^4(\theta/2)$ function, Perreault and Akasofu (1978) fitted a function that could account for the small amount of geomagnetic activity observed when the IMF was slightly positive. This was then analytically derived by Kan and Lee (1979) from the geometric relationship between the electric and magnetic fields. The motivation for the study by Balikhin et al. (2010) was to understand why the ERR analysis resulted in $\sin^6(\theta/2)$ when most other studies and models preferred to use the southward component or $\sin^4(\theta/2)$ (Amariutei and Ganushkina, 2012; Boaghe et al., 2001; Akasofu, 1979).

Balikhin et al. (2010) revisited the arguments by Kan and Lee (1979) that deduced the $\sin^4(\theta/2)$ factor from first principles to determine why the results of Boynton et al. (2011b) were different. Like Kan and Lee (1979), Balikhin et al. (2010) started from the dayside reconnection electric field derived by Sonnerup (1974):

$$E_r = V_{MS} B_{MS} \sin\left(\frac{\theta}{2}\right),\qquad(9)$$

where the subscript MS indicates the magnetosheath velocity and magnetic field values. The reconnection electric field is assumed to be the only component of the magnetosheath electric field that is able to penetrate into the magnetosphere. The potential difference, Φ_M, across the polar cap can then be calculated from the perpendicular reconnection electric field:

$$E_{r\perp} = E_r \sin\left(\frac{\theta}{2}\right) = V_{MS} B_{MS} \sin^2\left(\frac{\theta}{2}\right)\qquad(10)$$

multiplied by the length of the X-line l_0, which is assumed to be constant, projected along the electric field. In Fig. 1, the length of the X-line projected along the electric field is the line $x_3 x_2$, which will be $l_0 \sin(\theta/2)$. Therefore, the cross-polar cap potential:

$$\Phi_M = V_{MS} B_{MS} \sin^2\left(\frac{\theta}{2}\right) l_0 \sin\left(\frac{\theta}{2}\right),\qquad(11)$$

$$\Phi_M = V_{MS} B_{MS} \sin^3\left(\frac{\theta}{2}\right) l_0 .\qquad(12)$$

The total power produced by the solar wind dynamo was then obtained by the square of the cross-polar cap potential divided by the resistance, R, assuming magnetic flux conservation so that $V_{MS} B_{MS} = V B$:

$$P = \frac{\Phi_m^2}{R} = \frac{V^2 B^2}{R} \sin^6(\theta/2) l_0 ,\qquad(13)$$

thus resulting in a theoretical explanation of the NARMAX results by Boynton et al. (2011b), which yielded the $\sin^6(\theta/2)$ factor as the most appropriate clock angle factor.

Equations (12) and (13) differ from the Kan and Lee equations for the potential and power. When Kan and Lee (1979) calculated the cross-polar cap potential they failed to account
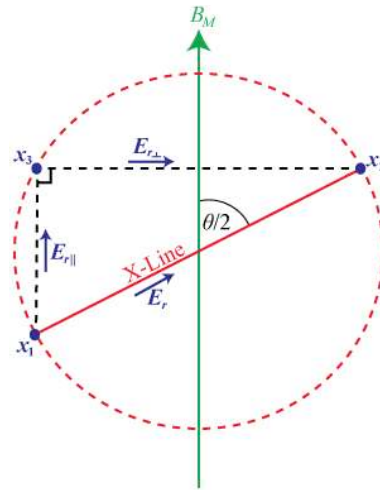

Fig. 1. The components of the reconnection electric field, where the line $x_1 x_2$ is the length of the X line, l_0.

for the fact that the potential should be calculated over the length in which the electric field is projected. In their calculation, they multiplied the perpendicular reconnection electric field by the entire length of the X line, line $x_1 x_2$ in Fig. 1. Consequently, their expression for the cross-polar cap potential missed a factor of $\sin(\theta/2)$ and their expression for the power, which resulted in $\sin^4(\theta/2)$, is also incorrect. Therefore, the application of the NARMAX ERR analysis found the correct solution and thus allowed for the amendment of a mistake made in the method by Kan and Lee (1979).

4.3 NARMAX Dst Model

Using the coupling function with the highest ERR in Table 3, Boynton et al. (2011a) derived a model of the Dst index that could estimate the following hours value. They analysed the model's performance, using data from the start of 1998 to the end of 2008, with three criteria: the correlation coefficient, the normalised root mean square error (NRMSE) and the coherency function. The model estimated Dst was shown to have a high correlation and a low NRMSE, however, their objectives were to identify a model that could forecast the onset, magnitude and duration of magnetic storms. They used the coherency function to illustrate how well the model achieved these goals, since it is able to determine the frequency dependencies between the measured and estimated Dst. The figures displayed that the model had a high coherency for the frequencies of a magnetic storm but did not perform as well for the higher frequencies. Boynton et al. (2011a) then compared the performance of their model to other Dst models that used a similar criteria, illustrating that the model using the NARMAX ERR derived coupling function had a higher correlation than the models employing $V B_s$ as the input. Figure 2 shows the model predicted output in blue and the measured Dst index in red for the period between March and May 2000.

4.4 Summary

The application of the NARMAX ERR approach to the Dst index has proved to be very successful in the studies by Boynton et al. (2011b), Balikhin et al. (2010) and Boynton et al. (2011a). In summary, Boynton et al. (2011b) was able to automatically derive a combination of solar wind parameters to form a coupling function by utilising the structure detection stage of the NARMAX ERR algorithm. This coupling function was then justified from first principles by Balikhin et al. (2010), where they derived the relationship of the solar wind power from the reconnection geometry. Finally, the NARMAX deduced coupling function was shown to give a better model performance than the commonly used VB_s function.

5 Electron fluxes at GEO

The radiation belts are a very hazardous environment for satellites and humans that transit the region. High relativistic electron fluxes within the radiation belts significantly increase the probability of detrimental effects to the onboard satellite systems and can even lead to permanent hardware damage. As such, the study of radiation belts is highly important for modern technological systems that require satellites. Although the radiation belts were discovered by very first in situ measurements (Van Allen, 1959), due to their complexity, we are not able to deduce the mathematical model from first principles with our current level of knowledge. The mechanisms behind the acceleration and loss of energetic particles need to be understood in order to have a complete model of the radiation belts. At present, there are two main theories on acceleration. One based on radial diffusion (Falthammar, 1968; Schulz and Lanzerotti, 1974), where due to the earthward diffusion of an initial seed population, the particles are accelerated by the conservation of the first and the second adiabatic invariants. The second theory is local diffusion (Temerin et al., 1994; Reeves et al., 2009), where particles are accelerated by interacting with waves within the radiation belt (e.g. chorus, magnetosonic, etc). The losses of particles within the radiation belts can be caused by magnetopause shadowing (Onsager et al., 2007; Ohtani et al., 2009; Matsumura et al., 2011), where the magnetopause is compressed to within the radiation belts, and can also be attributed to waves that cause losses (Loto'aniu et al., 2010).

Numerous studies have focused on obtaining the solar wind parameters that cause the acceleration and loss of the energetic particles within the radiation belts. Paulikas and Blake (1979) compared the daily averaged, 27 day averaged and 6 months averaged > 0.7, > 1.55 and > 3.9 MeV electron fluxes at geostationary Earth orbit (GEO) with the solar wind velocity, IMF components and sector polarity. They found that the solar wind velocity exhibited a strong correlation for all the energy ranges studied. Recently, these
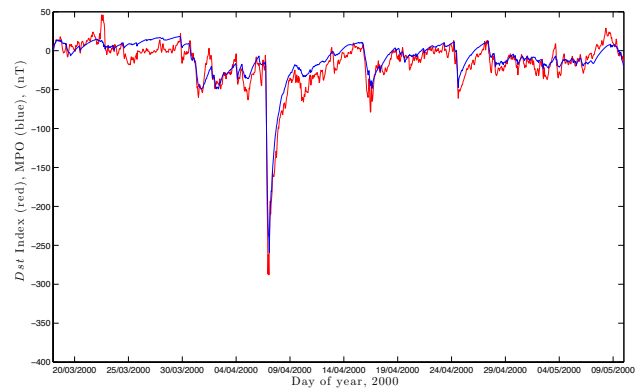

Fig. 2. The model predicted output in blue and the measured Dst index in red between March and May 2000.

results were revisited by Reeves et al. (2011). They analysed the long-term relationship between electron fluxes at GEO and solar wind velocity with the aid of scatter plots. These showed a much more complex relationship than the one suggested by Paulikas and Blake (1979), where, instead, the fluxes exhibited a triangular distribution with the velocity. On average the higher fluxes are a result of higher velocities and show a velocity dependant lower limit, but have an upper limit that is autonomous of the velocity. This complex triangular relationship between the electron flux and velocity motivated Boynton et al. (2013) and Balikhin et al. (2011) to investigate the solar wind parameters that control the evolution of electron fluxes at GEO using the ERR analysis.

5.1 ERR Analysis of electron fluxes at GEO

Boynton et al. (2013) employed the structure detection stage of the NARMAX algorithm to determine the solar wind parameters that control 14 different energies of the electron flux at GEO, ranging from 24.1 keV to 3.5 MeV. Similar to Boynton et al. (2011b), many different solar wind parameters were used as inputs to the algorithm, since it is not fully known what parameters influence the fluxes. These parameters included the solar wind velocity, density and pressure, north–south IMF component and values based on the daily variation of the north–south IMF component; these were the fraction of time in each day that the IMF had a southward orientation, the average southward IMF (B_s) within each day and the variance of B_z for each day. Table 4 displays the results from Boynton et al. (2013), employing a NARMAX algorithm that used a second degree nonlinearity and 5 time lags.

There are two interesting results from the analysis by Boynton et al. (2013). The first is that the solar wind density accounts for the majority of the variance for the energy range between 1.8 and 3.5 MeV and has an increasing influence on the fluxes from 925 keV. The other result is that as the energy of the electron flux increases, the time for the solar wind velocity to have an influence on the flux also increases. For 24.1–90 keV, the current day's velocity has the

Table 4. Results of the NARMAX analysis, showing top 3 terms in the order of ERR for the electron fluxes ranging from 24.1 keV to 3.5 MeV.

Energy	1st Term	ERR(%)	2nd Term	ERR(%)	3rd Term	ERR (%)
24.1 keV	$V(t)$	96.9	$V^2(t)$	2.82	$n(t)$	0.08
31.7 keV	$V(t)$	96.9	$V^2(t)$	2.83	$n(t)$	0.07
41.6 keV	$V(t)$	97.0	$V^2(t)$	2.82	$n(t)$	0.05
62.5 keV	$V(t)$	97.0	$V^2(t)$	2.80	$n(t)$	0.04
90.0 keV	$V(t)$	97.0	$V^2(t)$	2.77	$nV(t)$	0.03
127.5 keV	$V(t)$	74.8	$V(t-1)$	22.3	$V^2(t)$	2.08
172.5 keV	$V(t-1)$	65.7	$V(t)$	31.6	$V^2(t-1)$	1.74
270 keV	$V(t-1)$	97.4	$V^2(t)$	2.34	$B_z(t-1)$	0.02
407.5 keV	$V(t-1)$	84.1	$V(t-2)$	13.7	$V^2(t-1)$	1.63
625 keV	$V(t-1)$	75.9	$V(t-2)$	22.3	$V^2(t-2)$	0.61
925 keV	$V(t-2)$	96.2	$n(t)$	0.28	$V(t-4)$	0.24
1.3 MeV	$V^2(t-2)$	76.5	$nV(t-1)$	2.21	$n(t)V(t)$	1.90
2.0 MeV	$n(t-1)$	53.7	$nV(t-1)$	13.6	$n^2(t-1)$	5.55
1.8–3.5 MeV	$n(t-1)$	51.5	$n^2(t-1)$	15.1	$V^2(t-2)$	6.13

most influence on the electron flux, but at 127.5 keV the velocity of the previous day starts to effect the fluxes, having an ERR of 22 %. The ERR for the previous days velocity increases to 66 % for the higher energy of 172.5 keV electrons. This trend continues to 1.3 MeV electron fluxes, where the velocity recorded two days in the past is the controlling term.

5.2 Solar wind density

The relationship between the solar wind density, solar wind velocity and 1.8–3.5 MeV electron flux was investigated by Balikhin et al. (2011) to explain why the NARMAX ERR analysis resulted in the density having the most influence on the flux and not the velocity. They started by illustrating the relationship simply, via scatter plots of the density and velocity and showed that the high electron fluxes, above $10^{0.5}$ (cm^2 s sr keV)$^{-1}$, only occurred at at low densities, irrespective of the velocity value. Balikhin et al. (2011) then split scatter plots of velocity and electron flux into to six density ranges to examine how the distribution changed as the density is altered. They found that, for a fixed density, the electron flux increases with velocity until saturation, where the electron flux attains its maximum value. The velocity at which the saturation takes place and the maximum value of the flux decreases with increasing density. They concluded that the reason for the anti-correlation between density and electron flux could be because the growth rates of waves in local-wave particle interactions are effected by increases in density, thus, interfering with the acceleration of elections or causing the electrons to precipitate. Aryan et al. (2013) estimate the saturation velocity at different densities less than 6 cm^{-3} statistically by using the reverse arrangement test. They showed that there is a distinct anti-correlation between the saturation velocity of the electrons at GEO and solar wind density.

As mentioned by Aryan et al. (2013), a possible explanation for the density dependance could be magnetopause shadowing, since a high solar wind dynamic pressure, which is a function of density, can compress the dayside magnetopause to within the gyroradii of the electrons observed at GEO. Therefore, a high density could lead to the drift loss of electrons to the magnetopause. However, although the dynamic pressure was one of the inputs to the NARMAX algorithm, it was the density that had the highest ERR. So, why did the density have the highest ERR and not the pressure? Boynton et al. (2013) inspected the data to answer this question. They found a case where the electron flux decreased with no significant increase in pressure but a relatively large increase in density. Figure 3 has the same time period as Fig. 5 in the study by Boynton et al. (2013) and displays the daily averaged 1.8–3.5 MeV electron flux in the top panel a, 1 min solar wind velocity in panel b, 1 min solar wind density in panel c, the dynamic pressure in panel d and the magnetopause location, according to the model by Shue et al. (1997), in panel e with a black dashed line indicating GEO. Here, the electron flux data, from the Los Alamos National Laboratory (LANL) satellites, were only released in daily averaged format. However, since important information can be lost by daily averaging, the 1 min data is shown for the solar wind parameters and the estimated magnetopause position. For example, any magnetopause shadowing occurring within the day may be lost by averaging the data, thus indicating that no drift loss should occur, even though magnetopause shadowing is clearly shown in the 1 min data. An event in the electron flux can be seen in Fig. 3, with an increase of fluxes taking place between 7 and 12 November 2000, coinciding with a corotating interaction region, which can be seen by the increase of solar wind velocity for several days. After this, the fluxes plateau for 5 days before decreasing back to the initial levels.
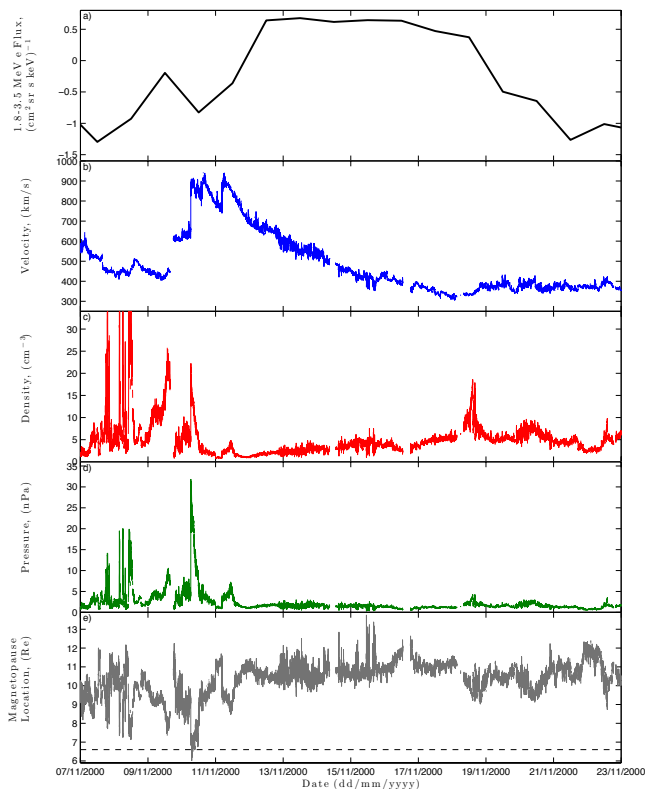
Fig. 3. The daily averaged 1.8–3.5 MeV electron flux **(a)**, 1 min solar wind velocity **(b)**, 1 min solar wind density **(c)**, the dynamic pressure **(d)** and the magnetopause location **(e)** according to the model by Shue et al. (1997). Starting on 7 November 2000 and ending on 23 November 2000, the same time period as Fig. 5 in the study by Boynton et al. (2013).

On 18 November 2000, Fig. 3 shows a steep decrease in electron fluxes. Meanwhile, the pressure increase is negligible in comparison to the increase on 10 November 2000 due to the lower solar wind velocity. However, the increase in density is large in comparison to the other increases in density. Also, according to the model of Shue et al. (1997) for the magnetopause location, during 18 November 2000 the magnetopause is always located beyond 9 R_E, well beyond the gyroradii of the electrons at GEO. With this evidence, Boynton et al. (2013) concluded that the loss of electrons is likely caused by density enhancement, at least in some cases, and that this could be due to the high densities resulting in waves that cause losses (Loto'aniu et al., 2010). It should be noted that during 10 November 2000, the Shue et al. (1997) model shows the magnetopause within GEO for a short period of the day, which corresponds to a decrease in flux.

These studies of the solar density influence on the electron flux at GEO illustrate how the NARMAX algorithm can indicate new paths of research by finding the significant parameters of a system. However, there is still much to understand in the relationship between the electron flux and solar wind

density, therefore, more in depth investigations into how the increases in density lead to the depletion of electrons at GEO are needed.

5.3 Solar wind velocity time lag

The second interesting result of the ERR analysis on the electron fluxes was that the time for the solar wind velocity to have an influence on the electron flux increased with the energy of the electrons. Although this had been observed before (Li et al., 2005), the NARMAX results allowed for the quantification of the lag vs. the energy. Balikhin et al. (2012) aimed to find the relationship between time lag and energy by solving the energy diffusion equation (Horne et al., 2005):

$$\frac{\partial F}{\partial t} = \frac{\partial}{\partial E}\left[A(E)D\frac{\partial}{\partial E}\left[\frac{F}{A(E)}\right]\right] - \frac{F}{\tau_L} \tag{14}$$

where F is a distribution function, t is the time, E is the kinetic energy, A is defined as

$$A = (E + E_0)(E + 2E_0)^{\frac{1}{2}}E^{\frac{1}{2}}, \tag{15}$$

D_{EE} is the bounce-averaged energy diffusion coefficient, τ_L is the effective timescale for losses to the atmosphere and E_0 is the rest energy of the electron. Horne et al. (2005) showed that the distribution function $F(E, \alpha_{eq})$ depends upon energy and the equatorial pitch angle, α_{eq}, and is related to the fluxes $J(E, \alpha_{eq})$ by

$$F(E, \alpha_{eq}) = \frac{E + E_0}{c(E + 2E_0)^{\frac{1}{2}}E^{\frac{1}{2}}}J(E, \alpha_{eq}). \tag{16}$$

Therefore, from Eqs. (14) and (16), Balikhin et al. (2012) estimated the upper limit of the timescale for the increase in electron flux as a function of energy. They assumed the energy diffusion coefficient D to be constant, the losses to be negligible ($\tau_L \to \infty$) and three cases for A: Case 1 $E \ll E_0$; Case 2 $E \approx E_0 \to E - E_0 \ll E_0$; and Case 3 $E \gg E_0$. Case 1 returns $A = E_0^{1/2}$; in Case 2, $A = E_0^0 = 1$; and for Case 3, $A = E^2$. For the second case, If $A = 1$, then the solution is the standard diffusion equation with constant coefficients; therefore, any changes in energy will be proportional to the square root of time. Thus, Balikhin et al. (2012) only solved for cases 1 and 3. For the sub-relativistic case (Case 1),

$$F = KE(t + t_0)^{-5/4}\exp\left(-\frac{E^2}{4DE_0^2(t + t_0)}\right); \tag{17}$$

while for the highly relativistic case (Case 3),

$$F = KE^2(t + t_0)^{-3/2}\exp\left(-\frac{E^2}{4DE_0^2(t + t_0)}\right), \tag{18}$$

where K is a constant and t_0 is the initial conditions for the time. Balikhin et al. (2012) noted that for these solutions to be valid, the seed population energies must be much lower than the energies being evaluated.
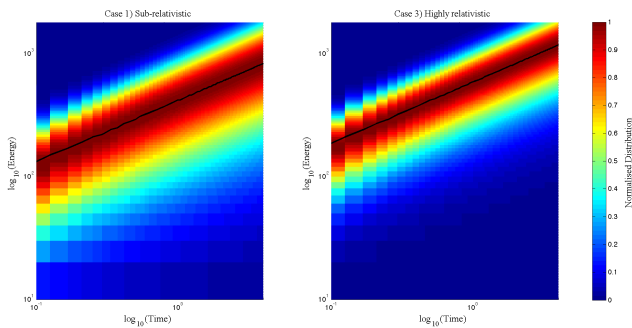
Fig. 4. Plots of the electron distribution with respect to energy and time from Eqs. (17) and (18), normalised between 0 and 1 for each time bin.

Fig. 5. Figure 1 from Boynton et al. (2013) displaying a log–log plot of the energy of the electron flux against the effective time delay of the solar wind velocity calculated from the NARMAX results. Also shown is the line of best fit in red.

Figure 4 is the log–log plot of the energy distributions for cases 1 and 3 as a function of energy and time, calculated from Eqs. (17) and (18). The distribution is normalised between 0 and 1 for each time bin to show the maximum of the distribution more clearly. As such, the gradients of these log–log energy distribution plots reveal the timescale for the increase in electron flux as a function of energy according to theory, which can then be compared to the lag vs. energy relationship found from the NARMAX results. In the sub-relativistic case the gradient is 0.4993, while it is 0.4996 in the highly relativistic case. Here, the gradients were found from the maximum of the distribution at each time, however, it is the same for all levels of the distribution function, including 10 % of the maximum that Balikhin et al. (2012) illustrated. Therefore, in all three cases of A, changes in the energy are proportional to the square root of time according to Eqs. (14) and (16).

However, the NARMAX results of Boynton et al. (2013), which revealed the statistical relationship between electron energy and velocity time lag does not concur with the solution of the energy diffusion equation. Figure 5 displays the relationship that was found from the NARMAX analysis, which can be compared to the gradients from the theory. The gradient from Fig. 5 is 1.05, which shows that the energy is proportional to the time delay not the square root. Accordingly, Balikhin et al. (2012) concluded that the time scaling of the solution of the energy diffusion equation is not fast enough to explain the increase of fluxes at GEO. Therefore, a purely local diffusion acceleration does not happen at GEO and radial diffusion plays an equal or greater role in the acceleration of electrons. Therefore, the interpretation of the NARMAX results have helped in the understanding of the electron acceleration mechanisms at GEO.
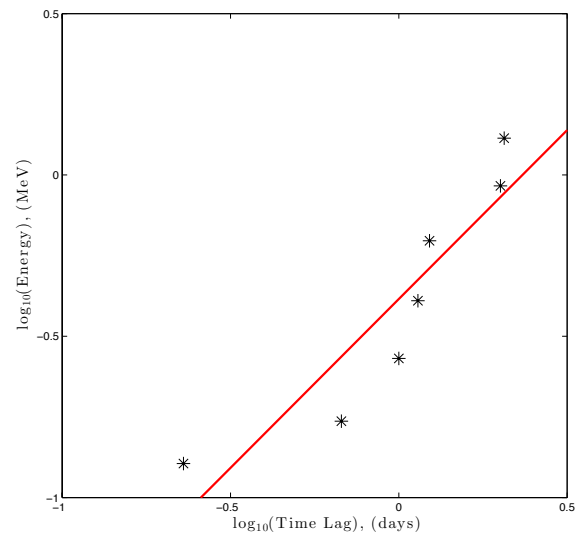
5.4 Summary

The NARMAX ERR analysis was applied to a range of electron flux energies by Boynton et al. (2013). From this analysis, there were two important results. The first was that the solar wind density had a major role in controlling the 1.8–3.5 MeV electron flux and the second was the quantification of the relationship between the electron flux and the time delay of the velocity.

The density relationship was confirmed by both Balikhin et al. (2011) and Aryan et al. (2013), where they found a statistical anti-correlation between the velocity at which the electron flux saturates and the density. Also, Boynton et al. (2013) showed that the depletion of the electron flux was not due to the increase in density causing magnetopause shadowing and, therefore, the decease in flux was most likely because of high densities resulting in waves that cause losses.

The energy diffusion equation was solved by Balikhin et al. (2012) to investigate if the solution agreed with the relationship between the electron flux energy and velocity time lag obtained by NARMAX. They found that for the solution of Eqs. (17) and (18), the energy was proportional to the square root of the time, which disagreed with the observed results interpreted by NARMAX.

From these studies, two online NARMAX electron flux models have been created for energies > 800 keV and > 2 MeV. With the identification of the main model parameters, achieved by Boynton et al. (2013), a NARMAX model was deduced, which uses real time data from the Advanced Composition Explorer (ACE) spacecraft to provide a 24 h ahead forecast of the electron fluxes. These forecasts are available at http://www.ssg.group.shef.ac.uk/USSW/UOSSW.html.

6   Conclusions

The NARMAX system identification technique has been shown to not only provide excellent models but also reveal insight into the physical processes of the system due to the interpretability of the results. This paper has reviewed how the NARMAX has been applied to the terrestrial magnetosphere, showing examples of how this approach can aid in understanding the physics of the magnetosphere.

For the Dst index, the application of the NARMAX approach by Boynton et al. (2011b), automatically derived a solar wind–magnetosphere coupling function from data, This coupling function was justified analytically from the geometry of dayside reconnection in the study by Balikhin et al. (2010). These two studies show how the physically interpretability of NARMAX can aid in the understanding of dayside reconnection. As well as providing insight into the physics, Boynton et al. (2011a) derived a NARMAX model for the Dst index, which also evidenced the superiority of this coupling function over others as an input for the Dst index.

The study of the electron fluxes at GEO, using the NARMAX algorithm, by Boynton et al. (2013) revealed a relationship between the solar wind density and 1.8–3.5 MeV electron flux, and quantified the timescale for the increase in electron flux as a function of energy. Balikhin et al. (2011) and Aryan et al. (2013) confirmed the density had an anticorrelation with the velocity at which the electron flux saturates, while Boynton et al. (2013) showed that the loss of electrons, in some cases, is not due to the density increases causing magnetopause shadowing. Balikhin et al. (2012) solved the energy diffusion equation and found that a change in energy of the electrons should be proportional to the square root of the time taken for this change. However, they concluded that since this disagreed with the observations from the NARMAX analysis, then local diffusion cannot be dominant at GEO.

References

Akasofu, S.-I.: Relationship between the growth of the ring current and the interplanetary quantity ε, Planet. Space Sci., 27, 1033–1041, 1979.

Amariutei, O. A. and Ganushkina, N. Yu.: On the prediction of the auroral westward electrojet index, Ann. Geophys., 30, 841–847, doi:10.5194/angeo-30-841-2012, 2012.

Aryan, H., Boynton, R. J., and Walker, S. N.: Analysis of trends between solar wind velocity and energetic electron fluxes at geo-stationary orbit using the reverse arrangement test, J. Geophys. Res.-Space, 118, 636–641, 2013.

Balikhin, M. A., Boaghe, O. M., Billings, S. A., and Alleyne, H. S. C. K.: Terrestrial magnetosphere as a nonlinear resonator, Geophys. Res. Lett., 28, 1123–1126, 2001.

Balikhin, M. A., Boynton, R. J., Billings, S. A., Gedalin, M., Ganushkina, N., Coca, D., and Wei, H.: Data based quest for solar wind-magnetosphere coupling function, Geophys. Res. Lett., 37, L24107, doi:10.1029/2010GL045733, 2010.

Balikhin, M. A., Boynton, R. J., Walker, S. N., Borovsky, J. E., Billings, S. A., and Wei, H. L.: Using the narmax approach to model the evolution of energetic electrons fluxes at geostationary orbit, Geophys. Res. Lett., 38, L18105, doi:10.1029/2011GL048980, 2011.

Balikhin, M. A., Gedalin, M., Reeves, G. D., Boynton, R. J., and Billings, S. A.: 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, 2012.

Billings, S. A. and Voon, W. S. F.: Correlation based model validity tests for non-linear models, Int. Control J., 44, 235–244, 1986.

Billings, S. A. and Zhu, Q. M.: Nonlinear model validation using correlation tests, Int. Control J., 62, 749–766, 1989.

Boaghe, O. M., Balikhin, M. A., Billings, S. A., and Alleyne, H.: Identification of nonlinear processes in the magnetospheric dynamics and forecasting of dst index, J. Geophys. Res., 106, 30047–30066, 2001.

Boynton, R. J., Balikhin, M. A., Billings, S. A., Sharma, A. S., and Amariutei, O. A.: Data derived NARMAX Dst model, Ann. Geophys., 29, 965–971, doi:10.5194/angeo-29-965-2011, 2011a.

Boynton, R. J., Balikhin, M. A., Billings, S. A., Wei, H. L., and Ganushkina, N.: 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, 2011b.

Boynton, R. J., Balikhin, M. A., Billings, S. A., Reeves, G. D., Ganushkina, N., Gedalin, M., Amariutei, O. A., Borovsky, J. E., and Walker, S. N.: The analysis of electron fluxes at geosynchronous orbit employing a NARMAX approach, J. Geophys. Res.-Space, 118, 1500–1513, 2013.

Burton, R. K., McPherron, R. L., and Russull, C. T.: An empirical relationship between interplanetary conditions and dst, J. Geophys Res., 80, 4204–4214, 1975.

Falthammar, C.-G.: Radial Diffusion by Violation of the Third Adiabatic Invariant, Reinhold, New York, 1968.

Horne, R. B., Thorne, R. M., Glauert, S. A., Albert, J. M., Meredith, N. P., and Anderson, R. R.: Timescale for radiation belt electron acceleration by whistler mode chorus waves, J. Geophys. Res., 110, A03225, doi:10.1029/2004JA010811, 2005.

Kan, J. R. and Lee, L. C.: Energy coupling function and solar wind-magnetosphere dynamo, Geophys. Res. Lett., 6, 577–580, 1979.

Klimas, A. J., Vassiliadis, D., Baker, D. N., and Roberts, D. A.: The organized nonlinear dynamics of the magnetosphere, J. Geophys. Res., 101, 13089–13113, 1996.

Leontaritis, I. J. and Billings, S. A.: Input-output parametric models for non-linear systems part I: Deterministic non-linear systems, Int. J. Control, 41, 303–328, 1985a.

Leontaritis, I. J. and Billings, S. A.: Input-output parametric models for non-linear systems part II: Stochastic nonlinear systems, Int. J. Control, 41 329–344, 1985b.

Li, X. L., Baker, D. N., Temerin, M., Reeves, G., Friedel, R., and Shen, C.: Energetic electrons, 50 keV to 6 MeV, at geosynchronous orbit: Their responses to solar wind variations, 3, S04001, doi:10.1029/2004SW000105, 2005.

Loto'aniu, T. M., Singer, H. J., Waters, C. L., Angelopoulos, V., Mann, I. R., Elkington, S. R., and Bonnell, J. W.: Relativistic electron loss due to ultralow frequency waves and enhanced outward radial diffusion, J. Geophys. Res., 115, A12245, doi:10.1029/2010JA015755, 2010.

Matsumura, C., Miyoshi, Y., Seki, K., Saito, S., Angelopoulos, V., and Koller, J.: Outer radiation belt boundary location relative to the magnetopause: Implications for magnetopause shadowing, J. Geophys. Res., 116, A06212, doi:10.1029/2011JA016575, 2011.

McCulloch, W. and Pitts, W.: A logical calculus of ideas imminent in nervous activity, B. Math. Biophys., 5, 115–133, 1943.

Newell, P. T., Sotirelis, T., Liou, K., Meng, C. I., and Rich, F. J.: A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables, J. Geophys. Res., 112, A01206, doi:10.1029/2006JA012015, 2007.

Ohtani, S., Miyoshi, Y., Singer, H. J., and Weygand, J. M.: On the loss of relativistic electrons at geosynchronous altitude: Its dependence on magnetic configurations and external conditions, J. Geophys. Res., 114, A01202, doi:10.1029/2008JA013391, 2009.

Onsager, T. G., Green, J. C., Reeves, G. D., and Singer, H. J.: Solar wind and magnetospheric conditions leading to the abrupt loss of outer radiation belt electrons, J. Geophys. Res., 112, A01202, doi:10.1029/2006JA011708, 2007.

Paulikas, G. A. and Blake, J. B.: Effects of the solar wind on magnetospheric dynamics: Energetic electrons at the synchronous orbit, Quantitative Modeling of Magnetospheric Processes, Geophys. Monogr. Ser., AGU, Washington, DC, 21, 180–202, 1979.

Perreault, W. K. and Akasofu, S.-I.: A study of geomagnetic storms, Geophys. J. Roy. Astr. S., 54, 547–573, 1978.

Reeves, G. D., Chan, A., and Rodger, C.: New directions for radiation belt research, Space Weather, 7, S07004, doi:10.1029/2008SW000436, 2009.

Reeves, G. D., Morley, S. K., Friedel, R. H. W., Henderson, M. G., Cayton, T. E., Cunningham, G., Blake, J. B., Christensen, R. A., and Thomsen, D.: On the relationship between relativistic electron flux and solar wind velocity: Paulikas and blake revisited, J. Geophys. Res., 116, A02213, doi:10.1029/2010JA015735, 2011.

Rumelhart, D. E. and MacClelland, J. L.: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, MIT Press, Cambridge, MA, 1986.

Schulz, M. and Lanzerotti, L. J.: Particle diffusion in the radiation belts, Physics and Chemistry in Space, Berlin: Springer, 1974.

Sharma, A. S.: Assessing the magnetospheres nonlinear behavior – its dimension is low, its predicability, Rev. Geophys., 33, 645–650, 1995.

Shue, J.-H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., and Singer, H. J.: A new functional form to study the solar wind control of the magnetopause size and shape, J. Geophys. Res., 102, 9497–9511, doi:10.1029/97JA00196, 1997.

Sonnerup, B. U. O.: Magnetopause reconnection rate, J. Geophys. Res., 79, 1546, doi:10.1029/JA079i010p01546, 1974.

Temerin, M. and Li, X.: Dst model for 1995–2002, J. Geophys. Res., 111, A04221, doi:10.1029/2005JA011257, 2006.

Temerin, M., Roth, I., Hudson, M. K., and Wygant, J. R.: New paradigm for the transport and energization of radiation belt particles, Eos T. Am. Geophys. Un., 7544, Fall Meet. Suppl, p. 538, 1994.

Valdivia, J. A., Sharma, A. S., and Papadopoulos, K.: Prediction of magnetic storms by nonlinear models, Geophys. Res. Lett., 23, 2899–2902, doi:10.1029/96GL02828, 1996.

Van Allen, J. A.: The geomagnetically trapped corpuscular radiation, J. Geophys. Res., 64, 1683–1689, doi:10.1029/JZ064i011p01683, 1959.