

LA-UR-05-6700

*Approved for public release;
distribution is unlimited.*

Title: Radiation Belt Data Assimilation and Parameter Estimation

Author(s): Josef Koller, ISR-1
Reiner H.W. Friedel, ISR-1
Geoffrey D. Reeves, ISR-1

Submitted to:



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Radiation Belt Data Assimilation and Parameter Estimation

J. Koller, R.H.W. Friedel, and G.D. Reeves

Space Science and Applications, Los Alamos National Laboratory, Los Alamos,
New Mexico, USA

Abstract. In this paper we apply the techniques of data assimilation to the problem of radial diffusion of electrons in the radiation belts. Unlike previous studies, the objective of this work is not to reproduce observations but rather to determine the diffusion parameters regulating the physical behavior of the radiation belts. We developed a code containing a radial diffusion solver and a Kalman filter and show how the assimilation of data could potentially pin-down the most likely diffusion parameters. We supplied our assimilation model with artificial data in an identical twin experiment in order to illustrate the strengths and capabilities of data assimilation techniques. We find that there is a degeneracy in the parameterized diffusion function and that these parameters are anti-correlated. We discuss how this degeneracy can be broken by decreasing the data uncertainty or by increasing the data sampling frequency.

1. Introduction

The highly energetic electron environment in the inner magnetosphere (geosynchronous orbit and inward) has received a lot of research attention in recent years to better understand the dynamics of relativistic electron acceleration, loss, and transport. Physical processes in the natural energetic electron environment in the Earth's radiation belts are important to understand because dynamic variations in this environment can negatively impact the space hardware that our society increasingly depends on.

It has been known since the 1970's that radial diffusion is a key process influencing radiation belt dynamics. Recently, new observations and increased monitoring [Friedel *et al.*, 2002] evidenced that other processes play an important role as well. Reeves *et al.* [2003] show that the net effect of geomagnetic activity on radiation belt dynamics is a delicate balance of acceleration, transport, and losses that can lead to either increased or decreased fluxes or to almost no changes at all. Despite uncertainties in the precise nature of all the processes controlling radiation belt dynamics, it is widely believed that radial diffusion is one of the critical factors that needs to be accurately specified. Boscher *et al.* [1996] and Bourdarie *et al.* [1996] show that radial diffusion accounts for 80% of the dynamics.

Although the physics behind the outer radiation belt flux variations is not yet fully understood, several correlations between the flux evolution and the power of ultra-low frequency (ULF) waves have been observed. Wave activity and particle dynamics are linked together by radial diffusion. ULF waves that are in resonance with particles motions, result in a violation of the third adiabatic invariant and lead to a particle drift or diffusion [Schulz and Lanzerotti, 1974]. Lyons and Thorne [1973] show that the radial structure and energy spectrum of the radiation belts are consistent with radial diffusion. This paper focuses on the radial transport process, which is only one part of the radiation belt physics, and provides a method on how to estimate the diffusion coefficients that are driven by various, known or unknown, physical mechanisms.

Our new approach is to extend available techniques of data assimilation that are widely used for other geophysical systems (meteorology, oceanography, ionosphere) to the radiation belt. The general purpose of data assimilation is to combine measurements and models to produce best estimates of current and future conditions. The resulting forecast or estimate is, either closer to the data or the model depending on their uncertainties. Correlations and uncertainties are incorporated and carried along automatically. The output is based on all measurements and the model. Forecasting including state and parameter estimation can be done with an increased fidelity.

One important method of data assimilation is the Kalman filter. It became popular because it is a recursive solution to the optimal estimator problem. However, the Kalman filter is just one way of finding an optimized solution. The least square method, for instance, can be extended to provide forecast predictions as well [Tarantola, 1987] and the two methods become basically equivalent. Here, we will use the Kalman filter to eventually test the competing theoretical processes that play a role in the radiation belt and distinguish between them. For an overview of data assimilation methods, see Bouttier and Courtier [1999].

After more than 50 years of successful application in meteorology, data assimilation started

to appear for ionospheric physics and space weather prediction. One of the earliest efforts was the AMIE (Assimilative Mapping of Ionospheric Electrodynamics) algorithm that ingests multiple data sets with a weighted least-squares fit to specify the electrodynamic state of the high-latitude ionosphere [Richmond and Kamide, 1988].

GAIM (Global Assimilation of Ionospheric Measurements) is another example of data assimilation for the ionosphere. It applies a Kalman filter to an ionosphere-plasmasphere model as a basis for assimilating a diverse set of real-time data. GAIM provides both specifications and forecasts on a global or regional grid [Schunk *et al.*, 2004] and is now undergoing transition into operation at the Air Force Weather Agency [McCoy, 2005].

It took more than a decade for data assimilation to move into the radiation belt domain. One of the first attempts was the “direct data insertion” technique used by Bourdarie *et al.* [2005]. They showed that by adding data of just one extra satellite into their simulations the updated model could achieve global fidelity on the order of the input data uncertainty - in effect overcoming the fundamental limitations of the underlying physics model. Bourdarie *et al.* [2005] also described that one of the most crucial, and often overlooked, requirements is the fidelity of satellite data inter-calibration. However, obtaining a well calibrated and inter-calibrated set of radiation belt particle data can be a very time consuming but an essential task [Friedel *et al.*, 2005].

A more sophisticated application of data assimilation to the problem of radiation belt dynamics was recently presented by Naehr and Toffoletto [2005]. They applied a Kalman filter combining a diffusion model with artificial data. Their goal was to improve the state estimate of the system and to forecast radiation belt particle distributions. Naehr and Toffoletto [2005] tested their algorithm with artificial data from an idealized magnetospheric storm and also compared the result with the direct insertion method that was used by Bourdarie *et al.* [2005]. They found that the Kalman filter more accurately reconstructs the global particle distribution from sparse observational data.

Our study builds upon the idea of Naehr and Toffoletto [2005] to use an “identical twin experiment” - a term that has been introduced by the data assimilation community. The method of the identical twin experiment is to create artificial data with simulated uncertainties and to test assimilation schemes in such a controlled environment. This has several advantages over using real data. First, we can create a “true” state, something that can never be found in reality: Data as well as models will always contain uncertainties and they will only be approximations to the true state of a system. Second, data can be artificially created with any kind of error statistics along arbitrary satellite orbits. Third, it is an important exercise to learn the properties and limitations of data assimilation methods before applying them to real data.

In contrast to Naehr and Toffoletto [2005], we don’t use the Kalman filter to study the forecasting capabilities but we ask a different question. Can the Kalman filter tell us something about the underlying physical processes controlling radial diffusion? Can we use limited data and uncertainties to recover the “true” diffusion coefficients? How well can we do this, that is, can we place confidence bounds on our best estimation of the parametrization? Further, what is the impact on our ability to recover the true diffusion coefficients on factors such as spacecraft orbit, location, number of spacecraft, and sampling frequency of observations? What uncertainty in the data is acceptable to gain any new insights? Answers to these questions are pivotal in pointing us toward the area requiring the most attention - do we need more data? If so, what orbits would be the

best ones? Or, is it sufficient to simply increase the fidelity of existing data for now? We use the identical twin experiment to investigate the limitations and conditions under which a data assimilation approach can yield useful results.

While the simple model that we use here cannot hope to represent the reality of the radiation belts, the results obtained here are invaluable learning steps. An eventual application of these methods to more complex codes will require the time-dependent determination of a whole range of processes - wave-particle interactions leading to pitch angle and energy diffusion, radial diffusion (the part treated here), losses due to wave particle interactions, geometrical effects such as magnetopause shadowing, and adiabatic effects. While diffusion is an important part in the radiation belt description, eventually a self-consistent representation is necessary that includes ring current development and its interaction with radiation belt particles through electromagnetic ion cyclotron (EMIC) waves and the changing geomagnetic field. This paper attempts to lay the foundation for the effort to combine all these processes into a Dynamic Radiation Environment Assimilation Model (DREAM) to understand acceleration, transport, and losses in the radiation belts [Reeves *et al.*, 2005]. DREAM is a Laboratory Directed Research and Development project at Los Alamos National Lab. It will develop a next-generation space radiation model using extensive satellite measurements, new theoretical insights, global physics-based magnetospheric models, and the techniques of data assimilation. This paper is a foundation for the DREAM project.

In the following Section 2, we will describe the radial diffusion solver and the data assimilation algorithm (Kalman filter) in more detail. We will keep this section more tutorial and provide some simple examples. After describing how we created artificial data in Section 3, we will discuss the assimilation of an idealized storm in Section 4 and explain how the diffusion parameters can be estimated in Section 5. A discussion and conclusion will follow thereafter in Section 6 and 7. The Appendix will contain details on solving the diffusion equation, error covariance matrices, and Gaussian error statistics.

2. Radial Diffusion and Kalman Filter

2.1. Solving the Diffusion Equation

The physics of the inner magnetosphere, here the radiation belts with relativistic electrons, can be described by the phase-averaged particle distribution function, i.e., phase space density, $f(L, \mu, J, t)$ [Schulz and Lanzerotti, 1973] where the quantities L, μ, J are adiabatic invariants at time t defining the drift motion, periodic gyration and bounce motion [Roederer, 1970]. The relationship between the distribution function f and unidirectional flux J_{\perp} is given by $f = J_{\perp}/p_{\text{rel}}^2$. The flux and the relativistic particle momentum p_{rel} are evaluated on a surface generated by the mirror points of electrons with constant values of μ and J , i.e., with their pitch-angle diffusion neglected [Roederer, 1970]. The Focker-Plank diffusion equation represents the temporal evolution of the phase space density. Here, we assume constant adiabatic invariants μ and J such that the phase space density is only a function of radial diffusion in L and t . The governing equation for $f(L, t)$ then reduces to the simple form [Roederer, 1970]

$$\frac{\partial f}{\partial t} = L^2 \frac{\partial}{\partial L} \left(\frac{D_{\text{LL}}}{L^2} \frac{\partial f}{\partial L} \right) \quad (1)$$

where we also neglect any source or loss terms. They are simply additive and we will include those in future work. The diffusion coefficient D_{LL} is independent of time but a function of L and we refer to it as the “diffusion function” $D_{LL}(L)$ from here on. As in *Selesnick et al.* [1997], we assume the parameterized form

$$D_{LL}(L) = \alpha L^\gamma \quad (2)$$

with α and γ as the diffusion parameters.

The distribution function $f(L, t)$ in Equation (1) is typically continuous. For solving the diffusion equation computationally we assume a discrete meshed grid of the form $(L_1, L_2, \dots, L_N)^T$ where T is the transpose, and create a vector \mathbf{f} containing $f(L_n)$ for each grid point n . This vector will have dimension N where N , in our case, is the number of L -shell bins at which we specify the phase space density. Hence

$$\mathbf{f} = \begin{bmatrix} f(L_1) \\ f(L_2) \\ \vdots \\ f(L_N) \end{bmatrix}. \quad (3)$$

The Focker-Plank Equation (1) is linear and the distribution function f or the distribution vector \mathbf{f} can be scaled by any factor c that is a positive real number. We define a new vector

$$\mathbf{x} = c\mathbf{f} \quad (4)$$

with the scaling factor c such that initially $\max(c\mathbf{f}) = 1$. Both vectors \mathbf{x} and \mathbf{f} are so-called “state vectors”, after standard terminology of data assimilation [*Ide et al.*, 1997]. In its most general form, the “state vector” can be quite large and contain several different variables defined on a multidimensional grid. Model parameters like the diffusion coefficients can be part of the state vector as well. This is the so-called augmented state vector approach [*Voss et al.*, 2004]. Our state vector is simply the phase space density at discrete values of L .

The diffusion Equation (1) can be integrated with a number of methods. We choose the Crank-Nicolson scheme [*Crank and Nicolson*, 1947] which is an implicit, numerically stable method that does not need to satisfy the Courant condition [*Press et al.*, 1986]. The discretized Crank-Nicolson version of Equation (1) needs appropriately centered values of the diffusion function. We introduce $D_{LL}(L_{n+1/2}) = D_{n+1/2}$ as short for the diffusion centered between evenly spaced grid points n and $n + 1$. $L_{n+1/2}$ is defined between grid points in the same way. The Crank-Nicolson scheme then transforms Equation (1) into the following discretized version

$$\begin{aligned} \frac{x_n^{i+1} - x_n^i}{\Delta t} = & \left[\frac{D_{n+1/2}}{L_{n+1/2}^2} (x_{n+1}^i - x_n^i + x_{n+1}^{i+1} - x_n^{i+1}) - \right. \\ & \left. - \frac{D_{n-1/2}}{L_{n-1/2}^2} (x_n^i - x_{n-1}^i + x_n^{i+1} - x_{n-1}^{i+1}) \right] \frac{L_n^2}{2(\Delta L)^2} \end{aligned} \quad (5)$$

where x_n^i is the state at grid point n at time i . Defining

$$\beta_n^\pm = (\Delta t D_{n\pm 1/2}) / [2(\Delta L)^2 L_{n\pm 1/2}^2] \quad (6)$$

where $L_{n\pm 1/2}$ is a cell centered grid between n and $n \pm 1$, the equation above can be written with all future states $i + 1$ on one side and past states i on the other side:

$$\begin{aligned} -L_n^2 \beta_n^- x_{n-1}^{i+1} + [1 + L_n^2(\beta_n^+ + \beta_n^-)] x_n^{i+1} - L_n^2 \beta_n^+ x_{n+1}^{i+1} = \\ L_n^2 \beta_n^- x_{n-1}^i + [1 - L_n^2(\beta_n^+ - \beta_n^-)] x_n^i + L_n^2 \beta_n^+ x_{n+1}^i \end{aligned} \quad (7)$$

One can now conveniently rewrite the equation above in a matrix form of $\mathbf{A}\mathbf{x}^{i+1} = \mathbf{B}\mathbf{x}^i$ where \mathbf{A} and \mathbf{B} are tri-diagonal matrices of dimension $(N \times N)$. Hence the solution of the Crank-Nicolson scheme is to calculate the inverse \mathbf{A}^{-1} and multiply with \mathbf{B} to yield $\mathbf{M} = \mathbf{A}^{-1}\mathbf{B}$, i.e.,

$$\mathbf{x}^{i+1} = \mathbf{M}\mathbf{x}^i. \quad (8)$$

In data assimilation jargon, \mathbf{M} is called the model, dynamics, or forecast operator. All applicable physics, here radial diffusion, is contained in \mathbf{M} . Note, $\mathbf{M}(\alpha, \gamma)$ is not only a function of the parameters in the diffusion function α and γ but also of the integration time step Δt .

The boundary condition is specified in the first and last element of \mathbf{M} by defining \mathbf{M}_{11} and \mathbf{M}_{NN} where N is the dimension of the state vector \mathbf{x} . Typically we choose $N = 180$ grid points for the numerical integration of the diffusion equation using a radial dimension $1 < L < 10$. We note that the system will be measured at far fewer points as discussed below.

A special note on stability and boundary conditions can be found in the Appendix A.

Before calculations can start, the grid has to be initialized with some choice of $\mathbf{x}^0 = \mathbf{x}(t_0)$. One can choose a steady state with $\mathbf{x}^0 = \text{const}$. Other non-constant choices are also possible but have no effect on the results of this study.

Although the radial diffusion operator \mathbf{M} contains the physical model, it can look slightly different for each numerical solver. \mathbf{M} will be used for the prediction part in the Kalman filter.

2.2. Data Assimilation Algorithm

There are several different definitions for data assimilation depending on the application (engineering, statistical, geophysical). Here, we use it as short for model-based assimilation of observations. This already explains that data assimilation is the combination of a physical model with observations. The purpose is to efficiently find the most likely approximation to the true state using the information provided by the model and the data considering both of their uncertainties. Data assimilation methods are based on, and can be derived from, Bayesian statistics, minimum variance, maximum likelihood, or least square methods [*Maybeck*, 1979; *Kalnay*, 2003; *Daley*, 1991; *Talagrand*, 1997; *Tarantola*, 1987; *Tarantola and Valette*, 1982].

2.2.1. A Simple Example of Data Assimilation Let us consider a simple estimation problem to understand the principles and the jargon of data assimilation as applied to the radial diffusion of radiation belt electrons: Assume we have two observation y_1^o and y_2^o made at the same time on a simple one dimensional grid L by two instruments that are placed at different grid locations. Both ‘‘observatories’’ try to measure a scalar quantity $x^{\text{true}}(L)$, the true state on that grid point but have uncertainties or measurement errors. Hence y_1^o and y_2^o are of the form

$$\begin{aligned} y_1^o &= x_1^{\text{true}}(L) + \epsilon_1^o \\ y_2^o &= x_2^{\text{true}}(L) + \epsilon_2^o \end{aligned} \quad (9)$$

where ϵ_1° and ϵ_2° are the “observational errors”. These errors are of course unknown, but we assume that the statistical properties of the instruments are available. For instance, if we would take the same measurement over and over again, we can determine the statistical mean or expectation E . Usually data assimilation assumes the following statistics of observations:

- (1) $E(\epsilon_i) = 0$ which means the observations are unbiased.
- (2) $E(\epsilon_i^2) = \sigma_i^2$ is the variance of the observational errors or the width of the Gaussian error distribution.
- (3) $E(\epsilon_i \epsilon_j) = 0$ for $i \neq j$ tells us that the observational errors are uncorrelated. This will be usually the case if, for instance, the two observations have been obtained with different instruments [Talagrand, 1997].

In the case of the radiation belts, real instruments on satellites often do show a bias, background, and saturation effects. Also, since fluxes are converted to phase space densities that require knowledge of the magnetic field, we expect systematic errors as well. For these reasons, we start out with an “identical twin experiment”, as explained in the section 3, that allows us to study data assimilation methods in a controlled environment with known uncertainties. Once real data is assimilated, all these effects will have to be addressed in more detail. Also, the identical twin experiment allows us to pursue multiple case studies whereas real data will provide only one result.

The information about the uncertainties or errors can be combined into an error covariance matrix R° . Given the example from above with two data points y_1° and y_2° and assuming they are unbiased and uncorrelated, the error covariance matrix (see appendix B) for observation will be simply diagonal

$$R^\circ = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}. \tag{10}$$

The equation relating the measurements to the true state (9) are usually written in a more general vector form

$$\mathbf{y}^\circ = \mathbf{H}\mathbf{x}^{\text{true}} + \boldsymbol{\epsilon}^\circ. \tag{11}$$

The true state vector \mathbf{x}^{true} has the same dimension N as the computational grid. The dimension of \mathbf{y}° and $\boldsymbol{\epsilon}^\circ$ is equal to the number of instantaneous observations m . The purpose of \mathbf{H} is to map from the computational dimension N into the usually much smaller observational dimension m . It is therefore called the “observational operator” or “observation matrix” and has the size $(m \times N)$. If the model variables are different from the observation variables, e.g. flux versus phase space density, \mathbf{H} will apply the necessary space-time interpolations and conversions. Conversions from count rates or fluxes into phase space densities using a chosen geomagnetic field will go in here as well. To give an example for \mathbf{H} in its simplest form, assume that the first observatory or satellite is located exactly on grid point number three and the second satellite provides measurements on grid point number five, then \mathbf{H} is a $(2 \times N)$ matrix

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix}. \tag{12}$$

In reality, $\mathbf{H}(t)$ will be time dependent since most satellites are not stationary.

Satellite locations don't have to exactly fall on a grid point. If they are somewhere between two points, then one needs to apply an interpolation technique and define a "radius of influence". To illustrate the radius of influence, *Daley* [1991] describes the example of a thunderstorm or tornado with a characteristic scale of 10 km or less in an observing network with grid spacing 1000 km. If the thunderstorm lies between stations, it will not be "seen" by the network at all. On the other hand, if it lies directly over one of the stations, it may be misrepresented as a much larger-scale phenomenon. This introduces also the "error of representativeness" [*Daley*, 1991; *Kalnay*, 2003; *Petersen and Middleton*, 1963]. The error of representativeness can be reduced with a denser observational network. Satellites usually deliver measurements on a timescale of seconds and minutes while they sweep through the radiation belt. In our simple one-dimensional example they would provide enough measurement that the error of representativeness would not play a role. In reality the coverage is very low in a three-dimensional space because the number of satellites is limited.

After this introductory example, we want to combine all observations including their uncertainties with a physical model. The methods of data assimilation provide an effective framework to do this. We note that the physical model, here radial diffusion, will have uncertainties as well. The model and the observations are treated on the same level but will be considered according to their uncertainties for the final best estimate. For instance, if we are very uncertain about our observations then the resulting estimate will be closer to the model and vice versa. We can also use data assimilation to learn something about the parameters in our physics model which is the main topic of this study. The next section describes one special method of data assimilation, the Kalman filter, in more detail.

2.2.2. The Kalman Filter One popular example of data assimilation is the linear Kalman filter [*Kalman*, 1960]. It is an optimal recursive data processing algorithm [*Maybeck*, 1979, p. 4] that has become a favorite for many engineering application including the navigational system on the Apollo mission, GPS stand-alone devices, and many more [*Sorenson*, 1985]. The Kalman filter is so widely used because it guarantees a convergence on the maximum likelihood state [*Brammer and Siffing*, 1989]. The word "filter" defines assimilation techniques that use only data from the past to obtain an analysis. A method that uses also future data would be called a "smoother" [*Bouttier and Courtier*, 1999].

The Kalman filter involves three steps that are summarized as follows:

$$\left. \begin{array}{l} \mathbf{y}^o(t_i) \\ \mathbf{x}^f(t_i) \end{array} \right\} \xrightarrow{\mathbf{K}} \mathbf{x}^a(t_i) \xrightarrow{\mathbf{M}} \mathbf{x}^f(t_{i+1}) \quad (13)$$

(1) Gain computation: which yields the "Kalman gain matrix" or "weight matrix" \mathbf{K} . (2) State estimate: which uses the Kalman gain \mathbf{K} to weight the "observational residual" (in the older meteorological literature) or the "innovation vector" $\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^f$ and computes the "state estimate" or "assimilated state" \mathbf{x}^a . It includes information from the model and the observations at time t_i and gives the best estimate. (3) State forecast or prediction: The next step is to apply a "forward model operator" \mathbf{M} which results in the "forecast state vector" $\mathbf{x}^f(t_{i+1})$ that can be compared with new observations at time t_{i+1} in the next cycle. Figure 1 summarizes all steps

in the Kalman cycle and Table 1 provides the dimensions of typical vectors and operators in the Kalman filter algorithm.

Variable	Dimension	Description
\mathbf{x}^f	$(N \times 1)$	forecast state vector
\mathbf{x}^a	$(N \times 1)$	assimilated state vector
\mathbf{y}^o	$(m \times 1)$	observation vector
\mathbf{P}^a	$(N \times N)$	ECV assimilation matrix
\mathbf{P}^f	$(N \times N)$	ECV forecast matrix
\mathbf{K}	$(N \times m)$	Kalman gain matrix
\mathbf{H}	$(m \times N)$	observation operator
\mathbf{M}	$(N \times N)$	forward model operator
\mathbf{Q}^m	$(N \times N)$	model ECV
\mathbf{R}^o	$(m \times m)$	observation ECV
\mathbf{I}	$(N \times N)$	unity matrix

Table 1. Description of vectors and operators used in the Kalman filter. The value N represents the number of grid points, and m is the number of observations at time t_i . The term ECV stands for error covariance matrix.

In the following, we describe the three steps of the Kalman filter in more detail but start out with the initial condition in step zero.

0. Initial conditions: Choose initial values for the state vector $\mathbf{x}^f(0)$ and $\mathbf{P}^f(0)$. Initial conditions can be the result from a previous model run, from a rough estimate, or from a statistical “climatological” average (e.g. the AE-8 trapped particle flux maps [Vette, 1991] of the radiation belts). Note, it is sufficient to start the initial error covariance matrix $\mathbf{P}^f(0)$ with diagonal terms only. The Kalman filter will provide the cross correlations using the model operator \mathbf{M} in step three.
1. Gain computation: Given the error covariance matrix $\mathbf{P}^f(t_i)$ from the previous cycle or the initial condition, the first step is to compute the Kalman gain matrix \mathbf{K} . The observation operator \mathbf{H} from measurements at time t_i is required at this step as well. This will provide the Kalman filter with information from what L -bin the data comes from. The observational uncertainties are provided via the error covariance matrix \mathbf{R}^o of dimension $(m \times m)$. Since we assume that the measurements are uncorrelated, \mathbf{R}^o is a diagonal matrix.
2. Update estimate: The second step involves using the forecast $\mathbf{x}^f(t_i)$ that came out from the previous cycle or the initial condition. The forecast is compared to the observations \mathbf{y}^o using the Kalman gain matrix \mathbf{K} . This is the step where the actual observations $\mathbf{y}^o(t_i)$ pull the state \mathbf{x}^a away from the model toward the observations or vice versa depending on the confidence level of data and model. If the confidence in the observations is low, i.e., with large values in the observations error matrix \mathbf{R}^o , the estimate will favor the model. On the

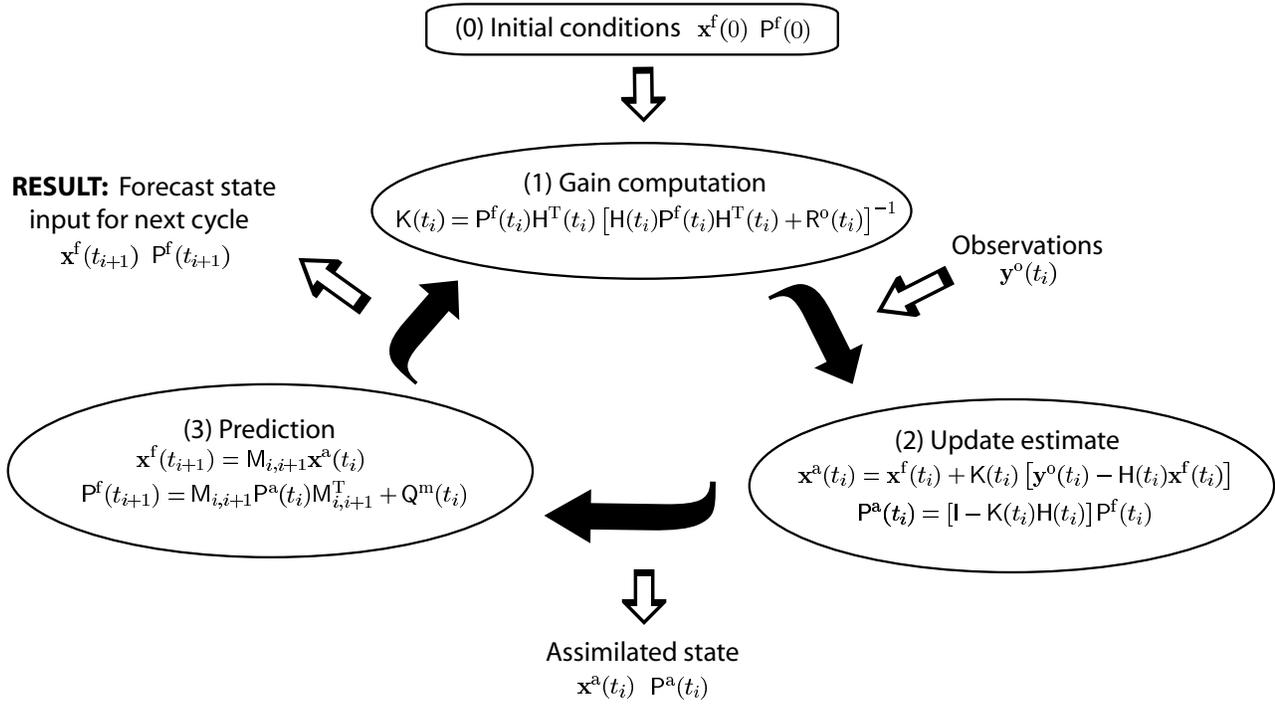


Figure 1. Flow diagram of the recursive Kalman filter algorithm. Starting with initial state vector $\mathbf{x}^f(0)$ and an initial assumption for the error covariance $\mathbf{P}^f(0)$, the first step is to compute the Kalman gain matrix \mathbf{K} . Then observations $\mathbf{y}^o(t_i)$ are used to calculate the state estimate \mathbf{x}^a . Step three yields the forecast state vector \mathbf{x}^f which is used as input for the next cycle.

other hand, if the errors of the model (e.g. the diffusion parameters) are large, then more weight will be given to the observations. This step also updates the error covariance matrix \mathbf{P}^a using the previous \mathbf{P}^f and the Kalman gain \mathbf{K} .

3. Prediction: The third and last step is to predict the new state vector $\mathbf{x}^f(t_{i+1})$ and the error covariance matrix $\mathbf{P}^f(t_{i+1})$ by advancing the old $\mathbf{x}^a(t_i)$ and $\mathbf{P}^a(t_i)$ from step two using the model operator $\mathbf{M}_{i,i+1}$. The model error matrix \mathbf{Q}^m is incorporated here. It is interesting to note that even though the initial error covariance matrix $\mathbf{P}^f(0)$ was diagonal, in this step \mathbf{M} is operated twice on $\mathbf{P}^f(0)$. Since \mathbf{M} has off-diagonal terms, the new $\mathbf{P}^f(t_1)$ will now contain off-diagonal correlation terms that are consistent with the model.

The loop starts again with step one using the output \mathbf{x}^f and \mathbf{P}^f from this cycle. New observations are compared to the forecast from this cycle.

3. Creating Data of an Idealized Storm

As we described before in Section 2.2.1, observation and true state are related by $y^o = x^{\text{true}} + \epsilon^o$. In reality we will always have observations y^o and some knowledge about their statistical uncertainties but we will never know x^{true} exactly. However, in the “identical twin experiment”

we can assume a certain “true state” and create an artificial data point y^o by adding statistical uncertainty to the “observation” using a random number generator. This section describes how we build artificial data sets for the identical twin experiment of an idealized geomagnetic storm.

The model operator $M(\alpha, \gamma)$ depends on a parameterized version of the diffusion coefficient similar to *Selesnick et al.* [1997] who use $D_{LL} = D_s(L/4)^{p_s}$ and find the following diffusion parameters $D_s = (2.1 \pm 0.2) \times 10^{-3} \text{ days}^{-1}$ and $p_s = 11.7 \pm 1.3$ provided a best fit to POLAR satellite observations. They applied a least square method to a one dimensional diffusion model and 3 to 8 MeV electron data from a geomagnetically quiescent period in mid 1996. Since *Selesnick et al.* [1997] combined data with a model, their approach can be considered as data assimilation as well. We only use their result as a starting point in our studies.

We take their mean diffusion parameters and use our own parameterization (Equation 2). The diffusion parameters are then $(\alpha^{\text{true}}, \gamma^{\text{true}}) = (8 \times 10^{-12} \text{ hours}^{-1}, 11.7)$ which we call the “true” coefficients from here on because these are the ones we used to create the data.

We can calculate now the evolution of an idealized storm similar to *Naehr and Toffoletto* [2005]. Before the beginning of the storm at $t = 50$ hours, the state vector \mathbf{x}^{true} consists of low but non-zero values. The time of the storm onset is marked by simply raising the outer boundary $\mathbf{x}_N^{\text{true}} = 1$. Raising the outer boundary and keeping it up high allows for fast inward diffusion. After five days of storm time and inward diffusion, at $t = 170$ hours we reduce the value of the outer boundary to $\mathbf{x}_N^{\text{true}} = 0$ which basically stops the inflow and the storm settles down (see Figure 2). These particular boundary conditions simulate a source at the outer boundary. Losses occur only at the boundary when they are set to zero. No other explicit sources or losses are included in the model since we chose Equation (1) to be source and loss free. The evolution of a virtual geomagnetic storm is shown as a color plot of normalized phase space density in Figure 2 including the satellite orbits.

With the temporal evolution of the true state $\mathbf{x}^{\text{true}}(t)$ now given, we can basically fly any kind of satellite through the idealized storm and create artificial data. These satellites sample the system in a similar way real satellites would sample the radiation belts. Such artificial data sets are commonly referred to as “virtual satellite data”. We assume geosynchronous and geosynchronous transfer orbits like the CRRES satellite and calculate the true state these satellites would measure based on their position at time t_i . Typically, we fly a geosynchronous satellite at $L = 6.5$ and a CRRES like spacecraft between $4 < L < 8$ (Figure 2). We experimented with different orbital periods. The real CRRES satellite had a period of ten hours. After we calculate the true value for each measurement, we fold a random Gaussian error onto the true state and create an artificial observation. The measurements have an error with a mean that is equal to the true state and a variance of $(\sigma^o)^2 = (\bar{\epsilon^o})^2$. In other words, if we could stop time and take the same observation over and over again, the mean of these observation would be the true state $\bar{y}^o = x^{\text{true}}$. The standard deviation σ^o from that mean value would be equal to the average observational error.

This setup allows us to choose the number of “weather stations”, the orbits and periods, and the data sampling frequencies of each satellite. Notice that the system is sampled at only two points at any given time and that the sampling rate is much more coarse than the simulation time step.

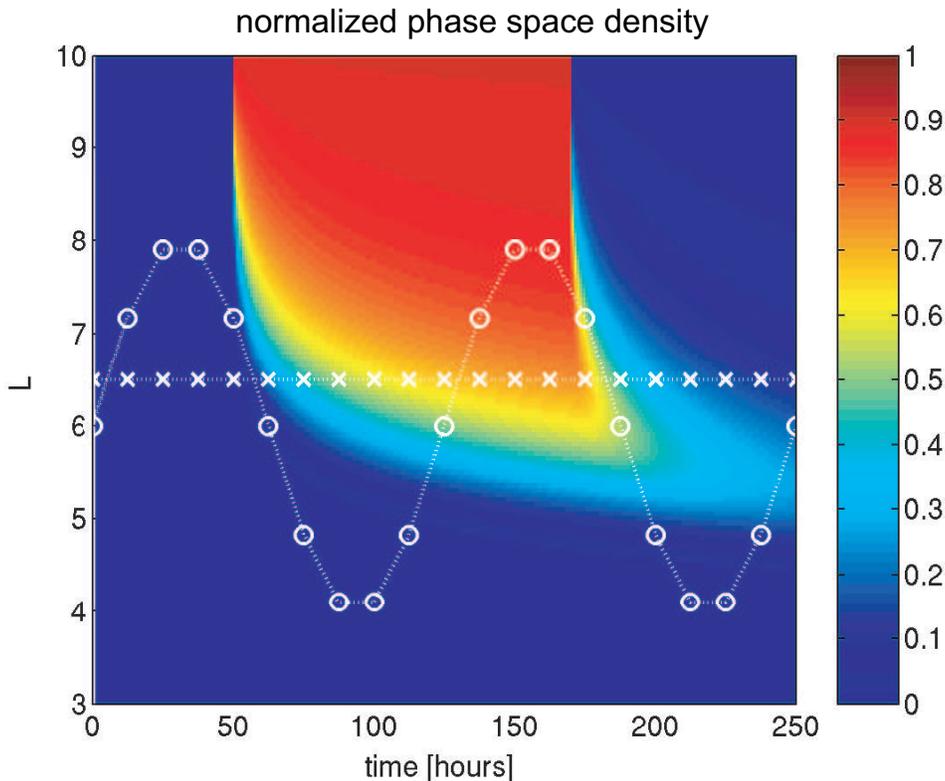


Figure 2. Evolution of the true phase space density with the measurement locations from two satellite orbits. The outer boundary is kept at the level of the storm beginning to allow for electron in-flow. After $t = 170$ hours, the outer boundary is set back to zero stopping the in-flow. Colors show the intensity of simulated storm. The location of satellite measurements are indicated by “ \times ” (geosynchronous at $L = 6.5$) and “ \circ ” (CRRES-like orbit between $4 < L < 8$).

We will now use these virtual data and assimilate them into a model. The next section describes how this can be done using a Kalman filter.

4. Assimilating an Idealized Storm

4.1. Results of an Assimilated Storm

We start to illustrate the data assimilation of an idealized storm using the true diffusion parameters, i.e., we use the same parameters in the model $M(\alpha^{\text{true}}, \gamma^{\text{true}})$ that we used to create the data. Consequently, the diffusion in the physics model $M(\alpha, \gamma)$ is perfect. This should give us a very close result to the true state. However, it will not be equal to the true state since we have uncertainties folded into the measurements. The result will be a best fit given the data.

For all simulations we chose a constant model error matrix $Q_{ii}^m = 0.1$ with diagonal elements only. The model error is included into the the data assimilation algorithm to allow for a fuzziness

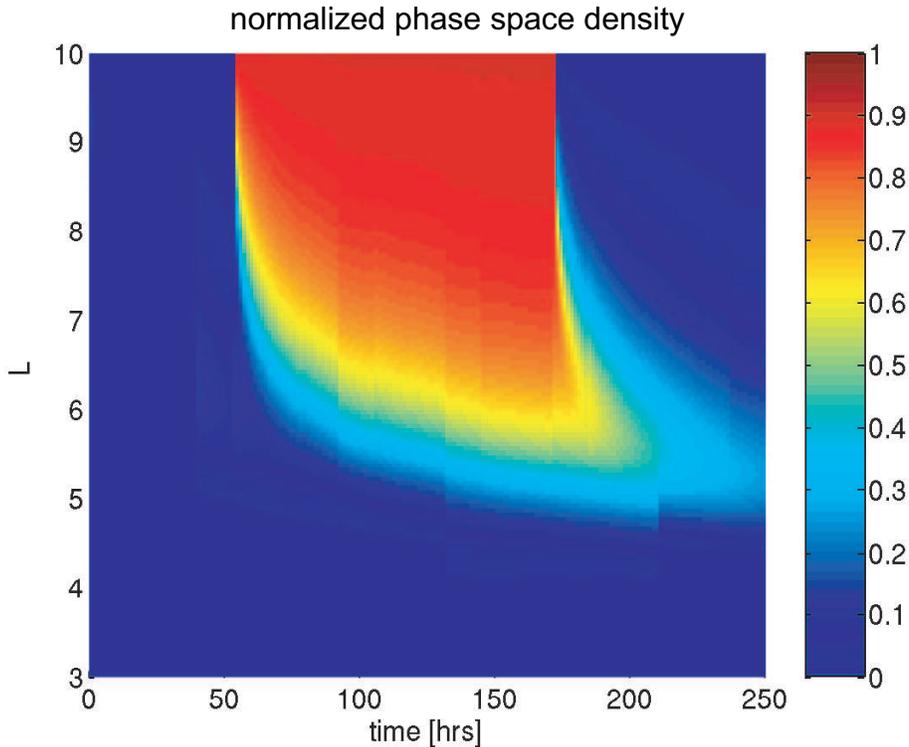


Figure 3. Evolution of the forecast state estimate \mathbf{x}^f applying the same diffusion parameters that were used for creating the virtual data. Hence we used a perfect model operator \mathbf{M} . The Kalman filter does an excellent job in reproducing a close result to the “true” state despite the large measurement errors in the artificial data.

of the forward model \mathbf{M} . Otherwise, \mathbf{M} would imply a deterministic model only.

The global picture of the forecast state \mathbf{x}^f filter in Figure 3 is indeed in very good agreement with the “true” state. The plot of the relative error in Figure 4 where the average error is about 10% supports that statement. However, the error is not zero because every single data point has an observational uncertainty as well.

Many symbols denoting the data sampling locations in Figure 2 are in the regions where the fluxes are extremely low (dark blue). We chose not to assimilate very low values to simulate the Poisson statistics of real particle counters which usually fly on radiation belt satellites (see Appendix C for details). That leaves only a few data points where the spacecraft fly through the storm with higher flux values. Yet, the output of the Kalman filter is already in good agreement with the true state early in the assimilation. This is not only due to the correct diffusion coefficients in $\mathbf{M}(\alpha, \gamma)$ but also due to the correct cross-correlations \mathbf{P}^f that are calculated in the Kalman filter in step three where $\mathbf{P}^f \propto \mathbf{M}\mathbf{P}^a\mathbf{M}^T$.

Figure 4 shows a large error around $t = 50$ hours. The reason for this lies in the random uncertainty of the observations. The random number generator created an outlier and the Kalman

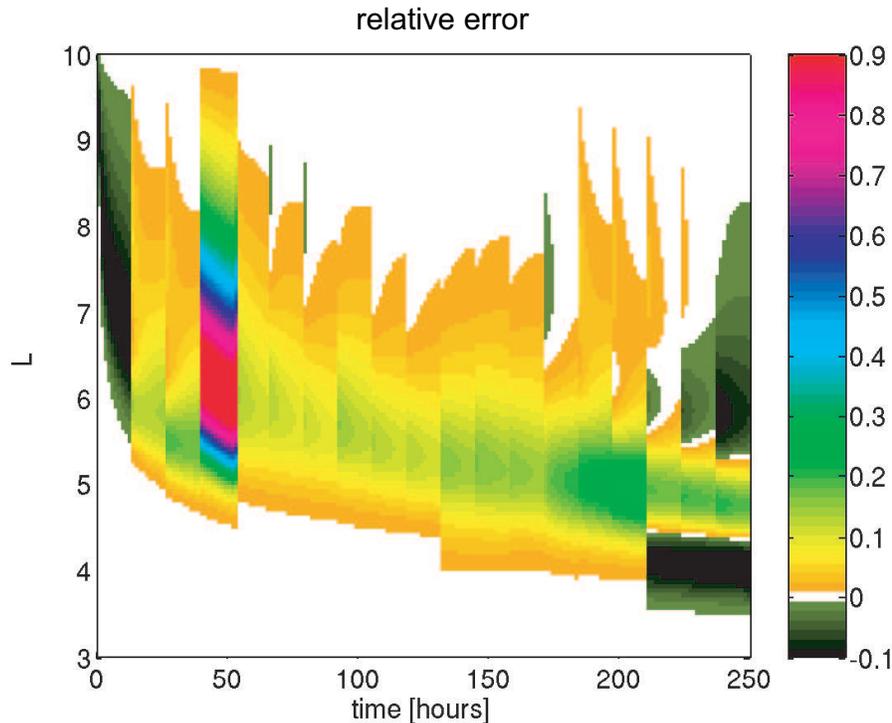


Figure 4. Error of the forecast compared to the true evolution $(\mathbf{x}^f - \mathbf{x}^{\text{true}}) / \max(\mathbf{x}^f)$. The average error is about 10%. The red slot just before $t = 50$ is due to the randomness of the data that happened to be on the high side (see also Figure 2). Once more data is assimilated, the Kalman filter produces a better output that is closer to the true state.

filter produced a matching forecast. Once more observations from the storm were assimilated after $t > 50$ hours, the Kalman filter adjusted the forecast downward accordingly. Note that although it is a large relative error, it is a small absolute error as we will see in the next section. Also we chose a low sampling rate to test the limit of the Kalman filter. With more data, the large error around $t = 50$ would be reduced to a more average value.

4.2. Time Series of Satellite Measurements

We have looked at the global radiation belt picture that resulted from using a Kalman filter. It would have been difficult to derive that picture from the satellite measurements alone.

The actual data that our virtual satellites take is plotted in Figure 5. The evolution of the “true state” that, by design, they fly through is shown as dashed lines. Since the measurements have errors associated with them, they don’t exactly follow the true state evolution but are randomly distributed around it. Only their mean value falls on the true state. The error bars are $\sigma^o = 0.1$. The resulting forecasts from the Kalman filter are indicated by the solid blue and the solid red lines.

As we discussed above, the random number generator produced a few measurement outliers

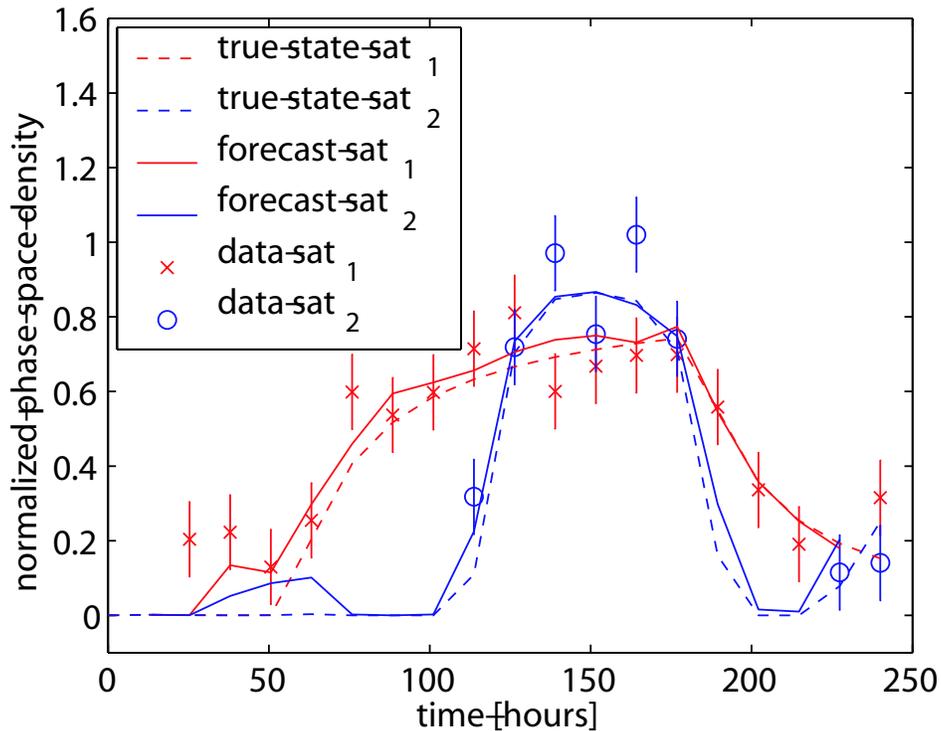


Figure 5. Satellite data, forecast, and true state with two satellites. The geosynchronous spacecraft measurements are plotted as red “ \times ” and data from the CRRES like satellite are indicated as blue “ o ”. The observations are randomly generated and have a Gaussian distribution with a mean of the true state and a variance $(\sigma^\circ)^2 = 0.1^2$. The evolution of the true state at the satellite locations are plotted as dashed lines. The solid lines show the forecast output from the Kalman filter. We used the “true” diffusion parameters, i.e., the data was created with the same D_{LL} .

that are too high although the true state is close to zero for $t < 50$ hours (Figure 5). These data points are assimilated with the Kalman filter and raise the flux estimate even before the storm has started. The effect can be seen in Figure 5 where the solid red and blue line of the forecast show a bump around $t \approx 50$ hours. The error plot in Figure 4 depicts also a strong deviation from the true state at that time. After more measurements from the storm are assimilated ($t > 50$ hours), also the forecast output is closer to the true state.

This example clearly shows the full power of the Kalman filter how only a few satellite data can reconstruct the global radiation belt picture (Figure 3). *Naehr and Toffoletto* [2005] compared the “direction insertion method” with the Kalman filter and found that the Kalman filter is far superior.

Before we apply the Kalman filter to estimate the diffusion parameters, we try an example where physical model $M(\alpha, \gamma)$ does not contain the correct diffusion parameters. Figure 6 shows the forecast based on wrong diffusion coefficients. The data, however, was created with the “true” diffusion coefficients α^{true} and γ^{true} . The Kalman filter calculates the best fit between the wrong

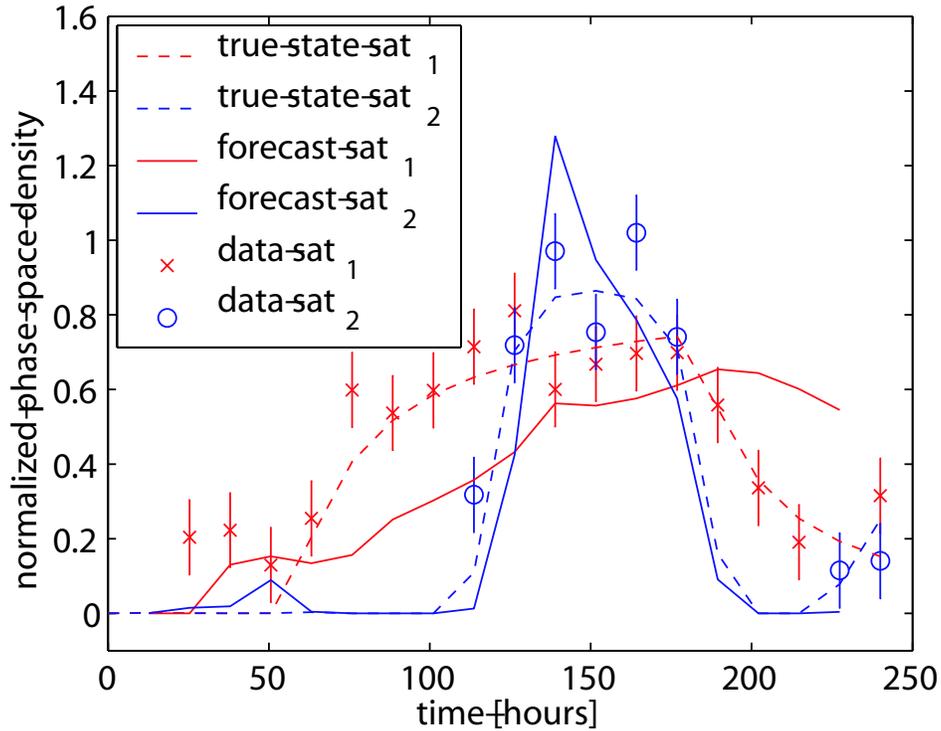


Figure 6. Satellite data, forecast, and true state with two satellites. This plot is the same as Figure 5 but the diffusion parameters in $M(\alpha, \gamma)$ are different compared to what the measurements are based on. The data points were calculated with the “true” diffusion parameters $(\alpha^{\text{true}}, \gamma^{\text{true}}) = (8 \times 10^{-12}, 11.7)$ but the model assumed $(\alpha, \gamma) = (1 \times 10^{-12}, 11.2)$. The forecasts (solid lines) deviate much more from the “true” state evolution (dashed lines) than in Figure 5.

model and the data. The forecast output is far off from the true state because of the mismatch between data based on $(\alpha^{\text{true}}, \gamma^{\text{true}})$ and the model based on a different set of (α, γ) .

This example of data-model combination illustrates again the power of data assimilation as applied to testing assumption in physics-based models. Without the model one would easily be tempted to just connect the data points. By combining data with a model we see that in this case the model cannot represent the data and hence some parameters in the model must be wrong or some of the physics is missing.

In the next section we will investigate if we can use the mismatch between forecast and data to determine the true diffusion parameters using the Kalman filter.

5. Estimating the Diffusion Parameters

5.1. Residual Error Surface

The essence of a model with its parameters is the prediction capability. To estimate the diffusion parameters, we need to evaluate the ability of a model to “describe” the observed data. Hence we introduce the “prediction error” or “innovation vector” \mathbf{d} . Data assimilation methods, including the Kalman filter, update the state estimate by

$$\begin{aligned}\mathbf{x}^a(t_i) &= \mathbf{x}^f(t_i) + \mathbf{K}(t_i) [\mathbf{y}^o(t_i) - \mathbf{H}(t_i)\mathbf{x}^f(t_i)] \\ &= \mathbf{x}^f(t_i) + \mathbf{K}(t_i)\mathbf{d}(t_i)\end{aligned}\tag{14}$$

where the innovation vector

$$\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^f.\tag{15}$$

We use \mathbf{d} in combination with the Kalman gain \mathbf{K} to provide us with the residual on how wrong the predictions are compared to the observations. Rewriting the analysis Equation (14) and transforming it into the observational space by multiplying with $\mathbf{H}(t_i)$ we get

$$\begin{aligned}\boldsymbol{\eta}(t_i) &= \mathbf{H}(t_i) [\mathbf{x}^a(t_i) - \mathbf{x}^f(t_i)] \\ &= \mathbf{H}(t_i)\mathbf{K}(t_i)\mathbf{d}(t_i).\end{aligned}\tag{16}$$

We call the vector $\boldsymbol{\eta}(t_i)$ the “model residual” at time t_i . The dimension of $\boldsymbol{\eta}$ is m which is the number of simultaneous observations. It has the same dimension as $\mathbf{y}^o(t_i)$.

If the absolute value or length of the model residual vector $|\boldsymbol{\eta}(t_i)|$ is large, then we know our choice of diffusion coefficients is unlikely because the prediction from the model is too far off. We evaluate all possible combinations of α and γ to determine which values best fit the data. Recall that the “data” were created with known α^{true} and γ^{true} .

At the end of our assimilation run, we want to have a single number to measure the “goodness” of our diffusion parameters in $\mathbf{M}(\alpha, \gamma)$. We simply sum over all model residuals of n^o measurements between $0 < t < 250$ hours and calculate

$$S_\eta = \sum_{i=1}^{n^o} |\boldsymbol{\eta}|_i.\tag{17}$$

where we sum over the absolute values of all vectors $|\boldsymbol{\eta}|_i$ which combines all measurements from multiple satellites. Since all measurements have a random error ϵ^o , the residual sum S_η has a variance as well. Hence we describe the statistical behavior with $\overline{S}_\eta \pm \Delta S_\eta$ where \overline{S}_η is the statistical mean and ΔS_η is the standard deviation of the residual sum.

Equipped with the formalism for \overline{S}_η , we can now study how different diffusion parameters affect \overline{S}_η . We can test several combination of (α, γ) and calculate the residual. This will give us a “residual surface” in the two dimensional parameter space of α and γ . Ideally we would like the residual error surface $\overline{S}_\eta(\alpha, \gamma)$ to have a well-defined single minimum at $(\alpha, \gamma) = (\alpha^{\text{true}}, \gamma^{\text{true}})$ but as Figure 7 shows instead of a single minimum point, we find a large area of minima. Since this

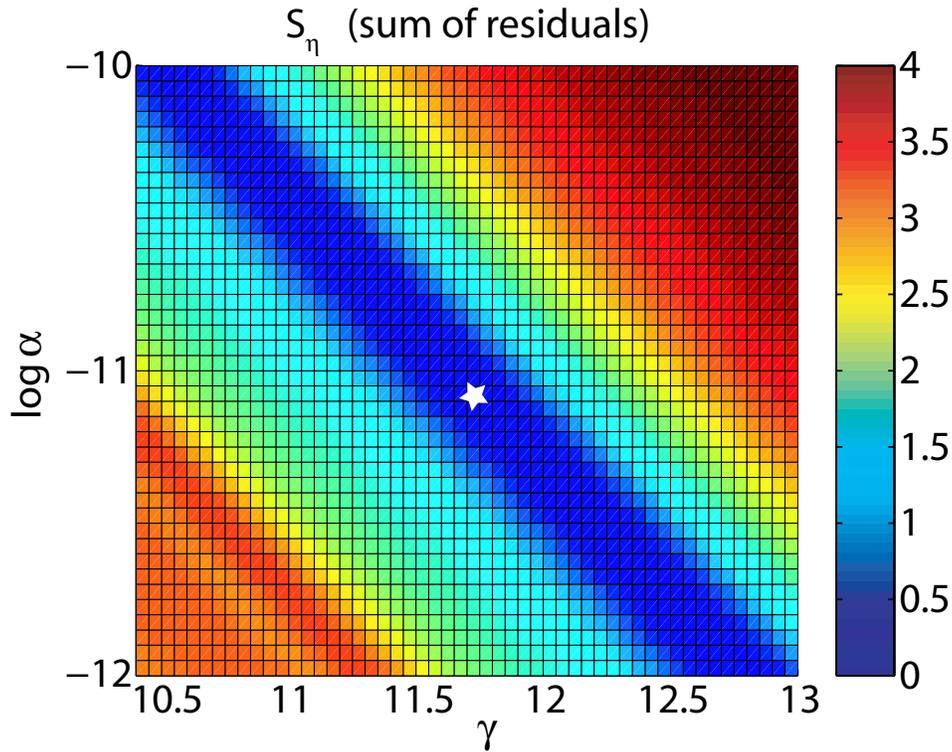


Figure 7. The S_η surface resulting from a parameter study of the (α, γ) -space. The center of the parameter space is the true set of coefficients ($\alpha^{\text{true}} = 8 \times 10^{-12}$, $\gamma^{\text{true}} = 11.7$), denoted by the white star. The figure shows a long stretch of values which we call the “minimum valley”. It can be attributed to diffusion parameters resulting in similar diffusion functions over a large L range.

area has the shape of a shallow valley with very steep slopes on the side, we call it the “minimum valley” from here on.

We interpret the minimum valley the following way: The true \bar{S}_η cannot be found due to the randomness of observational errors. Therefore, all combinations of (α, γ) giving a similar \bar{S}_η in the minimum valley are equally good representatives for the true $(\alpha^{\text{true}}, \gamma^{\text{true}})$. In other words, similar diffusion functions between $4 < L < 8$ can still have very different (α, γ) . We will discuss this in more detail in Section 6.

Obviously there is a problem to determine the true diffusion parameters along the minimum valley because the residuals are so similar. On the other hand, the slopes on the side of the valley are very steep. We can definitely rule out the diffusion coefficients (α, γ) that are perpendicular to the valley.

In the following section we want to investigate what can be done to better determine $(\alpha^{\text{true}}, \gamma^{\text{true}})$ along the minimum valley, i.e., how can we break the degeneracy.

5.2. Residual Along the Minimum Valley

We explained above the cumulative sum of the residual has an uncertainty associated with it. There are two ways to determine the mean residual and the standard deviation: (1) One is to make several runs with different random realizations of the observations, i.e., with a different random number initialization when creating the virtual data. Then the random observational errors will propagate into the residuals S_η . Using all residual from those runs, a mean \overline{S}_η and a standard deviation ΔS_η can be computed.

(2) We choose a different way to avoid multiple, time consuming simulations. When plotting a cumulative sum of each residuals $|\boldsymbol{\eta}(t_i)|$, they form a linear function in time. We calculate a linear fitting function and the standard deviation of each $|\boldsymbol{\eta}(t_i)|$ from that linear approximation. We use the cumulative sum of the linear approximation as our \overline{S}_η and the standard deviations from the linear approximation as ΔS_η .

We discussed above that ΔS_η strongly depends on the observational error σ° . To study that dependence, we plot the residual sum \overline{S}_η including the uncertainty ΔS_η for three different observational variances $(\sigma^\circ)^2$ along the minimum valley (Figure 8). We find that, indeed, with smaller observational variances, the uncertainty in ΔS_η is reduced as well. Also, the residuals \overline{S}_η along the minimum valley become less flat. Hence with better observations we can break the degeneracy of (α, γ) along the minimum valley. This makes sense because a higher fidelity in the data set will have less scattering around the true state. The Kalman filter will use the smaller observational uncertainty σ° to give the observations more confidence than the model.

This translates into the case of real observations as well: higher fidelity data in the identical twin experiment will break the degeneracy in the diffusion parameters. Better observations of the radiation belt will better estimate the physical parameters. However, reality is also more complicated due to the delicate balance of acceleration, sources, and losses.

6. Discussion

The question about the reason for the degeneracy in the residual surface in the (α, γ) -space can be answered by studying the parameterized form of the diffusion function (Equation 2). Since we are only interested in the diffusion function between $4 < L < 8$, we can always find a whole family of diffusion function that resemble the true diffusion very closely. This family of parameter combinations (α, γ) results in similar diffusive behavior that cannot be distinguished if the observational uncertainties are large. If there are measurements outside of $4 < L < 8$ then the family of parameters can be better constrained by excluding certain parameter combinations that do not match the true diffusion function over the whole range anymore.

The key point is that every result is based on data that has an uncertainty. It is important to keep in mind that therefore model parameters, here the diffusion parameters α and γ , have an uncertainty as well. Model parameters cannot be distinguished from the “true” parameters if they are within their uncertainties. One cannot determine different physical processes if the uncertainty of data is too large. The identical twin experiment can help us understand the limits of data uncertainties, i.e., the extend to what we can still gain new insights and learn new physics.

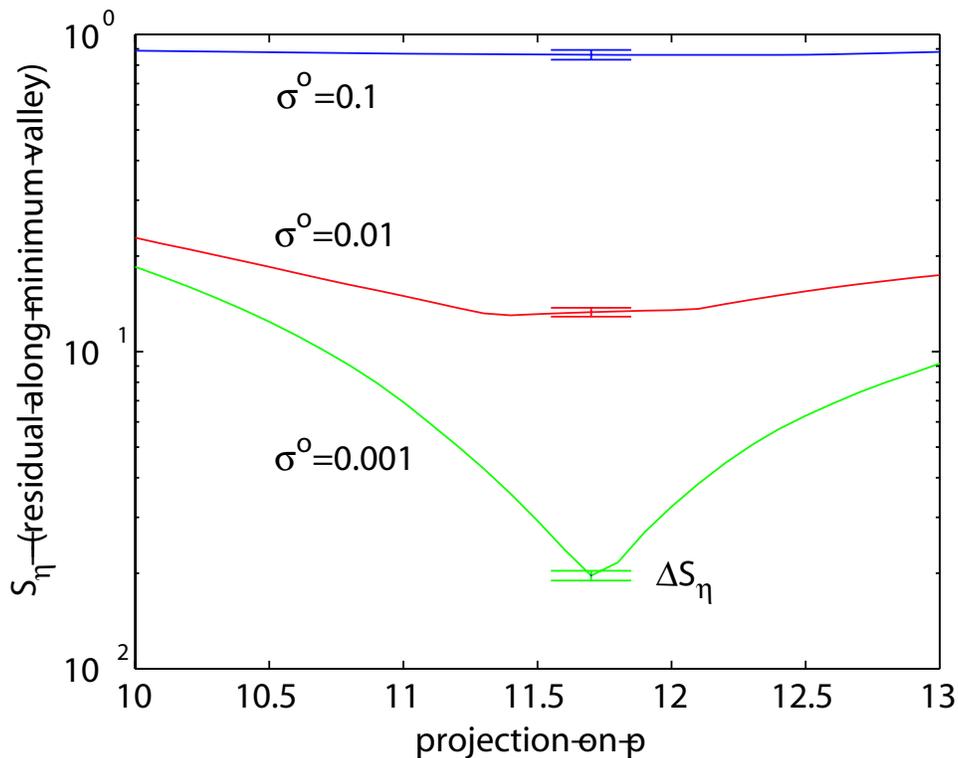


Figure 8. Cumulative residual S_η for three data assimilations with observational error $\sigma^o = 0.1$ (blue), 0.01 (red), 0.001 (green) along the minimum valley as a projection on the the p parameter. The error bar for $\sigma^o = 0.1$ indicates that the minimum cannot be determined because the error ΔS_η is so large and the residual so flat. The error bars for the data assimilations with more trustworthy observations (red, green) are smaller. One can find the diffusion parameters with the minimum residual that are closest to the “true” diffusion with $\gamma^{\text{true}} = 11.7$. This plot also illustrates how the area of degeneracy ϕ is affected by the observational error.

This is an important learning process before we apply the Kalman filter to real data in the DREAM project where we bring the data assimilation method to the next level. The identical twin experiment builds the foundation for the coupled radiation belt and ring current model to better understand limitation and constraints given by data.

Selesnick et al. [1997] describe their best fit with a least square method and include an uncertainty in their result. However, the residual surface plot with the minimum valley shows that the diffusion parameters α and γ are highly anti-correlated: If γ is small, then the likely α is large. The Kalman filter showed us how α and γ are correlated and provided us with a “confidence map”. Using the residual surface we can exclude large regions in the parameter space that are very unlikely to match the true diffusion parameters.

We looked especially at the uncertainty of observations and how an increased confidence in the data can reduce the uncertainty of the diffusion parameters. The minimum valley in the residual

surface plot remains even for high fidelity data, but the slope along the minimum valley becomes steeper and the uncertainty in the residuals becomes smaller as well. Hence we have a better idea of what the most likely diffusion parameters are and we have increased the confidence of these parameters.

We have also studied the dependence of the parameter uncertainties on the measuring frequency and the orbits of the virtual satellites. Naturally, more data and better coverage of the idealized storm increases the confidence of the diffusion parameters even when the data uncertainty remains the same. We made several simulation with varying data quality and find that up to some extend data abundance can make up for data quality. However, increasing the fidelity of data by a factor of two is more efficient than increasing the amount of data by a factor of two. Also, certain types of orbits like geosynchronous transfer orbits are better than geosynchronous orbits. Although many more measurements than the two unknown diffusion parameters are available, one cannot find an exact solution because of uncertainties in the data. The more data we include from more locations, the better the parameters will be determined. If there's only one satellite then one might think the problem is under determined. However, data is combined with the global model and its cross correlations which sets boundaries on the estimated diffusion parameters. A detailed study will follow in a future paper.

In reality, data quality is heavily influenced by the mapping from fluxes to phase space densities because of uncertainties in the geomagnetic field. It is also affected by background radiation, saturated instruments, and calibrations. Therefore, simply adding more data will not necessarily increase the confidence in the diffusion coefficients. A compensating factor is that our identical twin experiment limits knowledge to only that of the simulation. In reality other theoretical and observational studies (such as the proposed Radiation Belt Storm Probes) can provide additional constraints and information.

In any case, the Kalman filter is a valuable tool to properly estimate the diffusion coefficients and their uncertainties. It can use different data sets and implement them into a physics-based model. The Kalman filter automatically considers the uncertainties of model and data and can provide us with a higher confidence of the “best fit”. This can be achieved with a least-square method as well [see *Tarantola and Valette, 1982; Tarantola, 1987*]. We chose the Kalman filter because it can be expanded quite easily for more complex systems and data by using a so-called ensemble method. We will need this expertise for the DREAM project - the combination of radiation belt and ring current models.

7. Conclusion

We studied the combination of a radial diffusion code with a Kalman filter in an identical twin experiment of an idealized geomagnetic storm. We focused especially on parameter estimation of the diffusion function and found that the Kalman filter can be used as an efficient tool to determine the diffusion parameters even with relatively sparse data. We described in detail how the Kalman filter can be applied to the radiation belt problem because this application is relatively new and because it holds great promise for extracting new physical understanding.

The Kalman filter can provide us with a forecast based on previous observations and a physics

based model that we can compare to new measurements. The residual between the Kalman filter forecast and the new observations quantifies the uncertainty of the diffusion parameters.

Currently, we used a parameter space study to locate the range of diffusion coefficients that have the smallest residuals. We found that the residuals mostly depend on the uncertainty of the data but also on the coverage of the storm and the data sampling frequency. High fidelity data lead to a much smaller range of possible diffusion parameters.

We discussed the degeneracy of the parameterized diffusion function and how it leads to a minimum valley in the residual surface. All combination of parameters in the residual valley are from a family of diffusion functions that represent similar physical behavior. The key point of this study is that given data with uncertainties, we can only find a range of diffusion parameters that fit the data equally well. These diffusion parameters are anti-correlated and lie along a straight line in a log-plot as seen in the residual surface (Figure 7).

Eventually the Kalman filter can help us determine what satellites, flight paths, and data quality is required to gain more insight into the physics.

In the future we will apply an adaptive Kalman filter where the dimension of the state vector is extended by the diffusion parameters in a so-called “augmented state vector approach”. The parameters are then recursively estimated by the Kalman filter. This non-linear problem requires a so-called extended Kalman filter or an ensemble Kalman filter. We will also do a more detailed study on the effect of satellite orbits and sampling frequency.

Appendix A: Details on Solving the Diffusion Equation

To advance the model from one observation at time t_i to the next t_{i+1} the Crank-Nicolson scheme ensures numerical stability even when $\Delta t = t_{i+1} - t_i$ is very large. The Crank-Nicolson method provides second order accuracy in both space and time because it averages the spacial derivatives at the beginning and at the end of each time step [Press *et al.*, 1986]. However, very large time steps can still lead to spacial errors that are not negligible.

In our simulation we change the outer boundary to simulate a geomagnetic storm similar to Naehr and Toffoletto [2005]. At the beginning of the storm, we increase the boundary and keep it constant throughout the whole storm-time enabling strong inflow. At the end we return the outer boundary to zero to feature outward diffusion.

Such discontinuous boundary condition in the Crank-Nicolson method can lead to unphysical oscillations. However, positivity is ensured if $\tau/\Delta L \leq 3/2$ [Hundsdoerfer and Verwer, 2003, p. 124] where τ is the time step that is used for the model operator \mathbf{M} . This is for our simulations a factor of ten to one hundred larger than the Courant condition [see Press *et al.*, 1986] would require for explicit methods.

We test for the largest time step τ that still ensures positivity. This time step τ usually will not advance the model far enough to the time of the next observations. However, the forecast model \mathbf{M} can be expressed as the product of intermediate forecast steps, which reflects the causality of nature. The diffusion equation is linear and we can calculate $\tilde{\mathbf{M}} = \mathbf{M}^k$ where \mathbf{M} is a function of τ such that the time between observations $\Delta t = k\tau$.

Appendix B: Error Covariance Matrices

The information about the uncertainties or errors can be combined into an error covariance matrix. In general, the error covariance matrix can be obtained by multiplying an error vector of one set of observations

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix} \quad (\text{B1})$$

by its transpose $\boldsymbol{\epsilon}^T = (\epsilon_1 \ \epsilon_2 \ \cdots \ \epsilon_m)$, and averaging over many cases, to obtain the error covariance matrix

$$\mathbf{R}^o = \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T} = \begin{pmatrix} \overline{\epsilon_1\epsilon_1} & \overline{\epsilon_1\epsilon_2} & \cdots & \overline{\epsilon_1\epsilon_m} \\ \overline{\epsilon_2\epsilon_1} & \overline{\epsilon_2\epsilon_2} & \cdots & \overline{\epsilon_2\epsilon_m} \\ \vdots & \vdots & & \vdots \\ \overline{\epsilon_m\epsilon_1} & \overline{\epsilon_m\epsilon_2} & \cdots & \overline{\epsilon_m\epsilon_m} \end{pmatrix} \quad (\text{B2})$$

The over-bar represents the expected value or statistical mean E . The covariance matrix is positive definite and symmetric. The diagonal elements are the variances of the vector error components $\overline{\epsilon_i\epsilon_i} = \sigma_i^2$. Normalizing, i.e., dividing each component by the product of the standard deviation $\overline{\epsilon_i\epsilon_j}/\sigma_i\sigma_j = \text{corr}(\epsilon_i, \epsilon_j)$, we obtain the correlation matrix. See also *Kalnay* [2003].

Appendix C: Gaussian Error Statistics

Satellite instruments for the radiation belts are usually particle counters. Hence measurement errors have to be described with Poisson statistics for low count numbers [*Wilks*, 1995; *Lyons*, 1986]. However, high flux and large particle count numbers can be approximated with a Gaussian error [*Bevington*, 1969]. Since the linear Kalman filter, in its general form, is based on Gaussian error statistics [*Brammer and Siffing*, 1989], we should only assimilate data that is based on high count rates.

We introduce a cutoff level for a background b below which we do not assimilate data. If a measurement $y_j^o(t_i)$ of satellite j shows a flux level below b , the data is simply ignored. We arbitrarily define that boundary to be equal to the statistical measurement error or standard deviation $b = \sigma^o$. An observational error of $\sigma^o = 0.1$ would correspond to a factor of two at the lower flux levels $x^{\text{true}} \approx 0.1$. This is the same order of magnitude reported by *Friedel et al.* [2005, and priv. comm.]. However, we conduct several tests with $\sigma^o = \bar{\epsilon}^o \in [0.001, 0.1]$ simulating different ranges of data quality.

References

- Bevington, P. (1969), *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.
- Boscher, D., S. Bourdarie, and T. Beutier (1996), Dynamic modeling of trapped-particles, *IEEE Trans. Nuclear Sci.*, 43, 416–425.

- Bourdarie, S., D. Boscher, T. Beutier, J. A. Sauvaud, and M. Blanc (1996), Magnetic storm modeling in the Earth's electron belt by the salammbô code, *J. Geophys. Res.*, *101*, 27,171–27,176.
- Bourdarie, S., R. H. W. Friedel, J. Fennell, S. Kanekal, and T. E. Cayton (2005), Radiation belt representation of the energetic electron environment: Model and data synthesis using the salammbô radiation belt transport code and Los Alamos geosynchronous and GPS energetic particle data, *Space Weather*, *3*, S04S01, doi:10.1029/2004SW000065.
- Bouttier, F., and P. Courtier (1999), Data assimilation concepts and methods.
- Brammer, K., and G. Siffing (1989), *Kalman-Bucy Filters*, Artech House, Norwood.
- Crank, J., and P. Nicolson (1947), A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type, *Proceedings of the Cambridge Philosophical Society*, *43*, 50–67.
- Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge University Press, Cambridge.
- Friedel, R., S. Bourdarie, and T. Cayton (2005), Inter-calibration of magnetospheric energetic electron data, *Space Weather*, p. in press.
- Friedel, R. H. W., G. D. Reeves, and T. Obara (2002), Relativistic electron dynamics in the inner magnetosphere – a review , *J. Atmos. Terr. Phys.*, *64*, 265–82.
- Hundsdoerfer, W., and J. Verwer (2003), *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, Springer, Berlin.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc (1997), Unified notation for data assimilation: Operational, sequential and variational, *Journal of the Meteorological Society of Japan*, *75*(1B), 181 – 189.
- Kalman, R. E. (1960), New approach to linear filtering and prediction problems, *American Society of Mechanical Engineers – Transactions – Journal of Basic Engineering Series D*, *82*(1), 35 – 45.
- Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge.
- Lyons, L. (1986), *Statistics for Nuclear and Particle Physicists*, Cambridge University Press, Cambridge.
- Lyons, L. R., and R. M. Thorne (1973), Equilibrium structure of radiation belt electrons, *Journal of Geophysical Research*, *78*, 2142–2149.
- Maybeck, P. (1979), *Stochastic Models, Estimation, and Control - Volume 1*, Academic Press, Orlando, Florida.
- McCoy, R. P. (2005), Ionospheric data assimilation: A new paradigm, *The CEDAR Post*, *50*, 8.
- Naehr, S., and F. Toffoletto (2005), Radiation belt data assimilation with an extended kalman filter, *Space Weather*, *3*(6), S06,001, doi:10.1029/2004SW000121.
- Petersen, D., and D. Middleton (1963), On representative observations, *Tellus*, *15*, 384 – 405.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1986), *Numerical Recipes - The Art of Scientific Computing*, 1st ed., Cambridge Univ. Press, Cambridge, United Kingdom.

- Reeves, G. D., K. L. McAdams, R. H. W. Friedel, and T. P. O'Brien (2003), Acceleration and loss of relativistic electrons during geomagnetic storms, *Geophys. Res. Lett.*, *30*, 1529.
- Reeves, G. D., et al. (2005), Toward understanding radiation belt dynamics, nuclear explosion-produced artificial belts, and active radiation belt remediation: Producing a radiation belt data assimilation model, in *Global Physics of the Coupled Inner Magnetosphere*, edited by J. L. Burch, Geophys. Monogr. Ser., pp. –, AGU, Washington, D.C.
- Richmond, A. D., and Y. Kamide (1988), Mapping electrodynamic features of the high-latitude ionosphere from localized observations: technique., *Journal of Geophysical Research*, *93*(A6), 5741 – 59.
- Roederer, J. G. (1970), *Dynamics of Geomagnetically Trapped Radiation*, Springer-Verlag, New York.
- Schulz, M., and L. J. Lanzerotti (1973), *Particle Diffusion in the Radiation Belts*, 1st ed., Springer-Verlag, New York.
- Schulz, M., and L. J. Lanzerotti (1974), *Particle Diffusion in the Radiation Belts*, Springer-Verlag, Berlin.
- Schunk, R. W., et al. (2004), Global assimilation of ionospheric measurements (GAIM), *Radio Science*, *39*(1), 211 – 221, doi:10.1029/2002RS002794.
- Selesnick, R., J. Blake, W. Kolasinski, and T. Fritz (1997), A quiescent state of 3 to 8 mev radiation belt electrons, *Geophys. Res. Lett.*, *24*, 1,343–1,346, doi:10.1029/97GL51407.
- Sorenson, H. (Ed.) (1985), *Kalman Filtering: Theory and Application*, IEEE Press, New York.
- Talagrand, O. (1997), Assimilation of observations, an introduction, *Journal of the Meteorological Society of Japan*, *75*(1B), 191 – 209.
- Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Science Publisher, Amsterdam.
- Tarantola, A., and B. Valette (1982), Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion, *Reviews of Geophysics and Space Physics*, *20*, 219–232.
- Vette, J. I. (1991), The NASA/National Space Science Data Center Trapped Radiation Environment Model Program (1964–1991), *Tech. Rep. NSSDC/WDC-A-R & 91-26*, Goddard Space Flight Center, NASA/GSFC, Greenbelt, USA.
- Voss, H. U., J. Timmer, and J. Kurths (2004), Nonlinear dynamical system identification from uncertain and indirect measurements., *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, *14*(6), 1905 – 33.
- Wilks, D. (1995), *Statistical Methods in the Atmospheric Sciences*, Academic Press, San Diego.

J. Koller, R.H.W. Friedel, and G.D. Reeves, Space Science and Applications, Los Alamos National Laboratory, P.O. Box 1663, MS D466, Los Alamos, NM 87545, USA. (jkoller@lanl.gov, friedel@lanl.gov, reeves@lanl.gov)