

# Adaptive linear prediction of radiation belt electrons using the Kalman filter

E. J. Rigler, D. N. Baker, and R. S. Weigel

Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado, USA

D. Vassiliadis

Universities Space Research Association, Goddard Space Flight Center, NASA, Greenbelt, Maryland, USA

A. J. Klimas

Goddard Space Flight Center, Greenbelt, Maryland, USA

Received 9 October 2003; revised 30 October 2003; accepted 20 November 2003; published 12 March 2004.

[1] Prior studies have examined the time-stationary (and quasi-stationary) dynamic response of relativistic electrons in the Earth's outer radiation belt to changes in solar wind bulk speed using linear prediction filters [Baker *et al.*, 1990; Vassiliadis *et al.*, 2002]. For this study, we have implemented an adaptive system identification scheme, based on the Kalman filter with process noise, to determine optimal time-dependent electron response functions. The nonlinear dynamic response of the radiation belts can then be tracked in time by recursively updating the optimal linear filter coefficients as new observations become available. We demonstrate a significant improvement in zero-time-lag electron log-flux "predictions" relative to models that are based on time-stationary linear prediction filters, while incurring only a slight increase in computational complexity. We conclude by discussing modifications necessary for an operational specification and forecast model, including the assimilation of real-time data, more sophisticated model structures, and a more practical gridded description of the radiation belt state. **INDEX TERMS:** 2784 Magnetospheric Physics: Solar wind/magnetosphere interactions; 2720 Magnetospheric Physics: Energetic particles, trapped; 2722 Magnetospheric Physics: Forecasting; **KEYWORDS:** Kalman filter, electron radiation belt, adaptive system identification

**Citation:** Rigler, E. J., D. N. Baker, R. S. Weigel, D. Vassiliadis, and A. J. Klimas (2004), Adaptive linear prediction of radiation belt electrons using the Kalman filter, *Space Weather*, 2, S03003, doi:10.1029/2003SW000036.

## 1. Introduction

### 1.1. Linear Filters

[2] Linear prediction filters have been used to help understand dynamic coupling between the solar wind and the geospace environment for well over 20 years. Arnoldy [1971] first noted the strong linear relationship between the  $B_z$  component of the interplanetary magnetic field and the AE index, a measure of auroral electrojet activity. Iyemori *et al.* [1979] were among the first to calculate discrete linear impulse responses of several geomagnetic indices to  $B_z$ . A number of similar studies followed in the early-to-middle 1980s using various different inputs and outputs [e.g., Clauer *et al.*, 1981; Clauer *et al.*, 1983; McPherron *et al.*, 1984; Bargatze *et al.*, 1985]. Several excellent review articles can be found in Kamide and Slavin [1986].

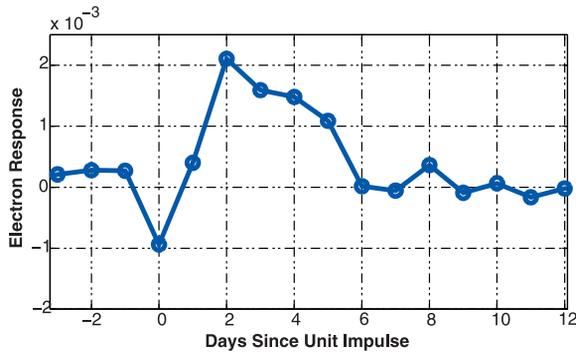
[3] The linear prediction filters used for these studies were typically comprised of simple vectors of linear coefficients that were convolved with an input time series to provide a zero-lag prediction (a "nowcast") of the linear response in a given output time series. This type of model can be thought of as a finite impulse response (FIR) function, since the response becomes zero as soon as there is no longer input to be filtered. This is represented

numerically in equation (1), where  $F$  represents the filter,  $u$  represents the input data, and  $\hat{y}$  is the estimated output.

$$\hat{y}_t = \sum_{j=p}^M F_j u_{t-j} \quad (1)$$

[4] Because of increased operational spacecraft use, a strong interest in practical radiation belt electron predictions developed in the middle-to-late 1980s. Thus, not surprisingly, linear prediction filters were one of the first techniques applied. Reasonably accurate electron flux predictions for geostationary altitudes were made using both causal (i.e.,  $p \geq 0$  [Nagai, 1988]) and acausal [Baker *et al.*, 1990] linear filters designed to operate on several different types of input data streams. Representative filter coefficients describing the discrete impulse response of relativistic radiation belt electrons at geostationary altitudes to changes in solar wind speed are shown in Figure 1.

[5] A more recent analysis by Vassiliadis *et al.* [2002] used observations of 2–6 MeV electrons from the Solar, Anomalous, and Magnetospheric Particle Explorer (SAMPEX) to study the spatially distributed (in L-shell) radiation belt response to changes in solar wind speed. Figure 2 is a recreation of the response function profile originally pre-



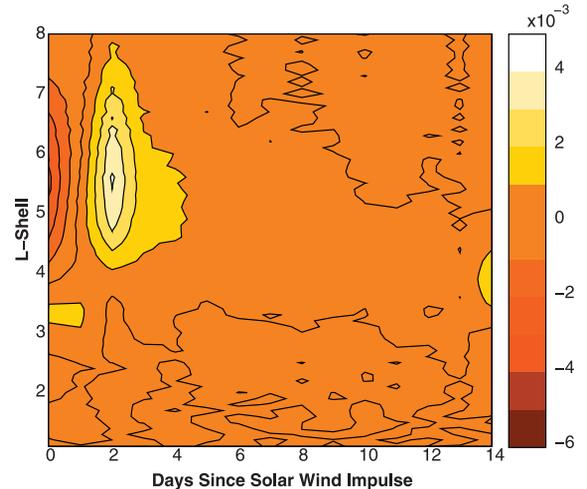
**Figure 1.** Linear prediction filter coefficients describing the time-invariant response of the log-flux of 2–6 MeV electrons at  $L = 6.6$  to a discrete unit impulse in the solar wind velocity.

sented by Vassiliadis et al., but uses SAMPEX electron data from  $L = 1.1$ – $8.0$  to calculate causal responses only. At least two spatially distinct responses are apparent. One, at a broad range of higher geomagnetic  $L$ -shells (i.e.,  $L \geq 4$ ), had a 1–2 day delay before it peaked, and was followed by an extended decay period. This was consistent with earlier studies based on geostationary data. The other response was observed near  $L = 3$ , occurred nearly simultaneously with changes in solar wind speed, and had little or no extended decay period. Significant solar cycle and seasonal variations in the response functions, as well as their prediction efficiencies, were also noted and discussed in terms of the well-known equinoctial and axial hypotheses [i.e., Boller and Stolov, 1970; Russell and McPherron, 1973; Cliver et al., 2000].

## 1.2. Adaptive System Identification

[6] All the previously mentioned studies employed some sort of batch process for system identification (parameter estimation). In other words, they accumulated a large regression matrix from all of the available input data, then inverted it to determine the filter coefficients that best predicted all of the observed output. The phasing of electron flux variations was predicted quite well with these time-invariant filters; however, the amplitudes, even for log-flux predictions, were often severely underestimated. This implied that even though solar wind speed was probably a significant driver of radiation belt dynamics the acceleration of relativistic electrons might be a fairly nonlinear process, all the relevant inputs were not being considered, or probably both.

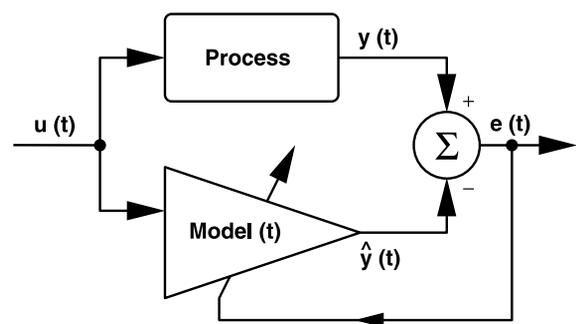
[7] The time variability of the response function profiles observed by Vassiliadis et al. [2002] suggested that an adaptive system identification technique might prove useful. Not only would it be able to systematically vary model parameters to track nonstationary dynamics in time, but the recursive nature of most adaptive algorithms is ideal for implementation in an on-line operational space weather model. Figure 3 illustrates graphically a generalized adaptive identification configuration for a deterministic model.



**Figure 2.**  $L$ -shell dependent, time-invariant impulse response profile of the log-flux of 2–6 MeV electrons.

[8] Most adaptive system identification techniques are closely related to formal data assimilation (DA) algorithms. One popular DA tool with extensive operational heritage is the Kalman filter (KF). This recursive algorithm provides the optimal estimate of a system’s state at particular time. It requires a baseline model and a statistical description of the uncertainty associated with both the model and observations that can be related to the state. What is not always appreciated about the Kalman filter, however, is that the state may include the parameters of the model itself. In fact, the state can be comprised solely of the model parameters. This provides an excellent adaptive system identification technique that might be applied to the relatively simple FIR prediction filters discussed previously.

[9] The primary purpose of the present paper is to (1) describe in some detail how the KF is used to adaptively



**Figure 3.** A general adaptive system identification configuration for a deterministic model. Input,  $u(t)$ , passes to both the physical system, or Process, and to a generalized deterministic model. The residuals ( $e(t) = y(t) - \hat{y}(t)$ ) are used to dynamically modify the model to better match new observations, but they do not drive the model directly (indicated by the arrow passing through the model block, not into it).

update the coefficients of simple linear prediction filters, (2) provide a brief demonstration of the predictive capability of the KF when it is configured to track nonlinear, non-time-stationary radiation belt dynamics, and (3) describe how real-time data and more sophisticated model structures might be employed to provide improved specification and forecasting of radiation belt electrons for use in space weather applications.

## 2. The Kalman Filter

[10] The Kalman filter is an approximate filter for nonlinear systems that is based on a first-order linearization of the state equation [Ljung, 1979]. It therefore requires that the system of interest can be described in a state-space configuration. Since we are presently only interested in the system identification aspect of the KF, our “state” will be comprised solely of the model parameters. Equation (2) illustrates this state-space configuration.

$$\begin{aligned}\theta_{t+1} &= \theta_t + v_t \\ y_t &= \phi_t^T \theta_t + e_t\end{aligned}\quad (2)$$

[11] For simplicity, and for consistency with the research described in section 1, the models used in this study are single input, single output (SISO) linear filters that describe the dynamic relationship between daily averaged solar wind speed and relativistic electron flux. Therefore  $\theta_t$ , the parameter vector, is simply a vector of linear filter coefficients designed to provide a zero-lag “prediction.”  $\phi_t$  is the regression vector, comprised of the time-lagged inputs to the model. The  $v_t$  is the process noise, which describes the error in the filter parameters as a function of time, and  $e_t$  is the observation error. Both  $v_t$  and  $e_t$  are usually considered independent white noise sequences. In an online (i.e., real-time) application, neither are known a priori, so it is necessary to represent them as stochastic processes that are defined by their respective covariance matrices ( $Q^v = E\{vv^T\}$  and  $Q^e = E\{ee^T\}$ ;  $E$  is the statistical expectation). The KF algorithm is then

$$\begin{aligned}\hat{\theta}_{t+1} &= \hat{\theta}_t + \mathbf{K}_t(y_t - \phi_t^T \hat{\theta}_t) \\ \mathbf{K}_t &= \mathbf{P}_t \phi_t (\phi_t^T \mathbf{P}_t \phi_t + Q^e)^{-1} \\ \mathbf{P}_{t+1} &= (\mathbf{I} - \mathbf{K}_t \phi_t^T) \mathbf{P}_t + Q^v\end{aligned}\quad (3)$$

[12] In other words, the best estimate of the model’s parameters for the next time step is the sum of the current model parameters and the current residuals multiplied by the Kalman gain,  $\mathbf{K}_t$ . The gain is updated using the state estimation covariance matrix,  $\mathbf{P}_t$ , and inversely scaled according to the uncertainty associated with the observations,  $Q^e$  (the larger the variance in the observation error, the less confidence we have in those observations and therefore the smaller the gain). Finally,  $\mathbf{P}_t$  is mapped to the next time step and should become smaller as more information is incorporated into the filter.

[13] The KF should remain stable so long as there is sufficient excitation (i.e.,  $\Delta\phi_t \neq 0$ ) in the input signal. If the input signal is actually nonexistent,  $\mathbf{P}_t$  will grow linearly according to  $\mathbf{P}_t + Q^v$ . Therefore prolonged periods without excitation may result in destabilization of the KF algorithm. On the other hand, if there is persistent excitation, but no process noise, the Kalman gain will approach zero in the limit  $t \rightarrow \infty$  as new observations are assimilated. This is only desirable if the true system dynamics are time-invariant, otherwise useful information may be lost if the system’s true nonlinear response and the model’s predictions diverge.

[14] The primary reason for this kind of divergence in a system identification scheme is a mismatch between the baseline model structure and the true system dynamics. It is therefore advisable to specify a finite, positive value for  $Q^v$ , thereby allowing the model parameters to vary with time as a sort of random walk. In other words, the process noise introduces a finite memory into the algorithm which allows the linear prediction filter to track nonstationary signals through time. This effective memory is somewhat ad hoc, but since  $Q^v$  is the variance of the noise in the model parameters, larger values generally imply faster tracking ability. Typically, values for  $Q^v$  are chosen by trial and error, where there is a trade-off between noise sensitivity and parameter tracking capability [Johansson, 1993].

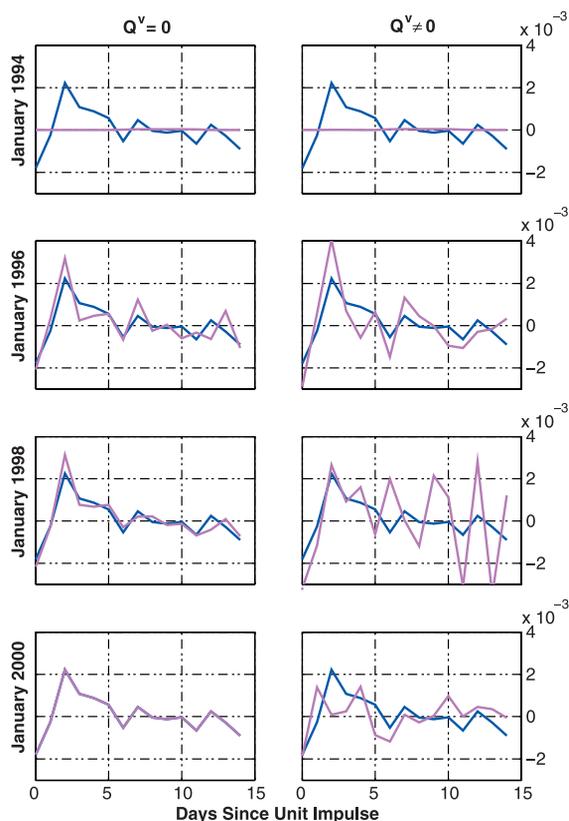
[15] Figure 4 illustrates graphically how the linear prediction filter coefficients evolve with time for two different KF configurations. The left panel of Figure 4 specifies a process noise to be zero, resulting in what amounts to a recursive linear least squares solution to the problem. The panel on the right in Figure 4 shows the KF solution when a fixed, finite process noise is allowed. Four snapshots were taken: one at the beginning (the initial guesses for  $\hat{\theta}$  were zero vectors), two in the middle, and one at the end of assimilating the training validation data set. The changing response functions illustrate (1) the convergence of the KF to a global solution, in the case where  $Q^v = 0$  and (2) the fact that when process noise is allowed, the linear filter coefficients vary with time, and are therefore able to track nonlinear dynamics that manifest as nonstationary processes in discrete time series.

## 3. Radiation Belt and Solar Wind Data

[16] It is important to compare and contrast the results attained in this study with those that came before. Therefore we chose to demonstrate the KF using the same data as Vassiliadis *et al.* [2002]. A brief description follows.

### 3.1. Relativistic Electrons

[17] The Solar, Anomalous, and Magnetospheric Particle Explorer (SAMPEX) satellite, launched in 1992, has a low-altitude ( $\sim 600$  km), high-inclination ( $\sim 82^\circ$ ) orbit, with a period of  $\sim 100$  min [Baker *et al.*, 1993]. As a result of its low altitude, it is possible to use the International Geomagnetic Reference Field (IGRF), an internal magnetic field model, to bin SAMPEX particle measurements according to a



**Figure 4.** The 2–6 MeV log-flux responses to impulses in solar wind speed determined using the Kalman filter with and without process noise (magenta). The fixed reference response (blue) is the global, or “batch,” solution obtained when all the data are included.

nominal geomagnetic L-shell. The ELO channel of the Proton Electron Telescope (PET) [Cook *et al.*, 1993] was used for this study, providing flux measurements for 2–6 MeV electrons (electrons/(cm<sup>2</sup> · sec · steradian)), a particle population with important space weather implications [Baker, 2001, 2002]. Strong diurnal variability is often observed in the ELO data, and is usually attributed to the rotation of the Earth’s imperfect dipole with respect to the relatively fixed orbit plane of the SAMPEX satellite. In order to avoid undesirable statistical effects resulting from strong auto-correlations associated with these diurnal variations, daily averages of electron flux observations were used. Finally, the log<sub>10</sub> of electron flux observations was used to reduce the asymmetry in the data’s probability distribution function (PDF), making it more amenable to linear estimation algorithms.

### 3.2. Solar Wind Speed

[18] The solar wind data used in this study was taken from the OMNIWeb database, a compilation of hourly solar wind magnetic field and plasma measurements, taken from a variety of different satellites over the last

several decades, and cross-normalized to provide multi-source uniformity (National Space Science Data Center, OMNIWeb, Near Earth Heliosphere Data, National Space Science Data Center, NASA, <http://nssdc.gsfc.nasa.gov/omniweb/>, 2003). We limited our current study to include only daily averaged solar wind bulk speed between 1994 and 1999, since the OMNI data density outside this range fell from an average of well over 80% to below 40%. The daily averaged solar wind speed PDF was not as asymmetric about its mean as the electron data and was therefore left as-is.

### 3.3. Data Gaps

[19] The SAMPEX radiation belt data set was fairly complete, however significant data gaps in the solar wind data existed throughout the time span of interest. This complicates the process of generating the regression vector for each iteration because only good data points may be included. The problem is exacerbated when dealing with filter-based models since they are comprised of a series of time delayed coefficients, which means a single bad datum in the input time series requires that a number of data points equal to the length of the filter must be dropped.

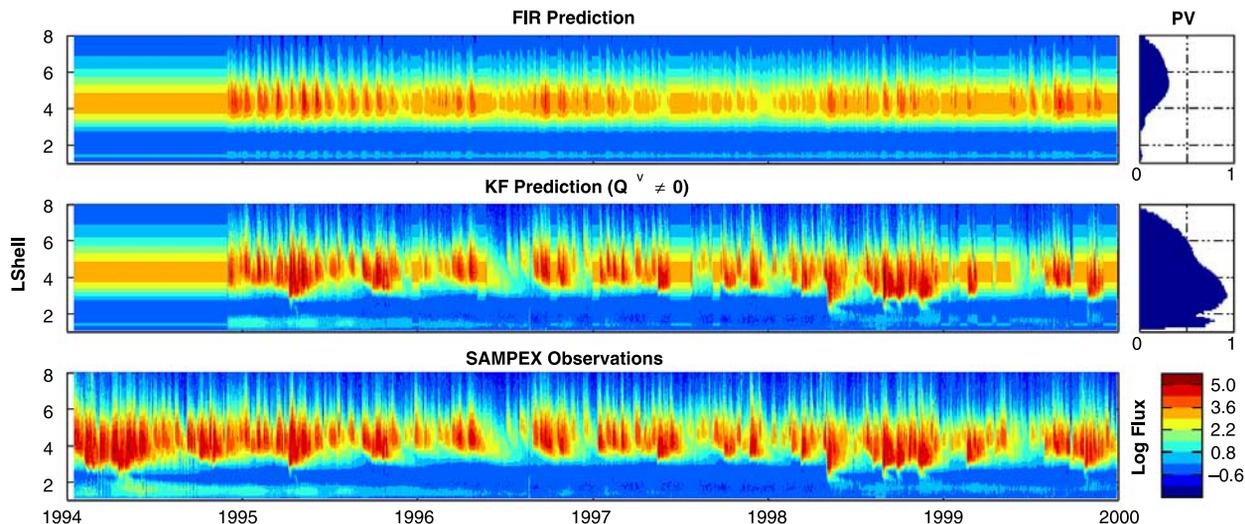
[20] The KF itself provides an excellent mechanism for dealing with this problem. The observation error covariance matrix,  $Q^e$ , is a measure of confidence in the observations that make up the training data. There is no reason that  $Q^e$  cannot be a function of time, and therefore specified very large for time steps corresponding to solar wind data gaps. This will result in the calculation of a very small Kalman gain, and effectively negate the influence of those particular observations on the parameter estimation.

[21] Once the optimal coefficients are determined, one must also take care to not pass bad data through the filter when making predictions. In order to keep the focus of this paper on the KF algorithm itself, we chose to apply a very simple interpolation technique to deal with this prediction problem. Time steps corresponding to data gaps in the solar wind input were simply filled with the average electron flux for that particular L-shell.

## 4. Model Performance

### 4.1. Prediction Quality

[22] The first and most straightforward test of appropriateness for any model is its ability to reproduce observed dynamics given relevant inputs. We determined a simple metric specifying the percent of observation variance (PV, an actual “prediction efficiency” only if the expected model error has zero mean) by subtracting the ratio of the variance of a model’s residuals to the variance of the observations from one ( $1 - (\sigma_{resid}^2 / \sigma_{obs}^2)$ ). This provides an effective skill score describing the model’s predictive ability relative to the time series average. Normally model validation is done using an “out of sample” data set. In the case where the KF is used to dynamically update linear prediction filter coefficients, however, it is necessary to use “in sample” data in order to drive the recursive identification algorithm. Therefore the same training/validation data must be used to calculate PV.



**Figure 5.** Static FIR and adaptive Kalman filter (KF) predictions of 2–6 MeV electron log flux for 1994 through 1999 are compared with observations. The percent of variance (PV) is plotted as a function of L-shell. Historical averages were used to substitute for model output during periods without reliable solar wind input (recognizable as horizontal contour lines in the predicted output), and were not included in the calculation of PV.

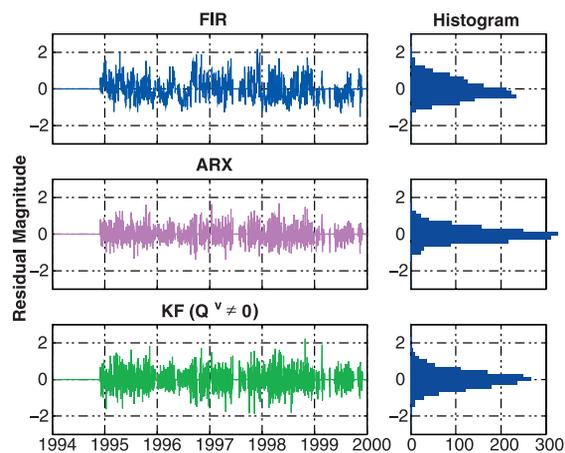
[23] Figure 5 compares SAMPEX observations from  $L = 1.1$  to  $L = 8.0$  with predicted output from both static and adaptive FIR linear prediction filters. The results from the KF with process noise show significant improvements in predictive ability over the time-invariant FIR linear filters across almost the entire range of L-shells. The improvements are particularly pronounced between  $L \approx 3$  and  $L \approx 5$ . It turns out, however, that a PV profile for simple persistence (i.e., today’s flux will be the same as yesterday’s) is very similar to the profile shown for our adaptive model. In point of fact, persistence is implicitly included in the KF with process noise. One might argue that a simpler static model which explicitly accounts for persistence might be used. A more sophisticated approach to residual analysis is necessary to determine which model is most appropriate.

#### 4.2. Residual Analysis

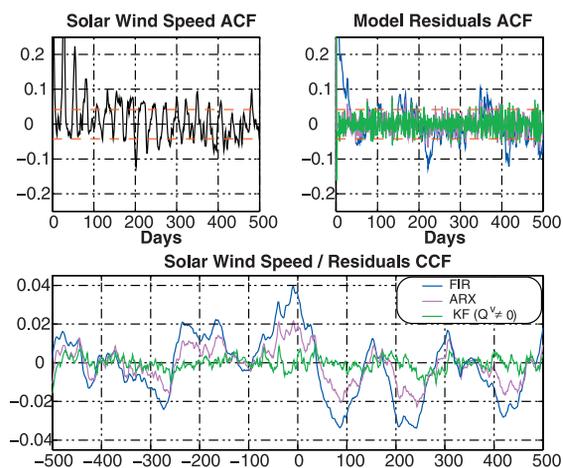
[24] Neither the model nor its training validation data are ever perfect (i.e., residuals  $\neq 0$ ). If the model is chosen wisely, however, there should be little or zero correlation or recurrent structure in the residuals. A simple test for determining the “whiteness” of the model residuals is to plot a histogram of the residual time series. This error distribution function should have a zero-mean, Gaussian profile. If the mean deviates significantly from zero, or the distribution function is significantly skewed to one side or the other, this is an immediate indication that the model is biased, and is unlikely to properly exploit the dynamical information available in the training validation signals.

[25] Figure 6 shows the residuals and corresponding distribution functions for three different models at  $L = 6.6$ : 1) a time-invariant Finite Impulse Response (FIR)

linear prediction filter; 2) a time-invariant Auto-Regressive (with exogenous input, ARX) linear model that filters both the solar wind input and time-lagged flux observations [see *Vassiliadis et al., 2000; Vassiliadis, 2000*, and references therein for a thorough description of this type of dynamical model]; and 3) our adaptive FIR model based on the



**Figure 6.** Residuals for three different models predicting 2–6 MeV electrons at geostationary altitude and their associated histograms. Note the times when the residuals appear to be nonexistent (early 1994, and subsequent smaller gaps). These correspond to gaps in the solar wind data and are not indicative of increased model accuracy.



**Figure 7.** Normalized autocorrelation functions (ACF) and cross-correlation functions (CCF, smoothed with 27-day moving average “boxcar” filter) for model residuals at geostationary altitude. Colors correspond to the residuals shown in Figure 6. The 95% confidence interval for a  $N(0, 1)$  distribution is designated by the dashed red lines.

KF with process noise. The first and third panels correspond to the predictions and data presented previously, just at  $L = 6.6$ . The ARX residuals in the second panel are presented in order to compare the KF output with a static model that accounts for persistence.

[26] If the system being modeled is linear and deterministic, the residuals (in units of  $\log_{10}$ -Flux) from the time-invariant FIR filter should be normally distributed about zero. They not only have a significant variance (0.34) and nonzero mean ( $-0.013$ ), but are very asymmetric about that mean (skewness = 0.53). The nondeterministic but time-invariant ARX model, on the other hand, generated residuals with a small variance (0.18) and a near-zero mean ( $-0.0046$ ). The skewness of the ARX residuals, however, is relatively high (0.33), indicating a bias in the model (i.e., it consistently over- or under-estimates the predicted output) which implies a true nonlinear relationship between the model input and output which cannot be accounted for with such a simple linear model structure. The KF residuals, while possessing a higher variance (0.27), have much less asymmetry (skewness = 0.10) than the ARX residuals.

[27] Finally, a time-lagged correlation analysis was performed to determine which model type is most appropriate. Normalized auto-correlation (ACF) and cross-correlation (CCF) functions were calculated (see *Johansson* [1993], or almost any modern time series analysis text for details). The top half of Figure 7 shows the ACFs for both the solar wind input and the residuals presented in Figure 6. The 95% confidence interval is based on a  $\chi^2$  distribution, and specifies the correlation limit at which we chose to consider the signal “white”. The FIR model residuals show significant 13, 27, 90, and 180 day periods, all often noted in

solar wind and magnetospheric observations [*Li et al.*, 2001a and references therein]. These oscillations also exist in the ARX residuals, but with much smaller amplitudes, often, but not always, falling below the 95% confidence limits. The KF residuals, on the other hand are almost perfectly white within the 95% confidence interval used here.

[28] Whiteness of residuals, in this case, is a strong indication that the model is able to account for nonlinear dynamics in the system. Such nonlinearities can be the result of either complicated internal dynamics (usually manifesting as recurrent structure in the time series itself), or actual nonlinear relationships that describe how the input data directly drives the output data. The relatively small cross correlations between the KF residuals and the solar wind speed input imply that this adaptive FIR filter was able to account for both types of nonlinear dynamics, while the ARX model mostly only accounted for internal and recurrent dynamics that were not directly related to solar wind speed. For “nowcasts” it may very well be best to use something like the ARX model. If, however, one intends to attempt to provide actual forecasts, bias in the model will ultimately lead to a degradation in prediction quality. As is almost always the case, it appears as though the choice of model type is a strong function of the intended application.

## 5. A Practical KF Implementation

[29] The preceding performance evaluation demonstrated the value of the KF as a powerful adaptive system identification and filtering technique. The small but significant correlations remaining in the KF residuals, however, indicate that there are deficiencies in the underlying model structure that cannot be accounted for with either persistence or simple nonlinearities that manifest as time-variant model parameters. A practical real-time radiation belt specification and forecasting tool will require a number of model enhancements. These can be broken into three distinct categories: 1) choice of appropriate real-time data sets; 2) improved dynamical model structures; and 3) a multidimensional, gridded radiation belt state.

### 5.1. Real-Time Data

[30] Possibly the most germane consideration for any empirical radiation belt model is the choice of data to incorporate. The OMNIWeb and SAMPEX data used in this study to validate the KF will not suffice for an operational model because: 1) they are not available in real-time; and 2) they have undergone a significant amount of post-processing to put them in a form that is more appropriate for scientific analysis than it is for assimilation into on-line specification and forecast models.

[31] The most likely candidate to provide real-time solar wind measurements is currently the Advanced Composition Explorer (ACE) satellite. ACE is stationed well outside of the Earth’s magnetosphere, in the solar wind, approximately 1.5 million kilometers ( $\sim 220$  RE) upstream of the Earth. While ACE was primarily designed for scientific

purposes, it provides nearly continuous real-time measurements of relevant solar wind parameters, many of which have already been incorporated into a number of space weather models [Li *et al.*, 2001b] (Space Environment Center, NOAA, REFM, Relativistic Electron Forecast Model, <http://www.sel.noaa.gov/refm/>, 2003). While the end of its likely operational life is approaching, both the scientific and engineering communities have come to acknowledge its vital importance to magnetospheric studies and space weather applications, so it is probable that similar missions will follow.

[32] A number of operational satellites already provide in situ energetic electron measurements. The National Oceanic and Atmospheric Administration (NOAA) geostationary GOES satellites, and the Los Alamos National Laboratory (LANL) geostationary satellites provide complete local-time coverage at geostationary altitudes. In addition, a number of the U.S. Department of Defense Global Positioning System (GPS) satellites also carry particle detectors on-board, and they pass through the heart of the outer electron radiation belt several times per day. Combined, these satellites can potentially provide real-time radiation belt measurements from  $L \approx 4-7$ . Finally, the National Polar-orbiting Operational Environmental Satellite System (NPOESS), first proposed in 1994, should see its first component launch by 2009. This multipurpose, multi-agency, multisatellite system will provide data on the radiation environment across all relevant L-shells in real-time (IORD II - National Polar-Orbiting Operational Environmental Satellite System (NPOESS), <http://npoeslib.ipo.noaa.gov/>, 2001).

## 5.2. Improved Dynamical Model

[33] A very simple model (the single input/single output (SISO) finite impulse response (FIR) filter) was chosen for the purpose of demonstrating the capabilities of the KF without focusing unnecessarily on its underlying dynamical structure. There are a number of more sophisticated dynamical model structures that should provide better predictions, and ultimately, improved insight into the dynamical characteristics of the radiation belts.

[34] First, while solar wind speed has a well-known correlation with electron fluxes in the inner magnetosphere, it is clearly not the only driver. Multiple solar wind measurements should be included as simultaneous inputs, especially those pertaining to the interplanetary magnetic field. Suitable inputs may be determined empirically [Weigel *et al.*, 2002, 2003], through expert analysis [Li *et al.*, 2001b], or using formal dimensional analysis [e.g., Perreault and Akasofu, 1978; Vasyliunas *et al.*, 1982; Gonzalez, 1990] when the objective is improved understanding and separation of the true physical processes.

[35] Next, the FIR structure requires a large number of time-lagged filter coefficients to describe the temporal evolution of a dynamic system with any accuracy. This, as we noted for the period early in 1994, results in a significant amount of good data that is necessarily thrown out with the bad. However, if one extends equation (1) to include dynamic feedback, it is often possible to recreate

the full response of the system with far fewer free parameters. Equation (4) is usually referred to as an *Output Error* (OE) model.

$$\hat{y}_t = \sum_{j=p}^M F_j u_{t-j} - \sum_{j=1}^N G_j \hat{y}_{t-j} \quad (4)$$

[36] In addition, much of the recurrent internal dynamics of the radiation belts can be reproduced with autoregressive filters, as in the case of the ARX model. By additionally filtering the time-lagged model residuals, however, a more flexible “noise model” is provided in the form of equation (5), often referred to as an Auto-Regressive Moving Average filter (ARMA; some contemporary literature refers to the ARX model as an ARMA model, but this notation is generally considered obsolete):

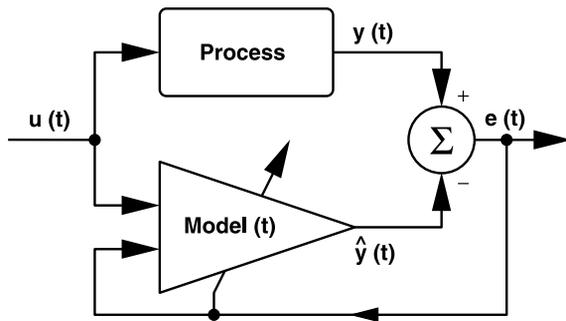
$$\hat{y}_t = \sum_{j=1}^M H_j e_{t-j} - \sum_{j=1}^N G_j \hat{y}_{t-j}; \quad e_t = y_t - \hat{y}_t \quad (5)$$

[37] Simple linear system identification techniques cannot be used for these types of models because they are nonlinear in their parameters (notice the  $\hat{y}$  on the right-hand side of these equations, indicating that the filters operate on previous model output rather than actual observations). Nonlinear estimation algorithms, based on iterative gradient decent techniques, must be used to determine the optimal model parameters. It is therefore necessary to carry along an estimate of the actual radiation belt state, in addition to the model parameters, if one wishes to track changes in model parameters with time. The so-called Extended Kalman filter (EKF) does this by design, and, if configured properly, behaves very similarly to more traditional adaptive maximum-likelihood algorithms [Ljung, 1979].

[38] Combining the ARMA with an FIR filter gives the popular ARMAX model [Johansson, 1993; Ljung, 1999; Nelles, 2001, and other standard texts]. Similarly, replacing the FIR with an OE filter results in the well-known *Box-Jenkins* model [Box and Jenkins, 1976]. These types of models are considered nondeterministic because they cannot generate output unless system observations are available to calculate the residuals. This is illustrated more clearly in Figure 8. Nondeterministic models may or may not provide better predictions than deterministic models, but they allow for better separation of the externally driven and recurrent internal dynamics that both govern radiation belt variability.

## 5.3. Multidimensional Radiation Belt State

[39] One assumption that has been made repeatedly throughout this paper is that one is working in a one-dimensional coordinate system (L-shell). This has obvious scientific merit and value, but it is of little use to satellite designers and mission operators, who will typically require data in a more standard 3-D cartesian coordinate system. The Earth’s external magnetic field is constantly



**Figure 8.** Block diagram describing an adaptive system identification configuration for a generalized nondeterministic model. Note that the residuals are used to dynamically modify the model to better match new observations as well as drive the model directly (indicated by the arrow pointing into the model).

changing, complicating the necessary coordinate transformation tremendously on very short time-scales. The best external field models to date are only accurate enough in a highly time-averaged sense.

[40] From a practical standpoint, it is better to forego a certain amount of physical understanding and simply start with a more geographically consistent, gridded coordinate system. This would be used to define finite volumes through which various radiation sensors might fly. Each finite volume would have its own time-adaptive filter(s) describing its unique dynamical characteristics as well as its relationship to other finite volumes within the overall grid. Reasonable accuracy can be achieved early with only a handful of grid points and sensor platforms due to the considerable dynamical coherence of the radiation belts [Baker et al., 2001; Kanekal et al., 2001]. This approach is quite scalable, so once the EKF and filter-based model framework is established, it is a simple matter to increase the resolution of the model as more powerful computers became available.

## 6. Summary and Conclusions

[41] We have shown that the Kalman filter (KF) provides a robust framework for combining available data with a dynamic model in order to generate the best possible estimate of the current, and possibly future, state of the electron radiation belt. A simple model structure (the finite impulse response (FIR) filter) was given tremendous flexibility by acknowledging the possibility that its parameters might vary with time in a controlled manner. This, in effect, allowed a linear model to reproduce output from a very nonlinear dynamical system to a high degree of accuracy.

[42] We observed significantly improved prediction quality when compared to a time-invariant FIR model, if not necessarily the ARX model. More importantly, however, we noted that there was almost zero auto-correlation in the residuals of the KF with process noise, which means

that nearly all of the useful information available in the electron flux time series was exploited to account for persistence and other recurrent dynamics. In addition, the cross-correlation between the KF residuals and the solar wind speed input were much lower than the residuals for either the time-stationary FIR, or the ARX model, implying that the improved predictive ability of the adaptive KF is not simply a matter of accounting for persistence.

[43] The correlations for the KF residuals were not perfectly white however, indicating that there is still room for improvement. More appropriate real-time data, more sophisticated baseline models, and a more realistic spatial representation of the radiation belts will all serve to enhance prediction quality and ultimately provide a state-of-the-art space weather specification and forecasting tool.

[44] **Acknowledgments.** Thanks are extended to the SAMPEX data team and the NSSDC at Goddard Space Flight Center for providing the high-quality data necessary to this research. This study was supported through NASA's Graduate Student Research Program (GSRP, award NGT5-132) and the NSF National Space Weather Program (NSWP, award ATM-0208341).

## References

- Arnoldy, R. (1971), Signature in the interplanetary medium for substorms, *J. Geophys. Res.*, 76(22), 5189–5201.
- Baker, D. N. (2001), Satellite anomalies due to space storms: The effects of space weather on spacecraft systems and subsystems, in *Space Storms and Space Weather Hazards*, edited by I. Daglis, pp. 285–311, Kluwer Acad., Norwell, Mass.
- Baker, D. N. (2002), How to cope with space weather, *Science*, 297, 1486–1487.
- Baker, D. N., R. L. McPherron, T. E. Cayton, and R. W. Kebedad (1990), Linear prediction filter analysis of relativistic electron properties at 6.6 RE, *J. Geophys. Res.*, 95(A9), 15,133–15,140.
- Baker, D. N., G. M. Mason, O. Figueroa, G. Colon, J. G. Watzin, and R. M. Aleman (1993), An overview of the Solar, Anomalous, and Magnetospheric Particle Explorer (SAMPEX) mission, *IEEE Trans. Geosci. Remote Sens.*, 31(3), 531–541.
- Baker, D. N., S. G. Kanekal, J. B. Blake, and T. I. Pulkkinen (2001), The global efficiency of relativistic electron production in the Earth's magnetosphere, *J. Geophys. Res.*, 106(A9), 19,169–19,178.
- Bargatze, L. F., D. N. Baker, R. L. McPherron, and E. W. Hones Jr. (1985), Magnetospheric impulse response for many levels of geomagnetic activity, *J. Geophys. Res.*, 90(7), 6387.
- Boller, B. R., and H. L. Stolov (1970), Kelvin-helmholtz instability and the semi-annual variation of geomagnetic activity, *J. Geophys. Res.*, 75(31), 6073–6084.
- Box, G. E. P., and G. M. Jenkins (1976), *Time Series Analysis Forecasting and Control: Revised Edition*, Holden-Day, Boca Raton, Fla.
- Clauer, C. R., R. L. McPherron, C. Searls, and M. B. Kivelson (1981), Solar wind control of auroral zone geomagnetic activity, *Geophys. Res. Lett.*, 8, 915.
- Clauer, C. R., R. L. McPherron, and C. Searls (1983), Solar wind control of the low-latitude asymmetric magnetic disturbance field, *J. Geophys. Res.*, 88(A4), 2123.
- Cliver, E. W., Y. Kamide, and A. G. Ling (2000), Mountains versus valleys: Semiannual variation of geomagnetic activity, *J. Geophys. Res.*, 105(A2), 2413–2424.
- Cook, W. R., et al. (1993), A proton electron telescope for studies of magnetospheric, solar, and galactic particles, *IEEE Trans. Geosci. Remote Sens.*, 31(3), 571–576.
- Gonzalez, W. D. (1990), A unified view of solar wind-magnetosphere coupling functions, *Planet. Space Sci.*, 38(5), 627–632.
- Iyemori, T., H. Maeda, and T. Kamei (1979), Impulse response of geomagnetic indices to interplanetary magnetic field, *J. Geomagn. Geoelectr.*, 31(1), 1–9.
- Johansson, R. (1993), *System Modeling and Identification*, Prentice-Hall, Old Tappan, N. J.

- Kamide, Y., and J. A. Slavin (Eds.) (1986), *Solar Wind-Magnetosphere Coupling, Astrophys. and Space Sci. Libr.*, Terra Sci., Tokyo.
- Kanekal, S. G., D. N. Baker, and J. B. Blake (2001), Multisatellite measurements of relativistic electrons: Global coherence, *J. Geophys. Res.*, *106*(A12), 29,721–29,732.
- Li, X., D. N. Baker, S. G. Kanekal, M. Looper, and M. Temerin (2001a), Long term measurements of radiation belts by sampex and their variations, *Geophys. Res. Lett.*, *28*(20), 3827–3830.
- Li, X., M. Temerin, D. N. Baker, G. D. Reeves, and D. Larson (2001b), Quantitative prediction of radiation belt electrons at geostationary orbit based on solar wind measurements, *Geophys. Res. Lett.*, *28*(9), 1887–1890.
- Ljung, L. (1979), Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Trans. Autom. Control*, *ac-24*(1), 36–50.
- Ljung, L. (1999), *System Identification: Theory for the User*, 2nd ed., Prentice Hall PTE, Upper Saddle River, N. J.
- McPherron, R. L., R. A. Fay, C. R. Garrity, L. F. Bargatze, D. N. Baker, C. R. Clauer, and C. Searls (1984), Coupling of the solar wind to measures of magnetic activity, *Eur. Space Agency Spec. Publ.*, *ESA SP*, *217*, 161.
- Nagai, T. (1988), Space weather forecast: Prediction of relativistic electron intensity at synchronous orbit, *Geophys. Res. Lett.*, *15*(5), 425–428.
- Nelles, O. (2001), *Nonlinear System Identification*, Springer-Verlag, New York.
- Perreault, P., and S.-I. Akasofu (1978), A study of geomagnetic storms, *Geophys. J. R. Astron. Soc.*, *54*, 547–573.
- Russell, C. T., and R. L. McPherron (1973), Semiannual variation of geomagnetic activity, *J. Geophys. Res.*, *78*(1), 92–107.
- Vassiliadis, D. (2000), System identification, modeling, and prediction for space weather environments, *IEEE Trans. Plasma Sci.*, *28*(6), 1944–1955.
- Vassiliadis, D., A. J. Klimas, J. A. Valdivia, and D. N. Baker (2000), The nonlinear dynamics of space weather, *Adv. Space Res.*, *26*(1), 197–207.
- Vassiliadis, D., A. J. Klimas, S. G. Kanekal, D. N. Baker, and R. S. Weigel (2002), Long-term-average, solar cycle, and seasonal response of magnetospheric energetic electrons to the solar wind speed, *J. Geophys. Res.*, *107*(A11), 1383, doi:10.1029/2001JA000506.
- Vasyliunas, V. M., J. R. Kan, G. L. Siscoe, and S.-I. Akasofu (1982), Scaling relations governing magnetospheric energy transfer, *Planet. Space Sci.*, *30*, 359–365.
- Weigel, R. S., D. Vassiliadis, and A. J. Klimas (2002), Coupling of the solar wind to temporal fluctuations in ground magnetic fields, *Geophys. Res. Lett.*, *29*(19), 1915.
- Weigel, R. S., A. J. Klimas, and D. Vassiliadis (2003), Solar wind coupling to and predictability of ground magnetic fields and their time derivatives, *J. Geophys. Res.*, *108*, 1298.

---

D. N. Baker, E. J. Rigler, and R. S. Weigel, Laboratory for Atmospheric and Space Physics, University of Colorado, 1234 Innovation Drive, Boulder, CO 80303, USA. (jrigler@colorado.edu)

A. J. Klimas and D. Vassiliadis, Goddard Space Flight Center, NASA, Code 692, Greenbelt, MD 20771, USA.