GEOPHYSICAL RESEARCH LETTERS, VOL. ???, XXXX, DOI:10.1029/,

Forecasting Change of the Magnetic Field using Core Surface Flows and Ensemble Kalman Filtering

C. D. Beggan*

School of GeoSciences, University of Edinburgh, Edinburgh, UK

K. A. Whaler

School of GeoSciences, University of Edinburgh, Edinburgh, UK

C. D. Beggan, School of GeoSciences, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh, UK, EH9 3JW. *now at the British Geological Survey (ciar@bgs.ac.uk)

K. A. Whaler, School of GeoSciences, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh, UK, EH9 3JW. (kathy.whaler@ed.ac.uk)

Accurate forecasting of the change of the Earth's internal magnetic field over short intervals of time (e.g. less than five years) has many applications for government, academic and commercial users. Forecasting can be achieved by making a number of reasonable assumptions about how the main field interacts with the flow in the liquid outer core. In particular, the magnetic field can be considered to be entrained in the large scale flow along the core-mantle boundary surface over short time periods, giving rise to measurable change of the field at the Earth's surface. The observed change (or secular variation) at or above the surface of the Earth can thus be inverted to produce flow models; these can be used to propagate fluid parcels threaded by the field forwards in time to forecast the non-linear change of the magnetic field. In addition to prediction of field change by flow models, it would be advantageous to include observations of the field from satellite measurements or ground-based observatories. We therefore present a method using Ensemble Kalman Filtering (EnKF) to produce an optimal assimilation between magnetic field change as forecast from core flow models and direct observations of the field. We show, by assuming a steady flow and assimilating field observations annually, it is possible to produce a forecast over five years with less than 30nT root mean square difference from the 'true' field – within an assumed error budget. The EnKF method also allows sensitivity analysis of the field models to noise and uncertainty within the physical representation.

1. Introduction

The slow temporal variation of the Earth's magnetic field is termed 'secular variation' (SV) and is related to advection and diffusion of the field within the liquid outer core. Forecasting the short term change of the field in an accurate and timely fashion is of great benefit to commercial users in areas such as mining, underground drilling and navigation, as well as for academic and civilian users, e.g. where access to real-time data may not be available. The International Geomagnetic Reference Field (IGRF) model enjoys widespread use for this purpose. The model is revised and updated every five years and forecasts secular variation for the future five year period [*Macmillan and Maus*, 2005]. Methods for forecasting the magnetic field change have previously relied upon extrapolation of ground-based observatory data and the forecasts can often be quite in error at the end of their desired lifetime.

Recently, high resolution magnetic field models such as GRIMM [*Lesur et al.*, 2008], POMME [*Maus et al.*, 2006] and xCHAOS [*Olsen and Mandea*, 2008] have been developed using data from the CHAMP, Ørsted and SAC-C satellite missions. These provide an excellent description of the field, SV and secular acceleration (SA) over the period 1999–2009. Detailed models of the large-scale surface core flows generating the observed SV have been developed by a number of researchers [e.g. *Hulot et al.*, 2002; *Holme and Olsen*, 2006]. If it is assumed on short time scales that advection by core flow of the magnetic field dominates diffusion then, in a manner analogous to weather forecasting, the evolution of the field can be forecast by propagation of the flow forwards in time.

Using this approach, *Maus et al.* [2008] generated SV from a series of flow models with differing physical constraints to investigate how well the field could be hindcast compared to the CM4 magnetic field model [*Sabaka et al.*, 2004]. They found the misfit between the hindcast field from core flow models and the CM4 model to be less than 100nT root mean square (RMS) difference after five years and up to 300nT after ten years. We improve upon this result by using a different flow model inversion technique and employing an Ensemble Kalman Filter.

The Ensemble Kalman Filter (EnKF) is a Monte-Carlo method for optimally combining models of and observational information about a physical process by statistical representation of the associated uncertainties [*Evensen*, 1994]. It is extensively used in weather and ocean dynamics forecasting to improve the accuracy of forecasts and to explore the sensitivity of systems to minor perturbations [*Evensen et al.*, 2007]. Data assimilation in geomagnetism is still in its infancy but has recently been investigated [*Fournier et al.*, 2007; *Kuang et al.*, 2008]. In this paper, we adapt the EnKF for magnetic field prediction using a simple steady flow model and assuming a relatively noisy field model from limited satellite coverage and ground-based magnetic observations is available. This scenario might occur at some point in the future where continuous satellite monitoring has ceased.

2. Methods

In the following we describe the methods used to derive a steady flow model that is used for forecasting, the implementation of an EnKF model and the resulting improvements of the field forecast using EnKF with assimilation compared to the forecast. We choose a steady flow as experiments by *Maus et al.* [2008] found that hindcasts from a steady flow model produced the best average long term fit to the CM4 field model and because it is the simplest assumption to make for a flow model. More complex flows (e.g. time-varying or different physical hypotheses) can be used if neccesary.

2.1. Flow Modelling and Forecasting

Magnetic main field models are typically represented as a vector of spherical harmonic Gauss coefficients ($\mathbf{g} = [g_l^m; h_l^m]$). Secular variation of the field can be inverted for toroidal and poloidal flow using the linear relationship between SV and flow spherical harmonic coefficients. The relation is through the Gaunt/Elsasser matrix ($\mathbf{H}$) whose elements depend on the main field coefficients [*Whaler*, 1986] which change with time. In this study, the main field, SV and flow coefficients are truncated at degree and order $l_{max} = 14$, thus we have assumed that only large scale flows are responsible for the large scale SV. Note that we invert SV data directly (as explained below) rather than using spherical harmonic models ($\dot{\mathbf{g}}$) of SV.

With knowledge of the data covariance, we seek the flow ($\hat{\mathbf{m}}$) which can be obtained from the SV using the standard L_2 least-squares minimisation norm. We then apply an additional step using an iterative L_1 norm minimisation technique as described in *Beggan et al.* [2009]. The L_1 norm technique improves the fit of the flow to the SV data by iterative reweighting of the residual differences. The flow is regularized by imposition of the so-called 'strong' norm *a priori* conditions [*Bloxham*, 1988], with a damping parameter controlling fit to the data versus flow smoothness.

In our first experiment, a series of 25 monthly SV data sets, over the period 2001.9–2004.0, were generated from CHAMP satellite data using the 'Virtual Observatory' method of *Mandea and Olsen* [2006]. The SV data were inverted for a steady flow model [*Voorhies and Backus*, 1985], with a tangentially geostrophic flow constraint. This produces a set of flow coefficents ($\hat{\mathbf{m}}_{SF}$) representing an 'average' flow over the period. The steady flow model was used to forecast the change of the magnetic field over the five year period from 2004.0 to 2009.0 and compared to the GRIMM, POMME and xCHAOS satellite field models.

The Gauss coefficients from the xCHAOS model for 2004.0 were used as the starting field model. The field was advected forward over successive months (k) for five years using the equation:

$$\mathbf{g}_{k+1} = \mathbf{g}_k + (\mathbf{H}_k \hat{\mathbf{m}}_{SF})/12 \tag{1}$$

with the $\mathbf{H}_k$ matrix updated at every timestep using the main field coefficients forecast from the previous timestep, making the system non-linear. To evaluate the validity of this forecast, the RMS difference (or misfit) metric ($\sqrt{dP}$) to a satellite field model is calculated by:

$$dP = \sum_{l=1}^{l_{max}} \sum_{m=0}^{l} (l+1)[(\mathbf{g}_l^m)_{field} - (\mathbf{g}_l^m)_{forecast}]^2 \tag{2}$$

Figure 1 shows the misfit of the forecast from the flow model to the GRIMM, POMME and xCHAOS satellite field models. Note the GRIMM model spline coefficients extend to 2006.5, while the POMME model is extrapolated beyond 2007.5 using constant SV.

We now show how to improve upon these results by employing an Ensemble Kalman Filter to assimilate field observations into forecasts from core flow models.

2.2. Data Assimilation in Ensemble Kalman Filtering

In an EnKF, the state of a dynamic process at any particular time can be represented as a vector in n-dimensional space, where n is the number of parameters in the system. The uncertainty of the process is represented by perturbing the inputs randomly by a known variance (with zero-mean) to produce an 'ensemble' of states – conceptually imagined as a 'cloud' of points in n-dimensional space. The evolution of the states though time is controlled by propagating the ensemble forward using model equations of the system behavior. When an observation is available, it can be optimally assimilated into the ensemble by applying the standard Kalman Filter equations [*Kalman*, 1960]. With a sufficiently large ensemble (determined through experimentation), the mean state represents the most likely value for the process at the time. The evolution of the ensemble can be explored by examining the 'spread' of the states about the mean.

A traditional Kalman Filter is implemented in two steps: (1) prediction of the evolution of the model state by dynamic equations believed to adequately represent the system and (2) assimilation of a measurement to correct any accumulated error from the model. At time k, the optimal blending of a forecast state ($\mathbf{x}_k^f$) and measurement ($\mathbf{z}_k$) to generate the assimilated state vector, $\mathbf{x}_k^a$, is through the so-called Kalman gain matrix ($\mathbf{K}_k$):

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{x}_k^f) \tag{3}$$

with

$$\mathbf{K}_k = \mathbf{P}_k^f(\mathbf{P}_k^f + \mathbf{Q})^{-1}. \tag{4}$$

where $\mathbf{P}_k^f$ is the covariance of the model and $\mathbf{Q}$ is the covariance of the data measurement. The balance between the error of the model and measurement controls the assimilation step. When the Kalman gain matrix has been calculated, the covariance of the assimilated state vector is calculated as:

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k)\mathbf{P}_k^f. \tag{5}$$

In the EnKF, $\mathbf{x}_k^f$ is a model forecast with noise $\mathbf{w}_k^f$, and $\mathbf{z}_k$ is a measurement with some associated measurement noise $\mathbf{u}_k$. The forecast, measurement and the newly assimilated estimate, $\mathbf{x}_k^a$, are related to the true state of the system, $\mathbf{x}_k^t$, by:

$$\mathbf{x}_k^f = \mathbf{x}_k^t + \mathbf{w}_k^f; \qquad \mathbf{x}_k^a = \mathbf{x}_k^t + \mathbf{w}_k^a; \qquad \mathbf{z}_k = \mathbf{x}_k^t + \mathbf{u}_k \tag{6}$$

with expectations (i.e. the mean of) $\overline{\mathbf{w}}_k^f = \overline{\mathbf{w}}_k^a = \overline{\mathbf{u}}_k = 0$, given a large enough ensemble. If we consider the covariance of an assimilated ensemble, it can be shown [*Evensen*, 1994]:

$$\begin{aligned}
\mathbf{P}^a &= \overline{(\mathbf{w}^a)^2} = \overline{(\mathbf{x}^a - \mathbf{x}^t)^2} \\
&= \left(\mathbf{I} - \frac{\mathbf{P}^f}{\mathbf{P}^f + \mathbf{Q}}\right)\mathbf{P}^f \\
&\quad +2\frac{\mathbf{P}^f}{\mathbf{P}^f + \mathbf{Q}}\left(\mathbf{I} - \frac{\mathbf{P}^f}{\mathbf{P}^f + \mathbf{Q}}\right)\overline{\mathbf{w}^f\mathbf{u}}.
\end{aligned} \tag{7}$$

This leads to the key result of the EnKF: when the expectation $\overline{\mathbf{w}^f\mathbf{u}} = 0$, Equation 7 is equivalent to Equation 5. This occurs when a suitably large number of ensemble states are employed.

2.3. Practical Implementation

There are three stages required to implement the EnKF for this problem: (1) generation of the initial ensemble, (2) forecasting the change of the field by driving the field model with SV predicted by core flow models and (3) assimilation of measurements e.g. from a 'true' field model. Each of these stages is explained in detail below.

Initiating the Ensemble

The ensemble is initiated by generating a perturbed set of Gauss coefficients. The mean value of the initial ensemble is equal to the input coefficients of the field. This is implemented as follows:

1. An initial state vector, at time $k = 1$, is set to be a vector of Gauss coefficients from a field model (e.g. xCHAOS).

2. If a time series of flow models are available, rather than a single steady flow, the variability of the the flow model coefficients can be used as additional information. To generate the perturbation to the $\mathbf{g}_l^m$ field coefficients, the standard deviation for each coefficient over the entire set of *flow* models is calculated (from the variability in each flow coefficient of $\hat{\mathbf{m}}$). However, with a single steady flow an alternative estimate of the variance must be made.

3. A matrix of normally distributed random numbers $N(0, 1)$ with size $[l_{max}(l_{max}+2) \times n_{ensembles}]$ is created, where $n_{ensembles}$ is the number of ensemble states.

4. The matrix of random numbers is multiplied by the standard deviation of the flow coefficients to give a perturbed flow coefficient matrix.

5. The perturbed flow coefficient matrix is pre-multiplied by the $\mathbf{H}$ matrix to produce a matrix of perturbed SV coefficients, correctly scaled to reflect the uncertainty in the flow models.

6. The perturbed SV coefficient matrix is then added to the initial state vector to produce an ensemble matrix (Ensemble$_1$).

Once the initial ensemble has been created, forecasting and assimilation can take place.

Driving the Ensemble Forecasts

The forecast (prediction) of the field is driven forwards by the summation of (1) the field coefficients and (2) the monthly SV coefficients from the flow model which are perturbed by a random matrix with zero mean and standard deviation computed from the variance of the flow over time. In addition, at each timestep, model noise is added to simulate the variance of the ensemble, forcing it to grow at each forecast iteration. The model noise is controlled by the size of the time-step (Δt), the standard deviation of the SV coefficients from the previous iteration, and a parameter ρ, which can be used to control the time correlation of the noise, if required [*Evensen et al.*, 2007].

1. The SV coefficients generated by the flow model for month k are calculated by multiplying the flow model coefficients by the $\mathbf{H}$ matrix.

2. The monthly SV coefficients are perturbed by the standard deviation of the flow converted into an equivalent SV.

3. Model noise is simulated by multiplication of a matrix of random zero-mean normally-distribution numbers (of size $[l_{max}(l_{max}+2) \times n_{ensembles}]$) with the square-root of the timestep $\sqrt{\Delta t}$ and ρ.

4. The matrix of perturbed SV coefficients and model noise are added to the ensemble from the previous timestep (Ensemble$_{k-1}$) to produce the forecast for the current ensemble (Ensemble$_k$).

These four steps are repeated until a measurement becomes available for assimilation into the ensemble.

Assimilation of Measurements into the Ensemble

Over time, the forecast field will begin to diverge from the actual field. To improve the forecast, data can be input into the ensemble to update (correct) it. The data have associated errors which are used to generate a perturbed data ensemble. These perturbed data are assimilated into the overall ensemble using the Kalman Filter algorithm.

1. Data, for example a set of Gauss coefficients ($\mathbf{z}_k$), are available with a certain (estimated or known) error for each coefficient.

2. A matrix of zero-mean Gaussian random numbers is generated and scaled with the data error.

3. The data are added to the matrix of scaled random numbers to produce a matrix of 'perturbed data', with mean equal to that of the data themselves.

4. Using Equation 3 the data perturbation matrix and the perturbed SV coefficients are optimally assimilated into the ensemble at this timestep.

The covariance matrices can be estimated from the ensemble and measurement errors [*Evensen*, 1994]. Note it is also possible to use non-synoptic (i.e. partial) measurements of the field in the assimilation step with an appropriate 'observation' operator. *Evensen et al.* [2007] outlines and demonstrates how to efficiently code and compute the matrix operations for the EnKF. The number of ensemble states was set to 1000 after experimentation, though it was found that any more than 500 is adequate. Typically, a measurement (i.e. Gauss coefficients from a field model) is assimilated every twelve months.

3. Applying the Ensemble Kalman Filter to Forecasting

In Figure 1, the steady flow model prediction slowly diverges from the main field models over the time period. Assimilating actual field measurements would be expected to improve the fit of the predicted field to the 'true' field. Any improvement is dependent on the errors of the input measurement. For example, a poor measurement allocated an associated small estimated error will increase the RMS misfit of the 'nowcast'. However, it is often difficult to correctly estimate the errors associated with each Gauss coefficient in a field model given that we do not have full knowledge of the field [*Langel et al.*, 1989].

The results of the forecast with data assimilation for the GRIMM and POMME (both extrapolated beyond 2006.5) and xCHAOS field models are shown in Figure 2. Each ensemble was initiated using the xCHAOS field model. Assimiliations of noisy measurments from the relevant field model are indicated by jumps in the curves. The solid black line represents the misfit (Equation 2) of the mean Gauss coefficients of the ensemble to the satellite field models, while the dashed lines are misfits of the Gauss coefficients one standard deviation above or below the mean. The middle and lower panels show that the mean ensemble (solid line) fits to better than 25nT for both the POMME and xCHAOS models over the entire period. Most of the misfit is from the difference between forecast and model at degrees $l = 1 - 4$.

From Equation 4 it should be clear that the calculation of the EnKF is sensitive to the estimates of input errors. Analysis of the factors affecting the forecast fidelity shows that the error associated with the assimilated Gauss coefficients is the major contributor. The error associated with the steady flow model coefficients is a secondary effect. In our example, after experimentation, the error on each of the field model coefficients was set to $\mathbf{z}/(2 \cdot 10^3)$. For the largest coefficient (g_1^0) this is a relative error of 15nT, equivalent to approximately two years of SV. Larger errors than this produce forecasts that are worse than predictions from steady flow alone. In this case, increasing the size of the error estimate of the measurement by two approximately doubles the size of the misfit. A ten-fold increase in the measurement error results in a poor input field estimate causing a large divergence from the 'true' field (the misfit after five years rises to over 400nT).

4. Discussion and Conclusion

The EnKF allows exploration of the system under consideration through examination of the 'spread' of the ensemble. In Figure 2, the ensemble models -1σ away from the mean are a poorer match to the 'true' model, though the $+1\sigma$ model is usually better than the mean for the GRIMM and POMME comparisons. Another note-worthy point is that certain measurement assimilations have little or no effect. For example, for POMME at 2008.0, the measurement assimilation barely alters the mean but does reduce the spread of the ensemble (the $\pm 1\sigma$ states become close to the mean).

With a steady flow model and annual data assimilations, the RMS difference between the forecast model and the 'true' field can be maintained at less than 30nT from 2004.0–2009.0 within assumed errors. This can result in a many-fold improvement e.g. compare the misfit of the forecast to xCHAOS in Figure 1 with the misfit of the mean forecast in Figure 2 (lower panel).

The use of the EnKF for this particular example is, perhaps, unneccesarily complicated. However, the method can be readily adapted for more complex flow regimes models and different data types.

In conclusion, we have demonstrated that forecasting of secular variation using a steady core flow model can achieve an acceptable match to the actual field. We have adapted the Ensemble Kalman Filter to improve forecasts and characterise their uncertainty by propagating a large number of possible field models forward in time using core flow models to control the evolution of the individual states. Optimal assimilation of measured data into the ensemble produces an improvement in the fit of the forecast to the actual field. Our approach thus offers a method to improve operational forecasting of the magnetic field.

Acknowledgments. We would like to acknowledge Susan Macmillan and Stefan Maus for their valuable advice and discussion on modelling and forecasting. We thank the two anonymous reviewers for their constructive suggestions for improving the manuscript. The CHAMP data used in this study were supplied by GFZ Potsdam. This research is part of the NERC GEOSPACE programme, funded under grant NER/O/S/2003/00674. CDB was funded under NERC studentship award NER/S/J/2005/13496.

References

Beggan, C., K. Whaler, and S. Macmillan (2009), Temporally varying residuals of core flow models from satellite-derived 'virtual observatories', *Geophys. J. Int.*, *177*, 463–475.

Bloxham, J. (1988), *The determination of fluid flow at the core surface from geomagnetic observations*, vol. Mathematical Geophysics, chap. 9, pp. 189–208, D. Reidel Publishing Company.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,14310,162.

Evensen, G., J. Hove, H. Meisingset, E. Reiso, K. Seim, and Ø. Espelid (2007), Using the EnKF for assisted history matching of a North Sea reservoir model, in *SPE 106184*.

Fournier, A., C. Eymin, and T. Alboussiere (2007), A case for variational geomagnetic data assimilation: insights from a one-dimensional, nonlinear, and sparsely observed MHD system, *Nonlin. Processes Geophys.*,, *14*, 163–180.

Holme, R., and N. Olsen (2006), Core surface flow modelling from high-resolution secular variation, *Geophys. J. Int.*, *166*, 518–528.

Hulot, G., C. Eymin, B. Langlais, M. Mandea, and N. Olsen (2002), Small-scale structure of the geodynamo inferred from Oersted and Magsat satellite data, *Nature*, *416*, 620–623.

Kalman, R. (1960), A new approach to linear filtering and prediction problems, *Transactions of the ASME Journal of Basic Engineering*, *82*, 35–45.

Kuang, W., A. Tangborn, W. Jiang, D. Lui, Z. Sun, J. Bloxham, and Z. Wei (2008), MoSST_DAS: The first generation geomagnetic data assimilation framework, *Communications in Computational Physics*, *3*, 85–108.

Langel, R., R. Estes, and T. Sabaka (1989), Uncertainty estimates in geomagnetic field modeling, *J. Geophys. Res.*, *94*, 12,281–12,299.

Lesur, V., I. Wardinski, M. Rother, and M. Mandea (2008), GRIMM: the GFZ Reference Internal Magnetic Model based on vector satellite and observatory data, *Geophys. J. Int.*, *173*, 382–394, doi:10.1111/j.1365-246X.2008.03724.x.

Macmillan, S., and S. Maus (2005), International Geomagnetic Reference Field: the tenth generation, *Earth Planets Space*, *57*, 1135–1140.

Mandea, M., and N. Olsen (2006), A new approach to directly determine the secular variation from magnetic satellite observations, *Geophys. Res. Lett.*, *33*(15), L15,306, doi:10.1029/2006GL026616.

Maus, S., M. Rother, C. Stolle, W. Mai, S. Choi, H. Lühr, D. Cooke, and C. Roth (2006), Third generation of the Potsdam Magnetic Model of the Earth (POMME), *Geochem. Geophys. Geosyst.*, *7*, Q07,008, doi:10.1029/2006GC001269.

Maus, S., L. Silva, and G. Hulot (2008), Can core-surface flow models be used to improve the forecast of the Earth's main magnetic field?, *J. Geophys. Res.*, *113*, B08,102, doi: 10.1029/2007JB005199.

Olsen, N., and M. Mandea (2008), Rapidly changing flows in the Earth's core, *Nature Geoscience*, *1*, 390–394.

Sabaka, T., N. Olsen, and M. Purucker (2004), Extending comprehensive models of the Earth's magnetic field with Oersted and CHAMP data, *Geophys. J. Int.*, *159*, 521–547.

Voorhies, C., and G. Backus (1985), Steady flows at the top of the core from geomagnetic field models: The steady motion theorem, *Geophys. Astrophys. Fluid Dynam.*, *32*, 163–

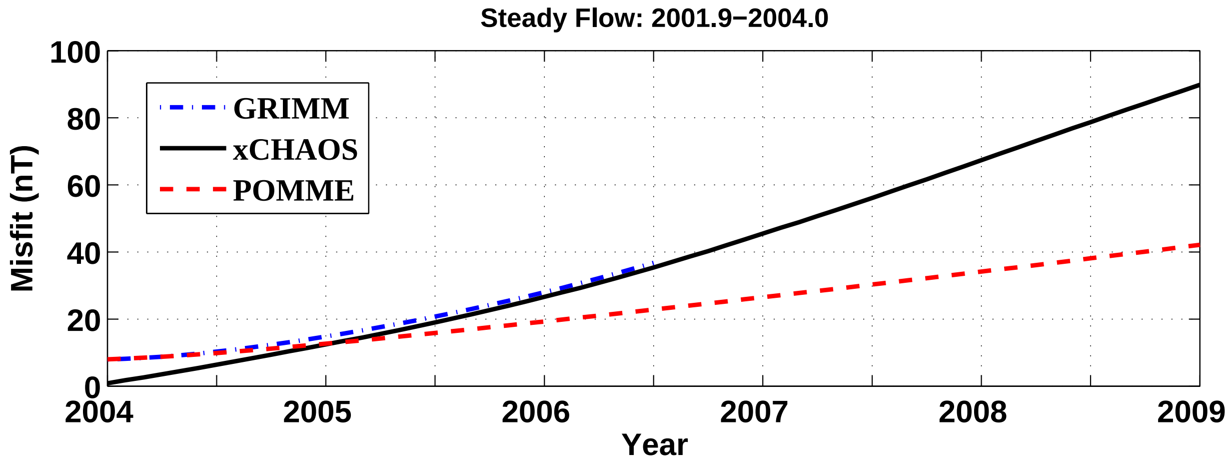

Figure 1. RMS difference (in nT) between the forecast field from a steady flow model generated from data over the period 2001.9–2004.0 and the GRIMM, POMME and xCHAOS satellite field models.

173.

Whaler, K. A. (1986), Geomagnetic evidence for fluid upwelling at the core-mantle boundary, *Geophys. J. R. Astr. Soc.*, *86*, 563–588.

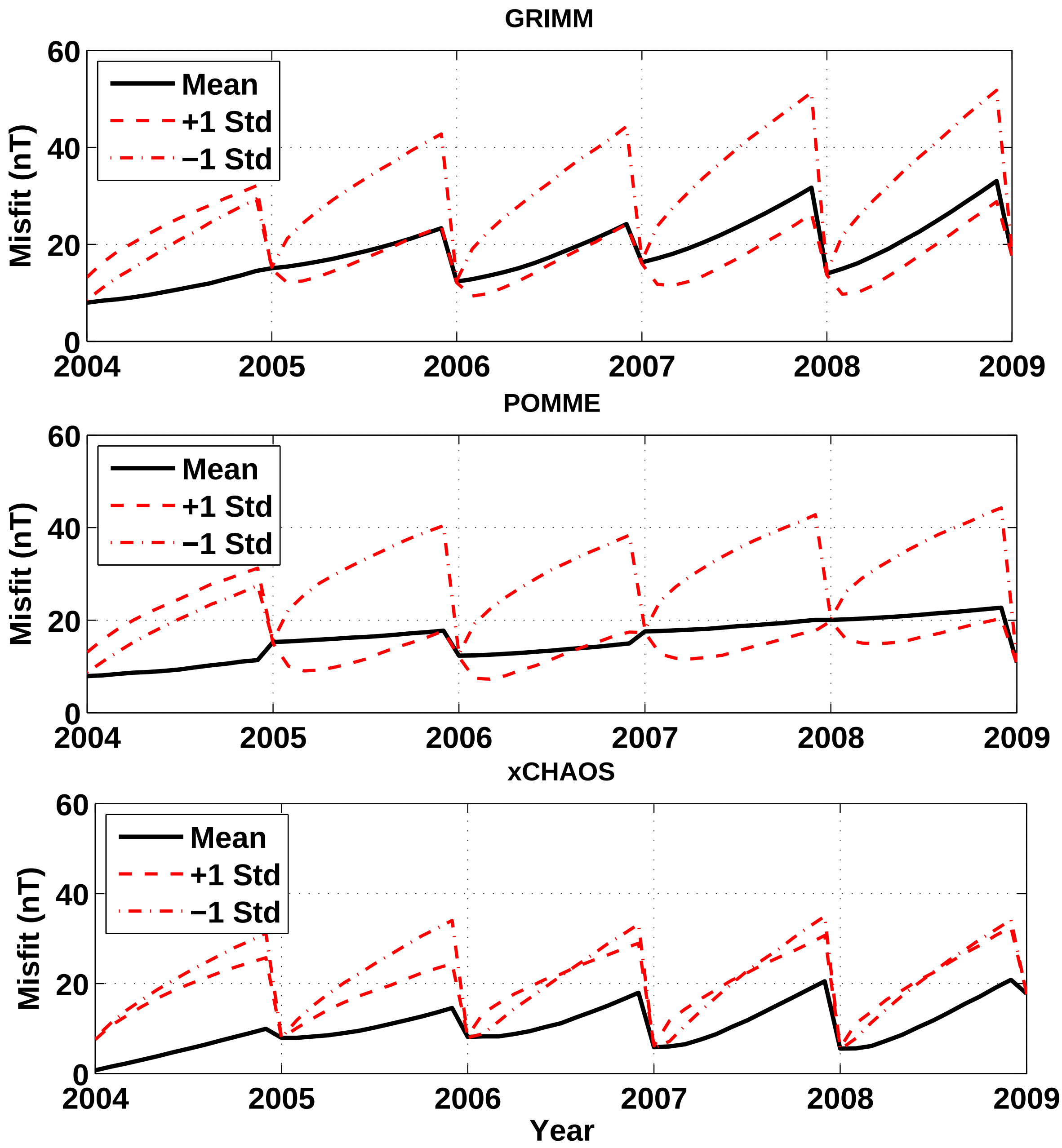

Figure 2. RMS difference (in nT) between a EnKF field forecast with annual assimilation derived from SV generated by a steady flow model from CHAMP satellite data over the period 2001.9–2004.0 and the GRIMM (top panel), POMME (middle panel) and xCHAOS (bottom panel) field models.