Log-normal Kalman filter for assimilating phase-space density data in the radiation belts

D. Kondrashov,[1] M. Ghil[1,3] and Y. Shprits,[1,3]

D. Kondrashov, Department of Atmospheric and Oceanic Sciences, 405 Hilgard Ave, Box 951565, 7127 Math Sciences Bldg. UCLA Los Angeles, CA 90095-1565, U.S.A. (dkondras@atmos.ucla.edu)

[1]Department of Atmospheric and Oceanic Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, U.S.A.

[2]Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and IPSL), Ecole Normale Supérieure, F-75231 Paris Cedex 05, FRANCE.

[3]Department of Earth and Space Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, U.S.A.

Abstract. Data assimilation combines a physical model with sparse observations, and has become an increasingly important tool for scientists and engineers in the design, operation and use of satellites and other high-technology systems in near-Earth's space environment. Of particular importance is predicting fluxes of high-energy particles in the Van Allen radiation belts, since these fluxes can damage space-borne platforms and instruments during strong geomagnetic storms. In transiting from a research setting to operational prediction of these fluxes, improved data assimilation is of the essence. The present study is motivated by the fact that phase-space densities (PSDs) of high-energy electrons in the outer radiation belt — both simulated and observed — are subject to spatio-temporal variations that span several orders of magnitude. Standard data assimilation methods that are based on least-squares minimization of normally distributed errors may not be adequate for handling the range of these variations. We propose herein a modification of Kalman filtering that uses a log-transformed, one-dimensional radial diffusion model for the PSDs and includes parameterized losses. The proposed methodology is first verified on model-simulated, synthetic data and then applied to actual satellite measurements. When the model errors are sufficiently smaller then observational errors, our methodology can significantly improve analysis and prediction skill for the PSDs compared to those of the standard Kalman filter formulation. This improvement is documented by monitoring the variance of the innovation sequence.

1. Introduction and Motivation

1.1. Data assimilation and operational prediction

In the process of moving from research to operations in the study of the Van Allen radiation belts, it is of the essence to properly understand and further improve data assimilation methodology — as applied to the filtering, smoothing and prediction of electron fluxes and phase space density (PSD) fields. The modern uses of data assimilation in the geosciences go back to the introduction of meteorological satellites and their application to numerical weather prediction (NWP) in the late 1960s and the 1970s (*Charney et al.* [1969]). It was *Bjerknes* [1904], following the earlier ideas of H. Helmholtz and others, who formulated the NWP problem as an initial-value problem for the partial differential equations that govern large-scale atmospheric flow.

As soon as the research group around J. von Neumann at the Institute for Advanced Studies in Princeton started working on experimental NWP (*Charney et al.* [1950]), it became apparent that the initial state of the atmosphere at any given time was known only very partially and inaccurately. The World Weather Watch introduced by the World Meteorological Organization after World War II was designed to provide as good a state of the atmosphere as possible twice a day, at noon and midnight Greenwich Mean Time, the so-called synoptic times. These synoptic (i.e., simultaneous) observations, however, were too costly or impractical to provide adequate coverage of the weather over the entire globe. The advent of asynchronous observations, via satellites and other unconventional observing platforms and instruments, sharpened the need for the time-continuous, rather than intermittent, blending of observations and models (*Ghil et al.* [1979]; *Bengtsson et al.* [1981]) .

To better understand this new point of view, consider a sequence of observations at discrete times $\{t_k : t_0 \leq t_k \leq t_K\}$ of a scalar or vector variable $x(t_k)$ or $\mathbf{x}(t_k)$. The vector variable $\mathbf{x}(t_k)$ will represent the spatially discretized values of a geophysical field, such as temperatures in NWP or PSD values in the radiation belts. *Wiener* [1949] defined filtering, smoothing and prediction of this variable $\mathbf{x}(t_k)$ as its estimate at: (i) the final observing time t_K; (ii) at all t_k over the observation interval $t_0 \leq t_k \leq t_K$; and (iii) at any time after the final observation, $t_K < t_k$.

In the real-time prediction problem, it is pretty easy to convince oneself that — under fairly general hypotheses on the process to be predicted and given observations up to time t_K — the best use one can make of the observations is to estimate as well as one can, with the knowledge one has, the state at the initial prediction time, i.e., at t_K. This is precisely how the so-called forecast-assimilation cycle proceeds in NWP; such an operational NWP cycle is illustrated in Figs 1(a,b).

Panel (a) of the figure shows the traditional blending of data and model, at the synoptic times, used from the beginnings of data assimilation in the 1960s until the late 1970s and early 1980s; such data windows are still used in so-called variational methods of data assimilation (e.g., *Courtier and Talagrand* [1987]). Panel (b) outlines the more recent approach, in which data are assimilated at any model time step at which they become available; this sequential approach includes a great variety of methods that generally fall these days in the broad class of Kalman-type filters (e.g., *Jazwinski* [1970] or *Gelb* [1974]).

In space physics, given the total absence — at any given time of day or night — of synoptic data that cover the entire domain of interest, it is natural to start relying on the time-continuous approach of data assimilation. This approach can address two types of applications that are of primary interest for the satellite design and operations community: nowcasting and short-term forecasting, as well as long-term reanalysis. For both types of applications, be it in a research or operational mode, data should be assimilated from the operating space platforms and instruments, when and where they become available.

The nowcasting applications help address issues linked to the state of the space environment's radiation properties at a given time and location, and thus provide post facto insight into the possible causes of particular anomalies. Moreover, a satellite operator could take preventive action, based on a reliable short-term forecast of the space environment, if a satellite is threatened; given the current lack of such reliable forecasts, such action is not a widespread practice at this time.

In a research mode, one can also consider the smoothing problem, which produces "movies" of the plasma properties, particle distribution functions, the magnetic and electric fields or the wave environment over the entire lifetime of a satellite or of a group of spacecraft, which may last over several solar cycles. Such a movie can help determine the average state and extreme conditions in a certain part of the space environment, and can be turned into satellite specifications. Our proposed improvement of assimilation methodology should thus help both operational and research aspects of space physics by providing better estimates of the radiation environment whenever observations are available.

1.2. The need for a log-density formulation

A striking feature of the radiation belts is that values of observed electron fluxes and modeled PSD vary by several orders of magnitude, and that the corresponding error distributions are, therefore, not Gaussian. Still, standard data assimilation methods —

such as the Kalman filter and its various adaptations to large-dimensional and nonlinear problems — are essentially based on least-squares minimization of Gaussian errors.

Even though this mismatch between the nature of the data and that of the method leads to substantial problems when applying standard assimilation methods to the radiation belts, there have been very few investigations to address these issues; we cite here the two that we are aware of: first, *Naehr and Toffoletto* [2005] relied on a log-based transformation of the PSD and on an extended Kalman filter to study sequential filter performance on synthetic data. Next, *O'Brien and Guild* [2010] proposed a variational data assimilation method based on a Maximum Likelihood Ensemble Filter (MLEF, *Zupanski* [2005]) that also uses a log-based transformation for both the measurements and the model state vector.

In the spirit of these two studies, we explore alternative ways to make Kalman-filter–type methods more efficient for use in PSD assimilation for the radiation belts by relying on an one-dimensional (1-D) version of the UCLA Versatile Electron Radiation Belt (VERB) diffusion model (*Shprits et al.* [2005], *Subbotin and Shprits* [2008]; see also Sec. 2 herein) and on observations from multiple satellites (Sec. 3). We introduce a log-normal PSD transformation in the UCLA VERB 1-D code to derive an analytical model equation for the transformed variable (Sec. 5) and use it in our extended Kalman filter formulation of Sec. 4.

First, we analyze the performance of the log-normal Kalman filter so derived on synthetic data in the "fraternal-twin" experiments of Sec. 6.1, and then apply it to spacecraft PSD measurements in Sec. 6.2. We conclude in Sec. 7 by analyzing under which conditions this log-normal formulation improves the accuracy of the reanalysis and prediction of the PSD field, as inferred from the variance of the Kalman filter's innovation sequence. The main factor influencing the performance of the log-normal filter is the ratio of the model error to the observational error. When this ratio is sufficiently small but not negligible, the log-normal formulation produces a more efficient modification of the model forecast in observation-void regions and better prediction, too.

2. The UCLA VERB Code

The version of the UCLA VERB code that we use here provide a 1-D description of the time evolution of the PSD $f = f(L^*, t; \mu, J)$ in the Van Allen radiation belts, at fixed values of the adiabatic invariants μ and J. The radial variable L^* is the distance — in the equatorial plane, measured in Earth radii R_{E} — from the center of the Earth to the magnetic field line around which the electron moves at time t, assuming that the instantaneous magnetic field is adjusted adiabatically to a pure-dipole configuration. In this study, the *Tsyganenko* [1989] T89 magnetic field model has been used to derive electron fluxes at a particular L^*-value. For simplicity from now on in the text and figures we drop the superscript and refer to this variable simply as L: both the radiation belt model and all satellite data are computed in L^*.

The PSD evolution in time is then governed by the following parabolic partial differential equation [*Shultz and Lanzerotti*, 1974; *Walt*, 1994] :

$$\frac{\partial f}{\partial t} = L^2 \frac{\partial}{\partial L}(L^{-2} D_{LL} \frac{\partial f}{\partial L}) - \frac{f}{\tau_L}. \quad (1)$$

The radial diffusion term in Eq. (1) represents the violation of the third adiabatic invariant, while the net effect of sources and losses due to violations of the μ and J invariants is modeled in this equation by the linear decay with a characteristic lifetime τ_L.

The parameters D_{LL} and τ_L in Eq. (1) vary rapidly in space and time, and depend on the background plasma density, as well as on the spectral intensity and spatial distribution of various plasma waves; all of these conditions are extremely difficult to specify accurately from limited point measurements. In this study, we adopt a commonly used empirical relationship due to *Brautigam and Albert* [2000] which is based on the magnetic field measurements at $L = 4$ *Lanzerotti and Morgan* [1973] and $L = 6.6$ *Lanzerotti et al.* [1978], between the radial diffusion coefficient D_{LL} and the geomagnetic activity index Kp:

$$D_{LL}(Kp, L) = 10^{(0.506Kp-9.325)} L^{10}; \tag{2}$$

this equation applies throughout the outer radiation belt.

For the lifetime parameter τ_L, we consider different parameterizations inside and outside the plasmasphere. The latter is a region of the inner magnetosphere that contains relatively cool and dense plasma at low energies; it is populated by the outflow of ionospheric plasma along the magnetic field lines, and consists of closed equipotential surfaces. The plasmapause that separates it from the regions of open equipotential surfaces lies, under quiet conditions, within the outer belt, at $L_{\mathrm{PP}} = 5 - 6R_{\mathrm{E}}$, where R_{E} is the Earth's radius. Under quiet conditions, the outer belt lies at about $3.5-6R_{\mathrm{E}}$ and the inner belt at about $1 - 2.5R_{\mathrm{E}}$, starting just above the ionosphere. Magnetospheric storms deplete the plasmasphere and L_{PP} can sink to 3 or, for particularly strong storms, even $2R_{\mathrm{E}}$, [*Baker et al.*, 2004].

As described in *Kondrashov et al.* [2007], distinct loss processes operate inside and outside of the plasmasphere, and so we account for them separately in the physical model. Inside we assume that $\tau_{LI} = 10$ days is constant in time, while outside we take

$$\tau_{LO} = \zeta / Kp(t). \qquad (3)$$

To discretize numerically Eq. (1), we use standard second-order centered difference approximations for spatial derivatives. We also utilize a fully implicit numerical method to advance the solution in time, which allows one to use much larger time steps than explicit schemes do.

3. Spacecraft Observations

This study covers a time interval of 120 consecutive days that starts on July 30, 1990, and includes measurements from four space missions: the Combined Release and Radiation Effects Satellite (CRRES), GEO-1989 (hereafter referred to as GEO), GPS NS18 (hereafter GPS), and Akebono. To perform data assimilation of the PSD distribution derived from the electron flux observations measured by the various spacecraft instruments, we need first to calculate the PSD in the appropriate phase-space coordinates (μ, K, L); here μ is the first adiabatic invariant, while K is a combination of the first two adiabatic invariants that is independent of the particle mass and charge.

The Kp data set is taken from the World Data Center for Geomagnetism in Kyoto, Japan, http://swdcdb.kugi.kyoto-u.ac.jp/aedir/. The T89 model is specified by the Kp-value and is valid only for relatively modest activity levels. Recently, *Ni et al.* [2009] have compared and mutually calibrated PSD data from the CRRES Medium Electron A (MEA) observations and those from the polar-orbiting Akebono Radiation Monitor

(RDM) by using the T89 model; they found in general good agreement between the PSD values inferred from the two sets of observations.

Recent, improved models of the magnetic field include parameterizations that use also *Dst* and solar-wind measurements. Though the latter are not generally available for the CRRES time period, *Kondrashov et al.* [2010] showed that Singular Spectrum Analysis can be used to fill in large gaps in past solar-wind and IMF data.

4. The Extended Kalman Filter (EKF)

In this section, we review the Kalman filter as applied to data assimilation in the radiation belts, following *Kondrashov et al.* [2007], *Shprits et al.* [2007], *Ni et al.* [2009], *Koller et al.* [2007] and *Daae et al.* [2011]. The summary here uses the filter's presentation for partial differential equations in the geosciences, as introduced by *Ghil et al.* [1981] and reviewed by *Ghil and Malanotte-Rizzoli* [1991]; in this presentation, both time and space have been discretized by finite differences.

The time evolution of the state vector $\mathbf{x}_k^{\mathrm{f,t}} = x^{\mathrm{f,t}}(k, \Delta t)$ is assumed to be governed by the numerically discretized system of equations whose right-hand side (RHS) is denoted by $\mathbf{F} = \mathbf{F}(\mathbf{x})$; here superscript "t" refers to *true*, "f" refers to model *forecast*, and k is a discretized time index. If the system is nonlinear, $\mathbf{F} = \mathbf{F}(\mathbf{x})$ has to be linearized to yield the model matrix $\mathbf{M}$ that will be used in advancing the forecast error covariances. Furthermore, the true state differs from the model forecast by a random error ϵ^{m}:

$$\mathbf{x}_k^{\mathrm{f}} = \mathbf{F}_k(\mathbf{x}_{k-1}^{\mathrm{f}}), \tag{4}$$

$$\mathbf{x}_k^{\mathrm{t}} = \mathbf{F}_k(\mathbf{x}_{k-1}^{\mathrm{t}}) + \epsilon_k^{\mathrm{m}}, \tag{5}$$

$$\mathbf{M} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}}. \tag{6}$$

For the radiation belt model of Eq. (1), the components of the state vector $\mathbf{x}_k$ are the PSD values at the discretized grid points in the independent variable L; since the partial differential operator in (1) is linear, no linearization, as in Eq. (5), seems to be required. In the next section, though, we will encounter a nonlinear version of this equation, thus justifying the use of the full set of Eqs. (4)–(6).

The model noise ϵ^{m} accounts for the net errors due to inaccurate model physics, such as errors in forcing, boundary conditions, numerical discretization, and subgrid-scale processes. Commonly, the column vector ϵ^{m} is assumed to be a Gaussian white-noise sequence, with mean zero and model-error covariance matrix $\mathbf{Q}$, $E[\epsilon_k^{\mathrm{m}}] = 0$ and $E[\epsilon_k^{\mathrm{m}} \epsilon_l^{\mathrm{mT}}] = \mathbf{Q}_k \delta_{kl}$, where E is the expectation operator, superscript "T" denotes the transpose, and δ_{kl} is the Kronecker delta.

The observations $\mathbf{y}_k^{\mathrm{o}}$ of the true system, where superscript "o" refers to "observed," are also perturbed by Gaussian white noise ϵ_k^{o} with mean zero and given covariance matrix $\mathbf{R}$, $E[\epsilon_k^{\mathrm{o}} \epsilon_l^{\mathrm{oT}}] = \mathbf{R}_k \delta_{kl}$:

$$\mathbf{y}_k^{\mathrm{o}} = \mathbf{H}_k \mathbf{x}_k^t + \epsilon_k^{\mathrm{o}}. \tag{7}$$

The observation matrix $\mathbf{H}_k$ accounts for the fact that usually the dimension of $\mathbf{y}_k^{\mathrm{o}}$ is less than the dimension of $\mathbf{x}_k^{\mathrm{t}}$, i.e., at any given time observations are not available for all grid points. It is often also the case that the values of the PSD or other state variable are not directly observable, and it is only some linear or nonlinear combination of such variables, such as weighted integrals over phase space, that can be measured.

The Kalman filter and its variants are sequential data assimilation methods. For given model and observation error covariances, $\mathbf{Q}$ and $\mathbf{R}$, the filter combines the model forecast with the observations so as to obtain the *analysis* that is closest in a least-square sense to the truth. The gain matrix $\mathbf{K}_k$ in Eq. (8) represents the *optimal* weights given to the observations in updating the model forecast, based on this least-square minimization:

$$\mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{f}} + \mathbf{K}(y_k^{\mathrm{o}} - \mathbf{H}\mathbf{x}_k^{\mathrm{f}}), \tag{8}$$

$$\mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^{f}\mathbf{H}^{T} + \mathbf{R})^{-1}, \tag{9}$$

$$\mathbf{P}_k^{\mathrm{f}} = \mathbf{M}_k\mathbf{P}_{k-1}^{\mathrm{f}}\mathbf{M}_k^{\mathrm{T}} + \mathbf{Q}, \tag{10}$$

$$\mathbf{P}_k^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_k^{\mathrm{f}}. \tag{11}$$

The error-covariance matrices $\mathbf{P}^{\mathrm{f,a}}$ are the time-dependent error estimates for the forecast and the analysis, respectively. One expects, from Eqs. (10) and (11), that the analysis error be smaller than the forecast error, cf. *Ghil et al.* [1981] and *Carrassi et al.* [2008]. The sequential estimator for nonlinear systems that uses the linear dynamics operator $\mathbf{M}_k$ in the quadratic equation (10) for advancing the model covariances $\mathbf{P}_k^{\mathrm{f}}$ in time, while preserving the nonlinear evolution (4) of the state itself, is called the *extended Kalman filter* (EKF: *Jazwinski* [1970], Gelb [1974], Miller et al. [1994]). To estimate poorly known parameters of the system, Kalman filter can be applied to state vector augmented with the parameters values (*Kondrashov et al.* [2007,2008]).

It is logical to assume that the PSD is log-normally distributed since it is always positive, and generally its variations — as measured, for instance, by the standard deviation — increase as its mean value increases. Normally distributed variables, on the other hand, can be negative and have a standard deviation that does not change as the mean changes. Log-normal errors arise when variation sources accumulate multiplicatively, whereas normal errors arise when these sources are additive [*Crow and Shimizu*, 1988].

By assuming a log-normal distribution of errors at each location and errors that are uncorrelated between different locations, both $\mathbf{Q}$ and $\mathbf{R}$ can be specified as diagonal matrices, and their diagonal terms can be taken simply as $\alpha_{\mathrm{o,m}} f_{\mathrm{o,m}}^2$, where $f_{\mathrm{o,m}}^2$ is the observed or modeled PSD value, and $\alpha_{\mathrm{o,m}}$ is a specified factor that corresponds to observational or model error. Note that the exact values of $\alpha_{\mathrm{o,m}}$ are not important: it is their respective ratio that determines the weights given to the observations vs. the model solution in the analysis, or update, step of the data assimilation. In this study, we follow approach of *Ni et al.* [2009] with the value of α_{o} depending on the intercalibration of satellite data; we use $\alpha_{\mathrm{o}} = 200$ for Akebono and $\alpha_{\mathrm{o}} = 400$ for GEO, while $\alpha_{\mathrm{m}} = 25$.

Due to their reliance on least-squares minimization, the EKF and other Kalman-type filters may not be efficient in modifying the model forecast at locations where the observations and forecast differ by several orders of magnitude. In the next section (Sec. 5), we use a log-normal transformation of variables to derive the model equation for $\log(f)$ and the corresponding EKF. We then study in Sec. 6.1 the performance of the log-normal EKF for synthetic data in "fraternal-twin" assimilation experiments — when the true evolution of the system is known — and next, in Sec. 6.2, for actual space-borne observational data.

5. Log-Normal Model and Filter

By introducing the new variable $S = \log(f)$ in Eq. (1) and using the chain rule for partial derivatives in time and space, one can easily obtain the following evolution equation for the log-transformed PSD:

$$\frac{\partial S}{\partial t} = L^2 \frac{\partial}{\partial L}\left(\frac{1}{L^2} D_{LL} \frac{\partial S}{\partial L}\right) - \frac{1}{\tau_L} + D_{LL}\left(\frac{\partial S}{\partial L}\right)^2. \tag{12}$$

The first two terms on the RHS of the log-transformed Eq. (12) correspond to radial diffusion and losses respectively, as in Eq. (1). Even though the original Eq. (1) is linear, the last term on the RHS of Eq. (12), $(\partial S/\partial L)^2$, is due to the nonlinear transformation of variables and requires special attention in the numerical solution. This term becomes important in locations where strong PSD gradients occur; it can be understood as nonlinear advection of S with a velocity that depends on its gradient.

Note that our methodology is distinctly different from that of *Naehr and Toffoletto* [2005], where a log-normal transformation is applied to the *numerically discretized* equation for f. We derive instead the analytical equation (12) for the evolution of $S = \log(f)$; this equation does not depend on the details of a particular numerical scheme for solving the f-equation (1).

Note that the log-transformed Eq. (12) and the original equation (1) should both yield the same solution $f(t, L)$ — to within the accuracy of the spatial discretization and time integration scheme — when solved numerically. The numerical solution of the original Eq. (1) is typically a smooth monotone function in space, but "naive" approximation in space of the quadratic term in Eq. (12) — for example by using centered differences — will result in spurious local extrema when integrated numerically. This Gibbs phenomenon ultimately leads to unstable solutions and — in order to avoid such numerical instabilities and preserve the monotonicity of the solutions of the log-transformed Eq. (12) — we use an upwind approximation of $(\partial S/\partial L)^2$ that is second-order in space and total-variation diminishing (TVD: *Harten* [1983]). The TVD scheme, when combined with the implicit time integration, guarantees stable numerical solutions of Eq. (12). To solve numerically either Eq. (1) or Eq. (12), we use a uniform grid of 100 points in L; the number of grid points also determines the dimension of the state vector in the Kalman filter formulation.

Steep gradients are key features of the radiation belts, and there is therefore an additional improvement in numerical performance obtained by recasting the diffusion problem of Eq. (1) in terms of the transformed variable $S = \log(\mathrm{PSD})$: doing so helps avoid non-physical negative values, which may arise in numerical schemes that solve Eq. (1) in the original variable $f = \mathrm{PSD}$. There is of course a trade-off in difficulty, since the log-normal model requires one to solve the nonlinear Eq. (12); this will also present a special challenge when applying the proposed methodology to the 3-D VERB code (*Subbotin and Shprits*, 2008) that describes diffusion in energy, pitch angle and L.

Since Eq. (12) is nonlinear in S, we also need to linearize the model in order to implement the EKF of Eqs. (4)–(11). Moreover, for the EKF implementation, observational and model errors for f are modified in a manner appropriate for log-transformed variables by setting the diagonal elements $\mathbf{Q}$ and $\mathbf{R}$ equal to $\log(1 + \alpha_{\mathrm{m}})$ and to $\log(1 + \alpha_{\mathrm{o}})$, respectively [*Crow and Shimizu*, 1988].

6. Numerical Results

6.1. "Fraternal-twin" experiments

To compare the log-normal EKF scheme of Sec. 5 with the standard EKF implementation of Sec. 4, we conduct so-called "fraternal-twin" experiments in which both the "true" solution, from which the observations are drawn, and the forecast are produced by the same model, but with different values of the lifetime parameters in Eqs. (1) and (12). This type of experiment is a harder test for a given assimilation method than a so-called "identical-twin" experiment, in which the model used for the assimilation of partial data is identical to the one used to generate the data, and only the initial state may differ.

We obtain our true PSD distribution from a model run with $\tau_{LI} = 10$ days and $\zeta = 5$ days, cf. Eq. (3) and Fig. 2a; this run is also used to generate synthetic observations along the tracks of the GPS and GEO satellites with a 10-min resolution, as plotted in Fig. 2c. Our goal is to recover the true solution by assimilating these observations into a model simulation with an "incorrect" set of parameters, equal to $\tau_{LI} = 10$ days and $\zeta = 1$ day; these values correspond to higher losses, as shown in Fig. 2b.

The results of assimilating the synthetic data from Fig. 2c by applying the standard EKF formulation are plotted in Fig. 3a. The plot shows that even though assimilating this data set drives the model forecast with the wrong parameter values towards the true model's solution, significant differences remain, cf. Fig. 3c. These differences are largest for $4 \leq L \leq 6$, i.e. in the heart of the outer radiation belt, where the PSD gradients are strongest in Fig. 2b.

When using the log-normal EKF of Sec. 5, on the other hand, our data assimilation reduces the model forecast error much more efficiently in the region of strong PSD gradients. This can be clearly seen by comparing Figs. 3b and 3d with the preceding Figs. 3a and 3c; it is also confirmed by the time-averaged analysis error in PSD values, for the standard EKF formulation and the log-normal one, as shown in Fig. 4b.

Since both the original model equation (1) and the log-transformed equation (12) yield identical solutions for the PSD field in the absence of data assimilation, the difference in the results can only be due to the change in the data assimilation scheme, as outlined and explained in Sec. 5.

In addition to comparing the analysis errors, as discussed above, another useful and readily available means for assessing an assimilation scheme is studying the *innovation sequence*:

$$\mathbf{z}_k \equiv \mathbf{y}_k^{\mathrm{o}} - \mathbf{H}\mathbf{x}_k^{\mathrm{f}}, \tag{13}$$

which appears in the updating equation (8). The importance of considering the innovations $\mathbf{z}_k$ in sequential estimation — i.e., the time series of the differences between the observations and the model forecast — was noted already by *Kailath* [1968] and was emphasized recently by *Fukumori* [2006] in the geophysical context.

The innovation vector represents the filter's correction to the model dynamics. For a linear system with known coefficients and known noise covariances, the *innovation property* of the Kalman filter states that the innovation sequence has zero mean and is white in time, i.e., $E[\mathbf{z}_k^{\mathrm{T}}\mathbf{z}_l] = 0$ for $k \neq l$; this means simply that the filter extracts, at each time step, any and all useful information from the observations. *Dee et al.* [1985], for instance, have used systematically deviations from this property to infer unknown error covariances $\mathbf{Q}$ and $\mathbf{R}$ in the shallow-water model of *Ghil et al.* [1981]. In the space-plasma context, *Koller et al.* [2007] and *Shprits et al.* [2007] showed how the nonzero time mean of the innovation sequence can point to missing physics in the model's competing processes of losses and local acceleration.

Here we show how inspection of the innovation sequence can be used to diagnose the performance of a data assimilation scheme, even in the case of a nonlinear problem, like Eq. (12). As the forecast $\mathbf{x}_k^{\mathrm{f}}$ does not yet utilize the upcoming observations $\mathbf{y}_k^{\mathrm{o}}$, the variance of the sequence $\mathbf{z}_k$ provides an objective measure of *prediction skill* with respect to independent observations: a perfect model would predict exactly the incoming observation vector, so that $\mathbf{y}_k^{\mathrm{o}} = \mathbf{H}\mathbf{x}_k^{\mathrm{f}}$, while a particularly poor model might differ from the observations more than their long-term mean, i.e., than the climatology of the system.

The prediction skill is thus defined here as the variance $E[\mathbf{z}_k^{\mathrm{T}}\mathbf{z}_k]$ and it is plotted in Fig. 4a for both the standard EKF formulation(red curve) and the log-normal one (blue curve); for the latter Eq. (13) has been converted into PSD space to make the two estimates comparable. In addition, the straight model simulation of Fig. 2b (dashed black curve), without data assimilation, has been plotted as well. The prediction skill of these three types of forecast is compared in turn to the total variance of the PSD observations obtained from the control run of Fig. 2a; the latter should be reduced by the forecast model's interpolating the sparse data, even though this model is not perfect.

In practice, we see that for our fraternal-twin experiments, the model simulation with the wrong parameter value of $\zeta = 1$ day does not yield any useful prediction skill, as the variance of its innovation sequence is even higher than the variance of the "observations," i.e. of the control run. On the other hand, both EKF formulations reduce the variance of the innovation sequence, while the log-normal formulation has a substantially better prediction skill than the standard EKF at all L-shells, cf. Fig. 4a. In addition, the analysis obtained by the log-normal formulation has a substantially lower time-averaged error with a truth (cf. Fig. 3c,d) at all L-shells, cf. Fig. 4b.

The results in Figs. 2–4 have been obtained with observational errors set much larger than the model error: $\alpha_{\mathrm{m}} = 25$ and $\alpha_{\mathrm{o}} = 100\alpha_{\mathrm{m}}$. In this situation, the EKF can more easily correct the model forecast's state-vector components at grid points away from the observation sites [*Ghil and Malanotte-Rizzoli*, 1991]: When the model is assumed to be more accurate than the observations, then the EKF's weights in Eq. (9) for such locations are non-negligible, due to the spatial correlations inferred from the error covariance matrix. The log-normal formulation, due to its capability to capture better very large variations in PSD values, allows for much larger corrections of the model forecast where the gradients are steepest.

When the observational and model errors are comparable, say $\alpha_{\mathrm{m}} = \alpha_{\mathrm{o}}$, both formulations yield very similar data assimilation results, with but small differences in prediction skill: the smallness of the differences apparent in Fig. 5 is largely due to the fact that the model forecast is modified to a much lesser extent at grid points away from the observation sites. Finally, when the model error is much larger than the observational error, the EKF approximates the "direct-insertion method," in which the observations simply replace the model forecast at all the points where observations are taken. In this case (not shown), the EKF results are the same, regardless of the formulation chosen. At the opposite end of the error ratio scale, when the model errors are negligible, the EKF will ignore the observations completely.

These results suggest that there is a certain range of ratios between observational and model errors within which the log-normal EKF formulation will perform better than the standard one. In particular, based on our fraternal-twin experiments with synthetic data, the log-normal formulation of the EKF is expected to perform better when the observational errors are larger than the model errors. In the next section, we verify these results by assimilating actual satellite data, whose errors are quite large.

6.2. Spacecraft data assimilation

In this section, we compare the standard and the log-normal EKF formulations by assimilating PSD data derived from measurements on-board the Akebono and GEO spacecraft.

These are assimilated into the VERB-1D code with the loss parameters $\tau_{LI} = 10$ days and $\zeta = 5$ days; see Eq. (3) and Fig. 2a. Unlike in the fraternal-twin experiment of the preceding section, here we do not know the continuous spatio-temporal evolution of the true PSD field. Instead, we will consider as a comparison benchmark independent, high-quality observations from the CRRES spacecraft, with more complete coverage in time and space, as shown in Fig. 6a.

Based on intercalibration of PSD data we assume $\alpha_{\rm m} = 25$ for the model error in the VERB-1D code, while we take observational error $\alpha_{\rm o} = 200$ for Akebono and $\alpha_{\rm o} = 400$ for GEO. Such a choice of error parameters allows the Kalman filter to modify efficiently the full state vector, as described in Sec. 6.1. In agreement with our synthetic-data results there, the assimilation results (not shown) using the PSD data from the CRRES mission (shown in Fig. 6a) and from the GPS satellite (not shown) do not depend on the EKF formulation, since these two data sets are of higher quality and have smaller observational errors than the GEO and Akebono data.

First, we assimilate the AKEBONO RDM measurements which do not include the near-Earth region of steep PSD gradients, $L \leq 3$; the RDM observations are plotted in Fig. 6b. The assimilation results for the standard EKF formulation (Fig. 6c) have several non-physical PSD maxima at $L \approx 3$; these maxima are absent from the CRRES observations in Fig. 6a. The results for the log-normal EKF formulation in Fig. 6d, on the other hand, yield a smooth PSD field in much better agreement with the CRRES data of Fig. 6a.

The prediction skill is shown in Fig. 7 and it is improved by both EKF filter formulations in comparison with the model simulation without the benefit of data assimilation.

However, the skill for the log-normal EKF is only modestly better at all L-values, i.e. the innovation variance is somewhat smaller than for the standard EKF.

Unlike in the Akebono case, the GEO measurements cover only a very narrow L-range, at $L \approx 5$; see Fig. 8b. Such a limited data set presents a greater challenge for the EKF in realistically reconstructing the PSD profile at low L-shells, far away from the observations points, cf. 7c. Even in this case, the prediction skill of the log-normal EKF is uniformly better over the L-range sampled by GEO, as can be seen in Fig. 9.

7. Conclusions

This study was motivated by the recognition that both simulated and observed phase-space density (PSD) values in the radiation belts are subject to very large spatio-temporal variations, and that variations over several orders of magnitude may not be adequate for standard data assimilation methods based on least-squares minimization of normally distributed errors. We formulated therefore in Sec. 5 a model and filter version using the logarithm of the PSD as the dependent variable.

Our "fraternal-twin" experiments in Sec. 6.1 showed that the proposed log-normal formulation of the extended Kalman filter (EKF) can substantially reduce the assimilation errors in regions of steep PSD gradients; see Figs. 2 and 3. The proposed methodology demonstrates the most substantial improvements when model errors are smaller than the observational errors. Such an error ratio allows the log-normal EKF implementation to modify the model forecast very efficiently in observation-void regions; these modifications lead to much better PSD predictions, as inferred from the variance reduction of the innovation sequence in Figs. 4 and 5.

These findings have been confirmed by assimilating PSD measurements from the GEO and Akebono satellites (Figs. 6–9), which have large observational errors derived from intercalibration studies. In particular, the log-normal EKF applied to Akebono observations yields an assimilated PSD field in which non-physical maxima are absent, according to independent CRRES validation data. In addition, the prediction skill of the log-normal formulation is better for both the GEO and Akebono data. The results of this study should thus be useful to researchers, as well as to spacecraft designers and engineers, in the transition to operational prediction of the near-Earth space environment of satellites and other high-technology systems.

Our proposed rescaling methodology holds even greater promise for the data assimilation of multiple-satellite measurements for sophisticated, three-dimensional radiation-belt models. Such models describe much better competing loss and source mechanisms than the 1-D VERB code used in this study, thus reducing further model errors.

Acknowledgments. This work is supported by the UCLA-LANL Radiation Belt Reanalysis Project, 090LR-04-116720-SHPY, and NSF AGS-1102009 grant. We also thank two anonymous reviewers for their constructive comments.

References

Baker, D. N., S. G. Kanekal, X. Li, S. P. Monk, J. Goldstein and J. L. Burch (2004) An extreme distortion of the Van Allen belt arising from the Halloween solar storm in 2003, *Nature*, 878–880.

Bengtsson, L., M. Ghil, and E. Källén, Eds. (1981) *Dynamic Meteorology: Data Assimilation Methods*, Springer-Verlag, New York/Heidelberg/Berlin, 330 pp.

Bjerknes, V. (1904) Das Problem der Wettervorhersage, betrachtet vom Standpunkt der Mechanik und der Physik. *Meteor. Z.*, 21, 1–7.

Brautigam, D. H., and J. M. Albert (2000) Radial diffusion analysis of outer radiation belt electrons during the October 9, 1990, magnetic storm, *J. Geophys. Res.*, 105, 291.

Carrassi, A., M. Ghil, A. Trevisan and F. Uboldi [2008] Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system, *Chaos*, 18(2), 023112, doi: 10.1063/1.2909862.

Charney, J. G., R. Fjørtoft and J. von Neumann (1950) Numerical integration of the barotropic vorticity equation. *Tellus*, 2, 237–254.

Charney, J. G., M. Halem and R. Jastrow (1969) Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.*, 26, 1160–1163.

Courtier, P., and O. Talagrand (1987) Variational assimilation of meteorological observations with the adjoint vorticity equation. Part II: numerical results. *Q. J. R. Meteorol. Soc.*, 113, 1329–1347.

Crow, E. L. and Shimizu, K., Eds (1988) *Lognormal Distributions: Theory and Applications*. M. Dekker, New York, 387 pp.

Daae, M., Y. Shprits, B. Ni, J. Koller, D. Kondrashov, and Y. Chen (2011) Reanalysis of Radiation Belt Electron Phase Space Density using the UCLA 1-D VERB code and Kalman filtering: Sensitivity to assumed boundary conditions and changes in the loss model, *Adv. Space Res.*, 48 1327–1334.

Dee, D. P., S. E. Cohn, A. Dalcher, and M. Ghil (1985) An efficient algorithm for estimating covariances in distributed systems, *IEEE Trans. Autom. Control*, AC-30, 1057–1065.

Fukumori, I. (2006) What is data assimilation really solving, and how is the calculation actually done?, in "Ocean Weather Forecasting: An Integrated View of Oceanography", E. P. Chasignet and J. Verron, Eds., Springer, pp. 317–342.

Gelb, A. (Ed., 1974) *Applied Optimal Estimation*, M.I.T. Press, Cambridge, MA, 374 pp.

Ghil, M. (1989) Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework, *Dyn. Atmos. Oceans*, 13, 171–218.

Ghil, M., M. Halem, and R. Atlas (1979) Time-continuous assimilation of remote-sounding data and its effect on weather forecasting, *Mon. Wea. Rev.*, 107, 140–171.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson (1981) Applications of estimation theory to numerical weather prediction, in "Dynamic Meteorology: Data Assimilation Methods," L. Bengtsson, M. Ghil and E. Källén (Eds.), Springer-Verlag, pp. 139–224.

Ghil, M. and P. Malanotte-Rizzoli (1991) Data assimilation in meteorology and oceanography, *Adv. Geophys.*, 33, 141–266.

Harten, A. (1983) High-resolution schemes for hyperbolic conservation laws, *J. Comput. Phys.*, 49, 357–393, doi:10.1006/jcph.1997.5713.

Jazwinski, A. H. (1970) *Stochastic Processes and Filtering Theory*, Academic Press, New York, 376 pp.

Kailath, T. (1968) An innovations approach to to least-squares estimation—Part I: Linear filtering in additive white noise, *IEEE Trans. Autom. Control*, 15(6), 658–660.

Koller, J., Y. Chen, G. D. Reeves, R. Friedel, T. E. Cayton, and J. A. Vrugt (2007) Identifying the radiation belt source region by data assimilation, *J. Geophys. Res.*, 112, A06, 244.

Kondrashov, D., Y. Shprits, and M. Ghil (2010) Gap filling of solar wind data by singular spectrum analysis, *Geophys. Res. Lett.*, **37**, L15101, doi:10.1029/2010GL044138.

Kondrashov, D., Y. Shprits, R. Thorne and M. Ghil (2007) A Kalman filter technique to estimate relativistic electron lifetimes in the outer radiation belt, *J. Geophys. Res.*, 112, A10227, doi:10.1029/2007JA012583.

Kondrashov, D., C. Sun, and M. Ghil (2008) Data Assimilation for a Coupled Ocean-Atmosphere Model. Part II: Parameter Estimation, *Mon. Wea. Rev.*, 136, 5062–5076, doi: 10.1175/2008MWR2544.1

Lanzerotti, L.J., and C.G. Morgan (1973) ULF geomagnetic power near L=4, 2. In: Temporal variation of the radial diffusion coefficient for relativistic electrons, *J. Geophys. Res.*, 78, 4600–4610.

Lanzerotti, L.J., D.C. Webb, and C.W. Arthur (1978) Geomagnetic field fluctuations at synchronous orbit, 2. Radial Diffusion *J. Geophys. Res.*, 83, 3866–3870.

Miller, R. N., M. Ghil and F. Gauthiez (1994) Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.*, 51, 1037–1056.

Naehr, S. M., and F. R. Toffoletto (2005) Radiation belt data assimilation with an extended Kalman filter, *Space Weather*, 3, S06001, doi:10.1029/2004SW000121.

Ni, B., Y. Shprits, T. Nagai, R. Thorne, Y. Chen, D. Kondrashov, and H. j. Kim (2009) Reanalyses of the radiation belt electron phase space density using nearly equatorial CRRES and polar-orbiting Akebono satellite observations, *J. Geophys. Res.*, **114**, A05208, doi:10.1029/2008JA013933.

O'Brien, P., and T. B. Guild (2010) A Proposed Data Assimilation Scheme for a Nonlinear Model With non-Gaussian Errors and Nonlinear Measurements, *Aerospace Report*, ATR-2010(5529)-2.

Y. Shprits, D. Kondrashov, Y. Chen, R. Thorne, M. Ghil, R. Friedel, and G. Reeves (2007) Reanalysis of Relativistic Radiation Belt Electron Fluxes using CRRES Satellite Data, a Radial Diffusion Model, and a Kalman Filter, *J. Geophys. Res.*, **112**, A12216, doi:10.1029/2007JA012579.

Shprits, Y. Y. , R. M. Thorne, G. D. Reeves, R. Friedel (2005) Radial diffusion modeling with empirical lifetimes: Comparison with CRRES observations, *Annales Geophys.*, 23, 4, 1467–1471.

Shultz, M., and L. J. Lanzerotti (1974) *Particle Diffusion in the Radiation Belt*, Springer-Verlag, New York.

Subbotin, D. A., and Y. Y. Shprits (2009) Three-dimensional modeling of the radiation belts using the Versatile Electron Radiation Belt (VERB) code, Space Weather, 7, S10001, doi:10.1029/2008SW000452.

Tsyganenko, N.A. (1989) A magnetospheric magnetic field model with a warped tail current sheet, *Planet. Space Sci.,* 37, 5–20.

Walt, M. (2005) *Introduction to Geomagnetically Trapped Radiation.* Cambridge Univ. Press, Cambridge/New York, 192 pp.; available online (2009), `http://dx.doi.org/10.1017/CBO9780511524981.`

Wiener, N. (1949) *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*, M.I.T. Press, Cambridge, MA, 163 pp.

Zupanski, M., (2005) Maximum likelihood ensemble filter: Theoretical aspects, *Mon. Weather Rev.*, 133, 1710.

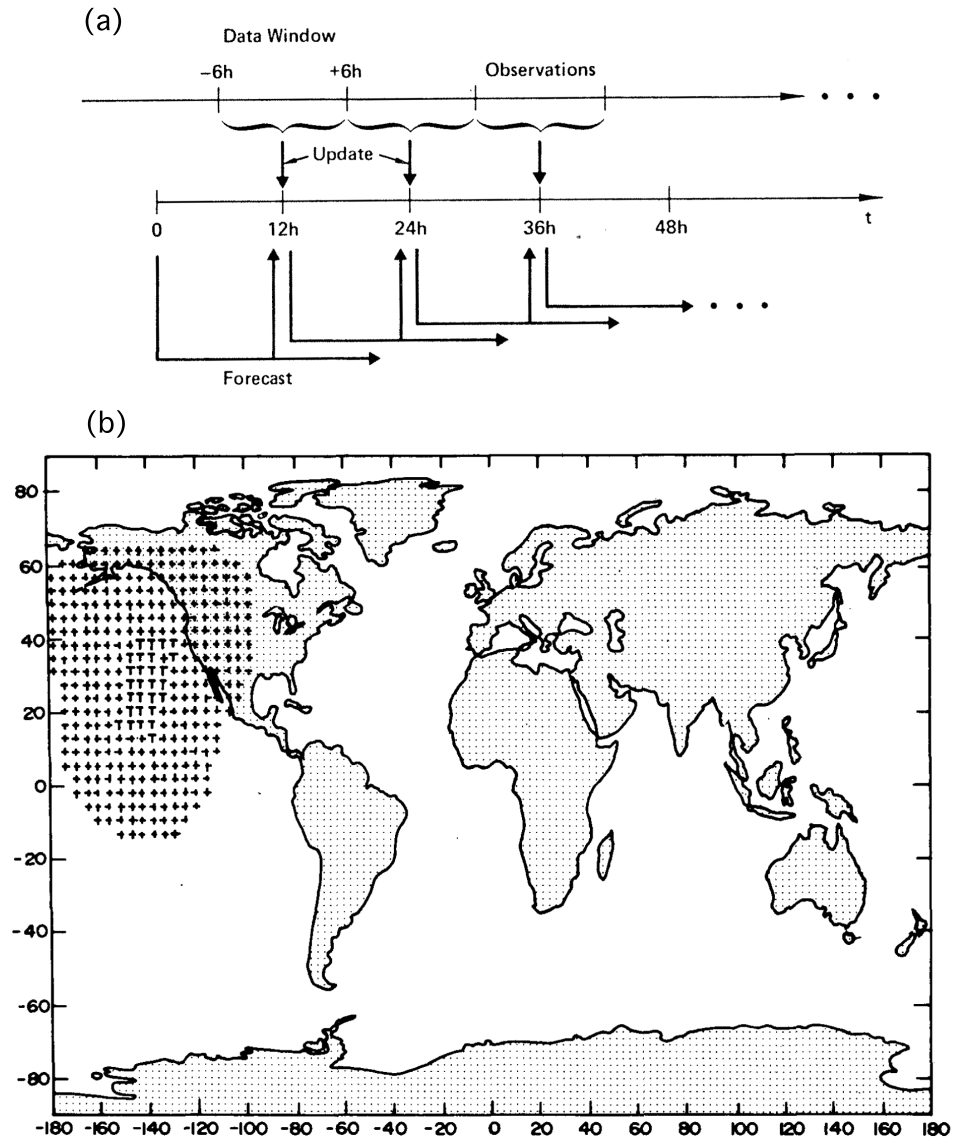

Figure 1. Operational forecast-and-assimilation cycle of a typical weather service that combines the prediction and data assimilation processes. (a) Data are gathered from a "window" of near-past and near-future data, at the synoptic times, 12 h apart; (b) data are assimilated as they become available, at any model time step. In panel (b), the letters 'T' stand for the locations at which temperature profiles become available from an infrared satellite sounder at a particular model time step, while the '+' signs indicate grid points that will be affected by those soundings due to the particular sequential filter applied to the soundings. Adapted from (a) *Ghil* [1989], and (b) *Ghil et al.* [1979], respectively.

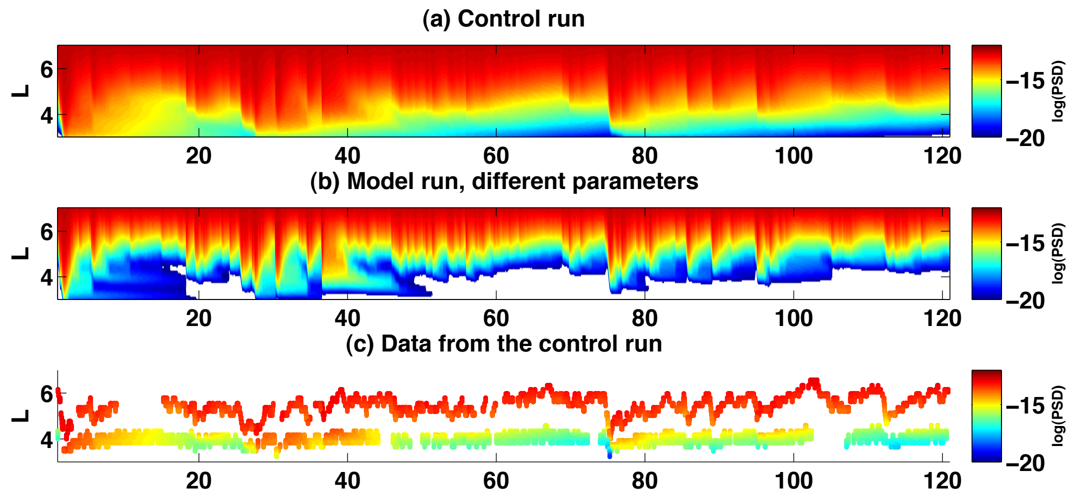

Figure 2. "Fraternal-twin" experiment using synthetic observations from a model simulation with different parameter values. The radiation belt model employs a Kp-dependent lifetime parameterization outside the plasmasphere, with $\tau_{LO} = \zeta/Kp(t)$, cf. Eq. (3). (a) "Truth" given by the model solution with $\zeta = 5$ days, also called the *control run* or *nature run*; (b) model simulation assuming higher losses, with $\zeta = 1$ day; (c) synthetic observations taken from the control run in panel (a).

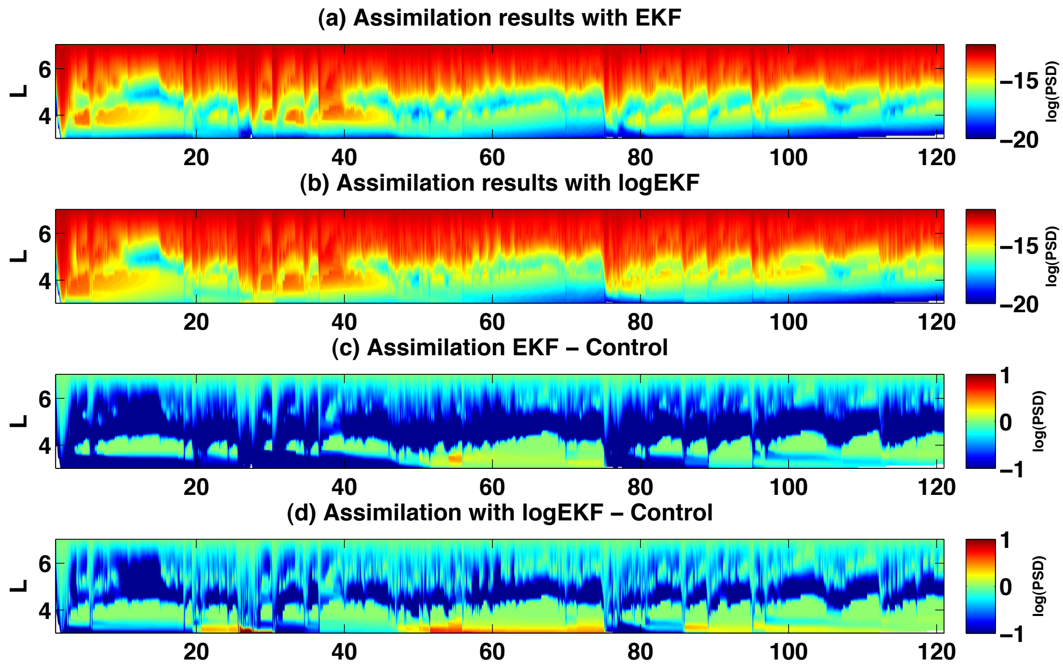

Figure 3. Assimilation results for "fraternal-twin" experiment with forecasts from the model in Fig. 2b and data from the control run in Fig. 2c: (a) using the standard EKF formulation of Sec. 4; (b) same as in panel (a) but for the log-normal EKF of Sec. 5; (c) difference between the assimilation results in panel (a) and the control run of Fig. 2a; (d) difference between the assimilation results in panel (b) and the same control run.

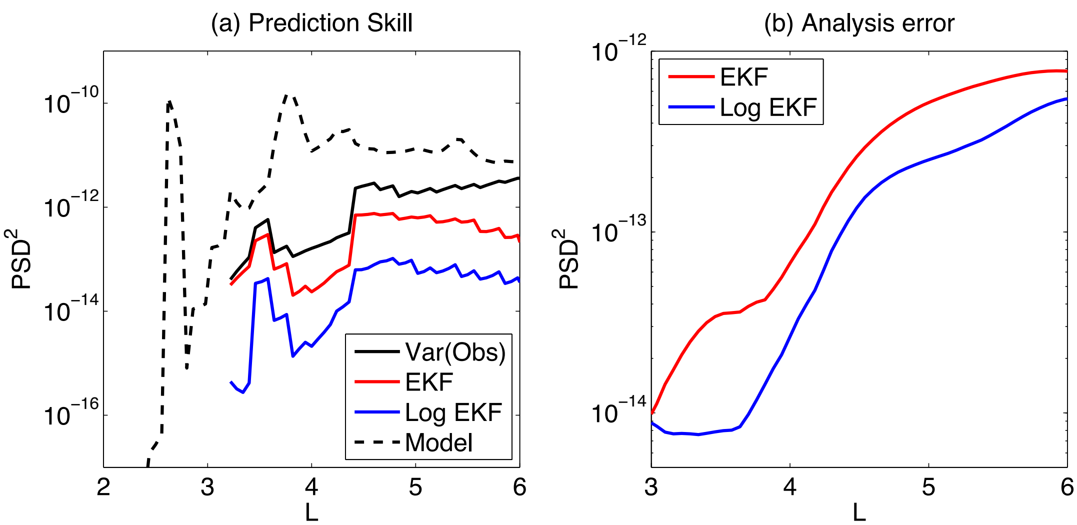

Figure 4. (a) Prediction skill of the models and the sequential estimation methods, defined as variance $E[\mathbf{z}_k^{\mathrm{T}}\mathbf{z}_k]$ of the innovation sequence with $\mathbf{z}_k$ given by Eq. (13); see text for details. Black solid curve: variance of the synthetic PSD observations sampled from the control run, cf. Fig. 2c; black dashed: model simulation with incorrect parameter values and no data, cf. Fig. 2b; red: standard EKF; blue: same for the log-normal EKF but converted into PSD values. Data assimilation clearly improves the models' forecasting ability of the data for all L-values, with the smallest variance of the innovation sequence for the log-normal formulation. These results are for observation errors much larger than the model errors; see text for details. (b) Error computed as time mean of the squared difference between assimilation results and control, given in both cases in terms of PSD values: for the standard EKF (red curve, cf. Fig. 3c), and log-normal EKF (blue curve, cf. Fig. 3d.)

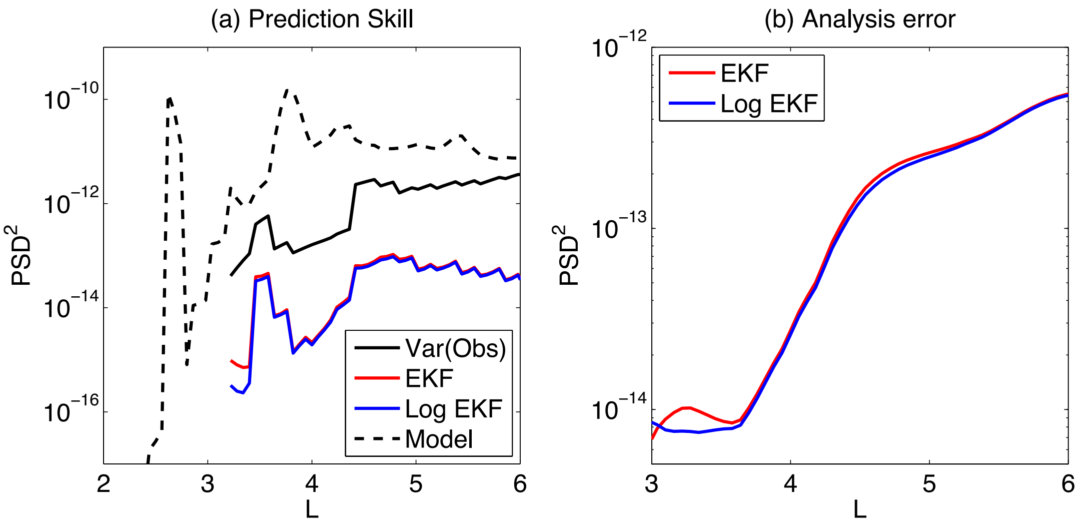

Figure 5. Same as in Fig. 4 but assuming equal model and observational errors; see text for details. In this case, the performance of the standard EKF and the log-normal EKF are quite comparable.

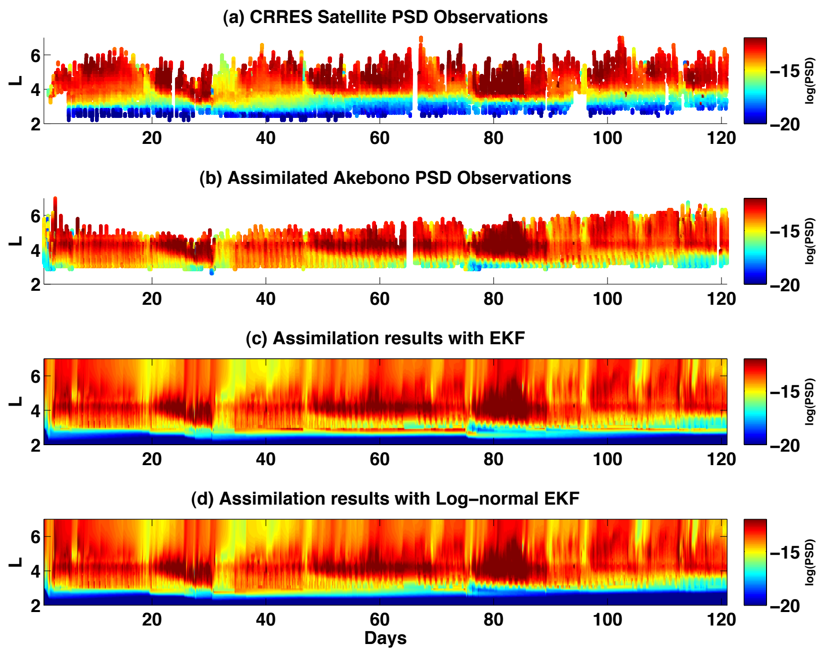

Figure 6. Data assimilation results using real spacecraft data. (a) CRRES PSD observations; (b) assimilated Akebono observations; (c) assimilation results with the standard EKF; (d) assimilation results with the log-normal EKF. The log-normal formulation provides better agreement of the assimilation results with the CRRES observations in the inner belt, $L < 3$, where the PSD gradients are strong.

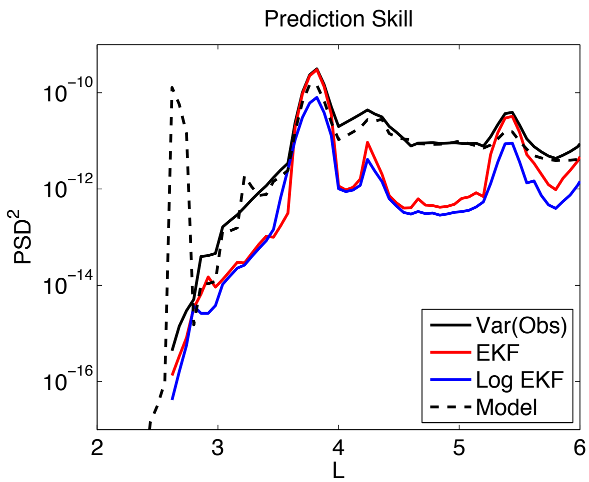

Figure 7. Prediction skill for real spacecraft data. Same color conventions as in Figs. 4a and 5a — solid black: variance of the Akebono PSD observations shown in Fig. 6b; dashed black: model simulation with no data (see again Fig. 2a); red and blue: standard and log-normal EKF, respectively. The Akebono data improve the model's forecasting ability over all L-values, and the log-normal formulation exhibits the smallest variance of the innovation sequence.

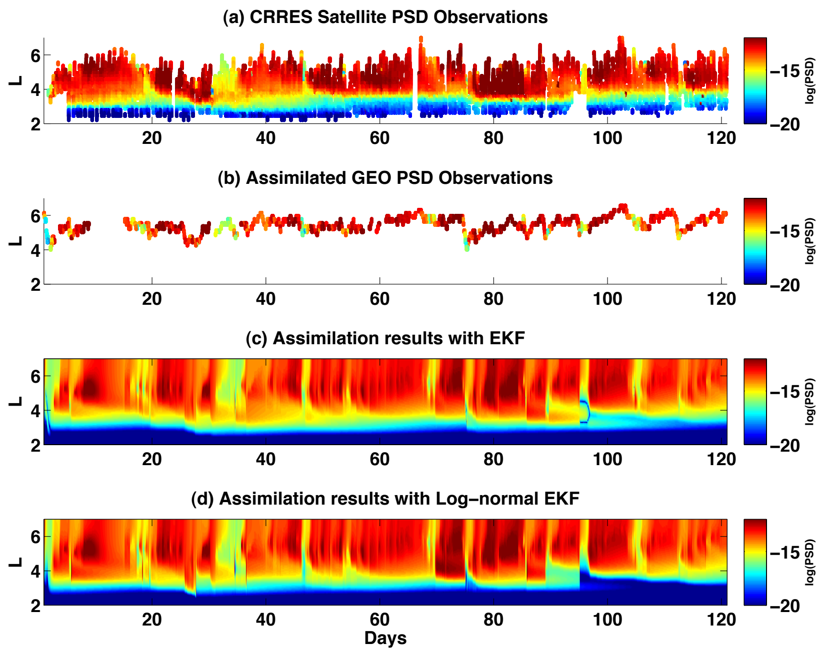

Figure 8. Same as in Fig. 6 but for assimilating observations from the GEO satellite.

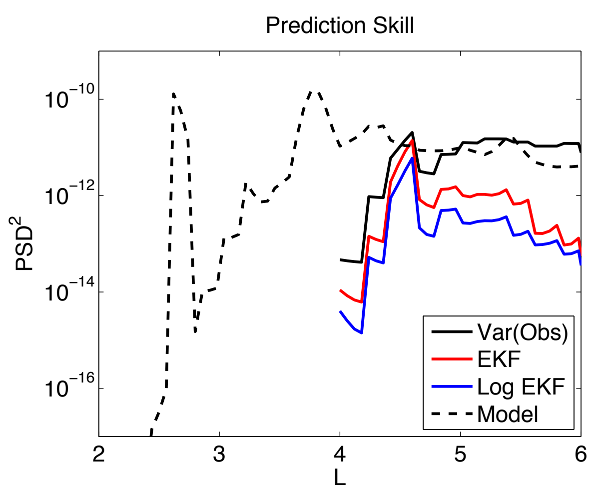

Figure 9. Same as in Fig. 7 but for assimilating observations from the GEO satellite.

The log-normal formulation provides again a smaller variance of the innovation sequence than the standard algorithm, and hence better prediction; compare the blue and red curves, respectively.