

MoSST _ DAS: The First Generation Geomagnetic Data Assimilation Framework

Weijia Kuang^{1,*}, Andrew Tangborn², Weiyuan Jiang², Don Liu²,
Zhibin Sun³, Jeremy Bloxham⁴ and Zigang Wei²

¹ *Planetary Geodynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA.*

² *Joint Center for Earth Systems Technology, University of Maryland-Baltimore County, Baltimore, MD, USA.*

³ *Department of Mathematics and Statistics, University of Maryland-Baltimore County, Baltimore, MD, USA.*

⁴ *Department of Earth and Planetary Science, Harvard University, Cambridge, MA, USA.*

Received 19 September 2006; Accepted (in revised version) 6 June, 2007

Available online 14 September 2007

Abstract. Constraining numerical geodynamo models with surface geomagnetic observations is very important in many respects: it directly helps to improve numerical geodynamo models, and expands their geophysical applications beyond geomagnetism. A successful approach to integrate observations with numerical models is data assimilation, in which Bayesian algorithms are used to combine observational data with model outputs, so that the modified solutions can then be used as initial conditions for forecasts of future physical states. In this paper, we present the first geomagnetic data assimilation framework, which comprises the MoSST core dynamics model, a newly developed data assimilation component (based on ensemble covariance estimation and optimal interpolation), and geomagnetic field models based on paleo, archeo, historical and modern geomagnetic data. The overall architecture, mathematical formulation, numerical algorithms and computational techniques of the framework are discussed. Initial results with 100-year geomagnetic data assimilation and with synthetic data assimilation are presented to demonstrate the operation of the system.

AMS subject classifications: 37M05, 65Z05, 76W05, 86A25

Key words: Geomagnetism, geodynamo, data assimilation, ensemble forecasting.

*Corresponding author. *Email addresses:* Weijia.Kuang-1@nasa.gov (W. Kuang), Tangborn@gmao.gsfc.nasa.gov (A. Tangborn), jiangw@umbc.edu (W. Jiang), donliu@LaTech.edu (D. Liu), sunzhib1@umbc.edu (Z. Sun), Jeremy_bloxham@harvard.edu (J. Bloxham), magstorm@umbc.edu (Z. Wei)

1 Introduction

The Earth has possessed an internal magnetic field (geomagnetic field) through much of its history. It is now widely accepted that this field is generated and maintained by convective flow in the Earth's liquid outer core (geodynamo).

Observation and study of geomagnetism can be traced far back in history. It was perhaps discovered more than 4000 years ago by Chinese [24]. While geomagnetic field properties were recorded decades earlier, one of the earliest scientific theories on the geomagnetism, *De Magnete*, was published by William Gilbert in 1600. Since then, geomagnetic studies have been developed along two separate tracks: understanding the spatial-temporal variation of the geomagnetic field (called the "kinematic track" in this paper), and understanding the origin of the geomagnetic field (called the "dynamic track"), though the latter appeared much later. In the early ages, the kinematic track was the main focus. An example is the work by Gauss on separation of internal and external magnetic fields in 1835. In this approach, the magnetic field is a potential field, and is described by a potential scalar. This scalar can then be represented by a spherical harmonic expansion and can be solved via Laplace's equation (the spectral coefficients in the expansion are called the Gauss coefficients in geomagnetism). The present work is only concerned with the internal field, so that the geomagnetic field in this paper implies only the part of the field originating in the interior of the Earth.

From surface observations it is found that the geomagnetic field varies on time scales ranging from as short as a year (e.g. geomagnetic jerks [7]), several decades (e.g. westward drift [11]), to as long as millions of years and beyond (e.g. field polarity reversal [22]). These can be described by time-varying Gauss coefficients. Combined with the spherical harmonic expansion, the geomagnetic field displays complicated spatial/temporal variation. Indeed, Gauss laid the foundation for modeling the global geomagnetic field.

In the dynamic track, dynamo theory, first proposed nearly 90 years ago [19], has been widely accepted as the most likely explanation for the origin of the geomagnetism. However, due to the complicated, nonlinear magnetohydrodynamic (MHD) processes involved, mathematical solutions (from numerical simulation) of self-consistent geodynamo action have been generated for only about the past decade. A detailed review can be found in [12].

Unfortunately, further interactions between geomagnetic field modeling and geodynamo modeling have been mainly on comparing numerical dynamo simulation results and observations, and possible geodynamic consequences [12]. There has been no attempt to combine geomagnetic field modeling and geodynamo modeling for studying the core dynamics. Perhaps the main reason is that numerical models are believed to be far from accurately simulating the real Earth's core. For example, while they display several properties similar to those of the geomagnetic field derived from surface observations, numerical model solutions cannot be labeled the "geodynamo" solutions: except the dominance of the dipole component at the core-mantle boundary (CMB), the higher

multipole components from the numerical solutions are very different from those observed at/near the surface of the Earth. Therefore, numerical dynamo simulation has been mostly focused on qualitative understandings of the dynamical processes in the Earth's outer core.

The "crudeness" of the numerical models has been mostly attributed to the fact that the parameters used in numerical modeling are far from those appropriate for the Earth's core, provided those values can be reasonably assessed. For example, the Ekman number E , which measures the fluid viscosity with respect to the Coriolis effect, and the magnetic Rossby number R_o , a measure of fluid inertia to the Coriolis effect, are currently at least 3 orders of magnitude larger in numerical models, than those appropriate for the Earth's core (the difference is even larger for the Ekman number E if the molecular value is used). In addition, the Rayleigh number R_{th} that measures the buoyancy force driving the convection is not known exactly for the Earth's core. Lack of detailed knowledge of the outer core also contributes to the model error, as the models are developed based on several physical approximations.

A number of simulations have been performed with a range of parameter values feasible for current computing facilities, with the purpose of understanding possible scaling rules that could lead to a better estimation of the physical state from the numerical model results [12,23]. In particular, such efforts could help us understand the physical properties associated with the parameter values (e.g. force balances in the core, small scale flow patterns) that are important to improve numerical geodynamo models (e.g. parameterization of sub-grid processes).

However, even if we could match the physical parameters appropriate for the Earth's core in a numerical simulation, this still would not imply that the model outputs represent the true physical processes, thereby allowing us to make direct comparisons between the model and observations. A useful analogy is in the field of numerical weather prediction (NWP) where modeling has advanced very far during the past 40 years. In spite of these advances, if one were to simply run an atmospheric model it would produce very realistic looking climate phenomena, including rainfall, storms and heat waves. However, none of these would occur at the time or place that they are actually happening in the real atmosphere. To tie the model to the physical atmosphere, NWP makes use of data assimilation, a general name for mathematical methods for combining model and observations.

Therefore, we believe that assimilation of surface geomagnetic observations to numerical geodynamo models could be the best approach to facilitate interactions between the kinematic and the dynamics tracks of geomagnetic research, thus significantly improving our understandings of the geomagnetism, and the related geophysical processes.

While it will be discussed in further detail later in the paper, data assimilation can be summarized as follows: observations are assimilated into the numerical model output (called the "forecast") in order to form a new "initial" state (called the "analysis"). This process is repeated a large number of times, resulting in numerical solutions that are drawn closer to the true physical states. Data assimilation has been shown to be very suc-

cessful in meteorology and in oceanography. Originally used to form better initial states for weather and climate forecasting, it has also been used to improve models through error diagnosis and parameter estimation.

The current state of geodynamo modeling is in a similar phase of development to NWP three decades earlier: the models are relatively primitive, in both physical and mathematical approximations. This has led to many comments from the community that geodynamo models are “not ready” for data assimilation. However, the example from NWP shows that in fact model development and data assimilation are highly interdependent, and that improvements in one area lead to improvements in the other. Further concerns have been expressed that model errors in geodynamo solutions will make geomagnetic field predictions impossible. Yet it is only through data assimilation that error growth due to model error can even be estimated. This should lead to improvements in numerical models, as it has done in NWP.

To carry out geomagnetic data assimilation, a modeling system (or framework) must include three components: a numerical geodynamo model, a data assimilation component, and a geomagnetic field model. The geodynamo model is developed to solve a set of partial differential equations that govern the dynamical processes in the Earth’s outer core; the data assimilation component provides an appropriate algorithm to combine the forecast and the observation to make the analysis for the geodynamo model; the field model produces a set of parameters (e.g. Gauss coefficients) describing the geomagnetic field at the surface of the Earth.

The data assimilation algorithm component has been under development during the past 3 years, while the other two have existing working models. For example, this system uses the numerical model originally developed by Kuang and Bloxham [14], and later by Kuang and Chao [16] with substantial improvements and revisions in many aspects and is henceforth referred to as the MoSST (Modular Scalable Self-consistent and Three-dimensional) dynamo model. There are several geomagnetic field models for different epochs. A paleomagnetic/archeomagnetic field model developed by Constable *et al* [6], and by Korte and Constable [13] is used for the data period from 5000BC to 1590AD; a geomagnetic field model by Jackson *et al* [10] is used for the observation from 1590AD to 1960; and a geomagnetic field model developed by Langel and Estes [18], and by Sabaka *et al* [26] is used for the period from 1960 to 2005. The actual data records of all models are longer and are being extended, but only the results for the cited periods are used in the present work.

The field models developed from surface and/or near surface observations cover the time varying geomagnetic field over 7000 years. This is two orders of magnitude longer than the time scales of the (theoretically estimated) fastest MHD waves in the Earth’s outer core: the torsional oscillations with the periods of several decades [4]. Thus, it is feasible to use data assimilation as a tool to understand decadal geomagnetic secular variation. These timescales are the focus of the work done using MoSST_DAS.

Two different sequential data assimilation algorithms are included in the assimilation component: an Optimal Interpolation (OI) algorithm with modeled error covariances

(estimates of forecast error), and a second scheme that uses error covariances estimated using a Monte Carlo technique within the same OI framework. The first algorithm is simple and can be easily implemented regardless of the details of a given model, while the latter attempts to provide a statistically optimal analysis for a given numerical model. Both algorithms have their pros and cons in terms of mathematical properties and computational efficiency.

In the past few years, we have carried out many geomagnetic data assimilation experiments, including those using the OI algorithm and with surface geomagnetic observations, Monte-Carlo techniques to estimate error covariance for multivariate data assimilation, and the use of observing system simulation experiments (OSSEs) to test the system. Two papers have been accepted for publication [20, 27], which include many of the details concerning the techniques used. However, we have not yet provided an overall description of the framework (numerical model), MoSST-DAS, used for the research. Therefore, this paper focuses on providing an overview of the framework, including the algorithms, the framework structure, the use of parallel computing. Several results are also reported in this paper to demonstrate the capabilities of the model.

This paper is organized as follows: the mathematics of the three components are given in Section 2. The program structures are described in Section 3. Numerical results and discussions are presented in Sections 4 and 5, respectively.

2 Mathematical formulation

As stated in the introduction, only the assimilation component is newly developed, while the other two have been transplanted from existing models. Therefore, we intend to only summarize the mathematics of the dynamo and the geomagnetic field models while the assimilation algorithms are discussed in more detail.

2.1 Dynamo model

Very simply speaking, an electrically conducting fluid moving in an existing magnetic field generates an electrical current within the fluid. This current induces its own magnetic field. A dynamo action exists if the generated field is maintained without the presence of the (initial) external field.

The geodynamo is in the Earth's fluid outer core, where convection (fluid motion) is driven by the buoyancy force arising from secular cooling of the Earth, in which lighter material and heat are released during the continuing solidification of the inner core. It has been recently argued that radiogenic heat may also be important in driving the geodynamo. Regardless, it can be approximated mathematically by a single density anomaly, which can be due to variations in composition, temperature, or both.

The details of the dynamo model used in this data assimilation system can be found in [15, 16]. Here we only summarize the key mathematical formulations of this model.

It should be pointed out that other dynamo models have been developed with various approximations and algorithms, which could also be used in this system.

In the model, the Earth's outer core is a spherical fluid shell with a finite electrical conductivity σ_c , and is bounded between

$$r_{\text{icb}} \leq r \leq r_{\text{cmb}},$$

where r_{icb} is the mean radius of the inner core boundary (ICB), and r_{cmb} is the mean radius of the core-mantle boundary (CMB). On top of the CMB is a thin layer (D'' -layer) with a much smaller electrical conductivity σ_d , and an upper boundary $r = r_{\text{dp}}$. The remaining solid mantle

$$r_{\text{dp}} \leq r \leq r_{\text{ea}}$$

is assumed electrically insulating (r_{ea} is the mean radius of the Earth).

The equations that describe MHD processes in the outer core are defined in the reference frame co-rotating with the solid mantle with an angular velocity

$$\mathbf{\Omega} = \Omega_0 \mathbf{1}_z + \mathbf{\Omega}_\epsilon, \quad (2.1)$$

where Ω_0 is the mean rotation rate, $\mathbf{1}_z$ is the unit vector of the mean rotation axis, and $\mathbf{\Omega}_\epsilon$ is the tiny variation in the Earth's rotation.

Formulating the geodynamo model in the mantle reference frame is convenient for geomagnetic data assimilation, mainly because the observations are made in this reference frame. Otherwise data reprocessing is necessary to account for any reference frame difference. In this reference frame, the Boussinesq MHD fluid can be described by the following equations [5, 16]

$$\left(\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \right) \mathbf{V} + 2\Omega_0 \mathbf{1}_z \times \mathbf{V} = -\nabla p + \frac{1}{\rho_0} \mathbf{J} \times \mathbf{B} + \frac{\rho}{\rho_0} \mathbf{g} + \nu \nabla^2 \mathbf{V} - 2\mathbf{\Omega}_\epsilon \times \mathbf{V} - \dot{\mathbf{\Omega}}_\epsilon \times \mathbf{r}, \quad (2.2)$$

$$\frac{\partial}{\partial t} \mathbf{B} = \nabla \times (\mathbf{V} \times \mathbf{B}) + \eta_c \nabla^2 \mathbf{B}, \quad (2.3)$$

$$\left(\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \right) \rho = \kappa \nabla^2 \rho, \quad (2.4)$$

where \mathbf{V} is the velocity field, \mathbf{B} is the magnetic field, $\mathbf{J} \equiv (\nabla \times \mathbf{B}) / \mu$ is the current density, ρ_0 is the mean density, ν is the fluid viscosity, κ is the dissipation for the density ρ (e.g. thermal conductivity if the density anomaly arises purely from temperature variation), μ the magnetic permeability, and \mathbf{r} is the position vector. With scaling rules specified in [15], the above equations can be non-dimensionalized as follows,

$$R_o \left(\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \right) \mathbf{V} + \mathbf{1}_z \times \mathbf{V} = -\nabla p + (\nabla \times \mathbf{B}) \times \mathbf{B} + R_{th} \Theta \mathbf{r} + E \nabla^2 \mathbf{V} - R_o (2\boldsymbol{\omega}_\epsilon \times \mathbf{V} + \dot{\boldsymbol{\omega}}_\epsilon \times \mathbf{r}), \quad (2.5)$$

$$\frac{\partial}{\partial t} \mathbf{B} = \nabla \times (\mathbf{V} \times \mathbf{B}) + \nabla^2 \mathbf{B}, \quad (2.6)$$

$$\left(\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \right) \Theta = -(\mathbf{V} \cdot \nabla) T_0 + q_\kappa \nabla^2 \Theta, \quad (2.7)$$

where R_{th} is the Rayleigh number, R_o is the magnetic Rossby number, E is the Ekman number and q_κ is the modified Prandtl number. For the detailed definitions of the parameters, we refer the reader to [15].

All variables in (2.5)-(2.7) are nondimensional, and unless otherwise specified, they remain nondimensional in the rest of the paper.

The magnetic field \mathbf{B} is decoupled into the poloidal \mathbf{B}_P and the toroidal \mathbf{B}_T components:

$$\mathbf{B} = \mathbf{B}_P + \mathbf{B}_T \equiv \nabla \times (\nabla \times P_B \mathbf{1}_r) + \nabla \times (T_B \mathbf{1}_r), \quad (2.8)$$

where P_B and T_B are called the poloidal and toroidal scalars, respectively. The velocity field \mathbf{V} can be also decoupled into poloidal and toroidal components, which are then described by the similar scalars P_V and T_V .

In the model, these scalars are approximated with the following spherical harmonic expansions,

$$\begin{bmatrix} P_B \\ T_B \end{bmatrix} = \sum_{m=0}^M \sum_{l=m}^L \begin{bmatrix} b_l^m(r_i, t) \\ j_l^m(r_i, t) \end{bmatrix} Y_l^m(\theta, \phi) + C.C. \quad \text{for } i=0, 1, \dots, N, \quad (2.9)$$

where (r, θ, ϕ) is the spherical coordinate, Y_l^m are the spherical harmonic functions of degree l and order m . Other variables are expanded similarly as (2.9). The integers (L, M, N) are the truncation order of the numerical model.

It should be pointed out that the toroidal scalar T_B , or more precisely, the coefficients j_l^m vanish in the insulating region. Therefore, they are not observable at the surface of the Earth. Only the poloidal coefficients b_l^m are observable. This will be discussed further in the geomagnetic field modeling subsection.

With the expansion (2.9), a 4th-order compact finite difference algorithm in radius r , and a 3rd order Adams-Bashforth/Adams-Moulton algorithm in time integration [16], the system is reduced to the following linear system:

$$\mathbf{A}_1 \mathbf{x}(t_{k+1}) = \mathbf{A}_2 \mathbf{x}(t_k) + \mathbf{f}(t_k), \quad (2.10)$$

where \mathbf{x} is the state vector that includes all discretized coefficients of the fields, and \mathbf{f} is the forcing vector that includes all nonlinear interaction of the system. The computational aspect of this model will be further discussed in the next section.

2.2 Data assimilation methodology

Data assimilation belongs to the field called *estimation theory* in which the state \mathbf{x} of a system is considered to be a stochastic variable. The purpose of data assimilation is to combine all possible sources of information, generally from models and observations, so as to produce the best possible estimate of \mathbf{x} .

The process can be qualitatively described as follows: a model will produce an estimate of the system at later time $t_a + \delta t_a$ when given an estimate at the analysis time t_a

$$\mathbf{x}^f(t_a + \delta t_a) = \mathbf{M}[\mathbf{x}^f(t_a)], \quad (2.11)$$

where $\mathbf{x}^f(t_a + \delta t_a)$ is called the forecast of the state at time $t_a + \delta t_a$, and \mathbf{M} is the (generally) non-linear numerical model (or discrete propagator). The true state of the system, \mathbf{x}^t , is unknown, and is propagated forward in time by the process

$$\mathbf{x}^t(t_a + \delta t_a) = \mathbf{M}[\mathbf{x}^t(t_a)] + \boldsymbol{\epsilon}(t_a)^m. \quad (2.12)$$

where the random vector $\boldsymbol{\epsilon}(t_a)^m$ describes the model error. It is applied to the true state propagation because there is a component of the true state evolution that cannot be captured by the numerical model.

Observations, \mathbf{y} , are made from the true state, and are modeled as

$$\mathbf{y} = \mathbf{H}\mathbf{x}^t + \boldsymbol{\epsilon}^o, \quad (2.13)$$

where the observation operator, \mathbf{H} , represents the transformation from the state variables to the observed quantities, and $\boldsymbol{\epsilon}^o$ is the observation error. In the present work, geomagnetic observations are the product of a field model and are represented by spherical harmonic coefficients up to certain degree L_{obs} . Because the observations can only be made at or above the Earth's surface and are only sensitive to the poloidal field \mathbf{B}_p (more specifically b_l^m), the matrix \mathbf{H} will act to zero out all of the other variables and \mathbf{B}_p at the grid points beneath r_{dp} . Determination of $b_l^m(r_{dp})$ from surface observations is described in Section 2.3.

The analysis step combines the state forecast and observations by calculating the observed minus forecast values ($\mathcal{O} - \mathcal{F}$) through the innovation equation,

$$(\mathcal{O} - \mathcal{F}) = \mathbf{y} - \mathbf{H}\mathbf{x}^f. \quad (2.14)$$

The $(\mathcal{O} - \mathcal{F})$ values represent the force exerted on the solution by the observations. The analysis equation spreads information from the $(\mathcal{O} - \mathcal{F})$ s back to the model grid through a gain matrix, \mathbf{K} ,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}[\mathbf{y} - \mathbf{H}\mathbf{x}^f]. \quad (2.15)$$

The gain matrix is determined by minimizing the estimated analysis variance (hence this is a minimum variance solution)

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T [\mathbf{H}\mathbf{P}^f \mathbf{H} + \mathbf{R}]^{-1}, \quad (2.16)$$

where \mathbf{P}^f is the forecast error covariance and \mathbf{R} is the observation error covariance.

In traditional OI assimilation schemes, the covariance \mathbf{P}^f is modeled in a way to insure that it retains the essential features of positive definiteness and diagonal dominance. In many cases, the error characteristics are not well understood, and a simple model is the most appropriate solution. However, when it is computationally feasible, a more sophisticated approach may be undertaken using an ensemble calculation of the covariances [8]. In this section we focus on the modeled covariances, and will discuss ensemble calculation in Section 4.2.

The algorithm for the gain matrix \mathbf{K} used in this section is qualitatively the same as an Optimal Interpolation (OI) algorithm in which observations are assumed “perfect”, i.e. no error is included in the data. In addition, no cross-correlation is introduced, so that only the poloidal field coefficients b_l^m for $l \leq L_{\text{obs}}$ are modified. However, this approach includes two arbitrary adjustable parameters to help tune the assimilation system.

For ease of discussion, we describe the algorithm with b_l^m , not in the more standard vector-matrix format shown in equation (2.15). The first adjustable parameter is the scaling factor α_b between the “observation” \mathbf{y} in (2.15) and the poloidal field coefficients $b_l^m(r_{\text{dp}})$ continued downward from surface observations. The other is the radial correlation length r_c in the outer core such that the analysis (2.15) only modifies the poloidal field coefficients b_l^m in the domain $r_{\text{dp}} \geq r \geq (r_{\text{cmb}} - r_c) \equiv r^*$. The details of derivation is given in a manuscript submitted separately. Here we only provide the final formula of the analysis:

$$b_l^{m(a)}(r) = \begin{cases} b_l^{m(f)}(r), & r < r^*, \\ b_l^{m(f)}(r) + h(r)\delta b_l^m [1 - l(r/r_{\text{dp}} - 1)], & r_{\text{dp}} > r \geq r^*, \end{cases} \quad (2.17)$$

where

$$\begin{aligned} \delta b_l^m &= \alpha b_l^{m(o)}(r_{\text{dp}}) - b_l^{m(f)}(r_{\text{dp}}), \\ h(r) &= \frac{(r - r^*)^2}{(r_{\text{dp}} - r^*)^3} [3(r_{\text{dp}} - r^*) - 2(r - r^*)]. \end{aligned} \quad (2.18)$$

As in the previous discussion, the superscripts “a”, “f” and “o” represent the coefficients of analysis, forecast and observation, respectively. From (2.17) and (2.18) one can find that the poloidal field coefficients b_l^m and their first order radial derivatives $\partial b_l^m / \partial r$ are continuous at $r = r^*$.

In this analysis, α is often chosen to be the ratio of the dipole coefficients of the observation and forecast

$$\alpha = b_1^{0(f)} / b_1^{0(o)}.$$

But other definitions can be used to ensure that the scaled coefficients $\alpha b_l^{m(o)}$ are comparable in magnitude to $b_l^{m(f)}$ of the numerical model output.

2.3 Geomagnetic field model

Magnetic measurements at and/or near the Earth’s surface include contributions from various sources, e.g. the magnetic field from the Earth’s core, the remnant magnetism in the crust, the induced field in the crust and mantle (due to the time varying magnetic environment), and the external fields (from ionosphere, magnetosphere and beyond) etc [26,28].

The magnetic field generated in the core contributes more than 90% of the magnetic signals at the surface, and is often referred to as the *Main Field*. The main field varies

slowly with time scales from a few years to millions of years [2]. Several models have been developed for the time-varying main field, using different measurements and over different epochs [6,10,26]. In this paper, we briefly describe the mathematical foundations used in these field models.

In the geomagnetic field model, \mathbf{B} is divided into the internal and external field:

$$\mathbf{B} = \mathbf{B}_I + \mathbf{B}_E. \quad (2.19)$$

In a source-free region (electrically insulating), it is a potential field, i.e.,

$$\mathbf{B} = -\nabla\Phi(r, \theta, \phi, t) = -\nabla\Phi_I - \nabla\Phi_E. \quad (2.20)$$

Since it is also divergence-free, this potential satisfies the Laplace equation

$$\nabla^2\Phi = 0. \quad (2.21)$$

In a spherical geometry, the solution can be conveniently described by the following spherical harmonic expansion:

$$\begin{aligned} \Phi(r, \theta, \phi, t) &= \Phi_I + \Phi_E \\ &= r_{\text{ea}} \sum_{l=1}^{L_i} \sum_{m=0}^l \left(\frac{r_{\text{ea}}}{r}\right)^{l+1} [g_l^m(t) \cos(m\phi) + h_l^m \sin(m\phi)] P_l^m(\cos\theta) \\ &\quad + r_{\text{ea}} \sum_{l=1}^{L_e} \sum_{m=0}^l \left(\frac{r}{r_{\text{ea}}}\right)^l [s_l^m(t) \cos(m\phi) + q_l^m \sin(m\phi)] P_l^m(\cos\theta), \end{aligned} \quad (2.22)$$

where (g_l^m, h_l^m) are the internal Gauss coefficients at $r = r_{\text{ea}}$, and (s_l^m, q_l^m) are the external Gauss coefficients. It should be pointed out that conventionally $P_l^m(\cos\theta)$ in (2.22) are the Schmidt-normalized associated Legendre polynomials of degree l and order m . L_i and L_e are the maximum degrees of the modeled internal and external fields, respectively.

The measured quantities differ in both type and accuracy over time. In the earliest measurements, the field declination D (the angle between the true north and the local field direction), and the inclination I (the dip of the field direction from the horizontal) were measured at the surface [24]. After Gauss developed a method for measuring absolute intensities in 1830s, the horizontal intensity H and the total intensity F were observed. With advances in technology, all three components of the magnetic field (the vector field) can now be measured accurately [17].

Since we are only interested in the main field, we focus on the coefficients g_l^m and h_l^m in (2.22). These coefficients are obtained via a minimization process [17]. Denote by C_i a measured quantity (e.g. a field component, the field intensity, or the field orientation) at the location \mathbf{r}_i , and $C(\mathbf{r})$ the global approximation to the measured quantity. A cost function

$$D^2 = \sum_i [C_i - C(\mathbf{r}_i)]^2 w_i \quad (2.23)$$

can be constructed. Obviously $C(\mathbf{r})$ and hence D^2 are functions of the Gauss coefficients. These coefficients are then determined by minimizing the cost function D^2 with respect to (g_l^m, h_l^m) . Naturally, different field models may be developed based on different measured quantities C_i and different cost functions D^2 .

After determining $g_l^m(t)$ and $h_l^m(t)$ at the Earth's surface, they can be downward continued to the top of the D'' -layer via (2.22). These coefficients are different from those used in geodynamo modeling (2.9). Their relationship can be derived from the radial component B_r of the main field \mathbf{B}_I . By (2.22),

$$\begin{aligned} B_r &\equiv -\frac{\partial}{\partial r}\Phi_I \\ &= \sum_{l=1}^L \sum_{m=0}^l (l+1) \left(\frac{r_{ea}}{r}\right)^{l+2} [g_l^m(t) \cos m\varphi + h_l^m(t) \sin m\varphi] P_l^m(\cos\theta). \end{aligned} \quad (2.24)$$

Obviously, if r_{ea} in (2.24) is read as the mean radius of the Earth scaled by the mean radius r_{cmb} of the CMB, then r is nondimensional. By (2.9), B_r can be also expressed as

$$B_r = \sum_{l,m} \frac{l(l+1)}{r^2} b_l^m Y_l^m(\theta, \phi) + C.C.. \quad (2.25)$$

At the top of the D'' -layer, (2.24) and (2.25) should be identical, leading to

$$b_l^m(r_{dp}) = \frac{r_{ea}^2}{l} \left(\frac{r_{ea}}{r_{dp}}\right)^l \sqrt{\frac{2\pi}{\delta_m(2l+1)}} (g_l^m - ih_l^m), \quad (2.26)$$

where r_{dp} is the scaled (relative to r_{cmb}) mean radius of the top of the D'' -layer, and

$$\delta_m = \begin{cases} 1, & \text{if } m > 0, \\ 2, & \text{if } m = 0. \end{cases}$$

It should be pointed out that (g_l^m, h_l^m) include contributions from both the core field and the crustal field that are not separable. In general, lower degree coefficients are dominantly from the core field. Careful examination of satellite data suggests that 87% of the degree $l = 12$ coefficients can be attributed to the core field [21]. This attribution decreases to 52% at the degree $l = 13$. Above that the crustal field dominates. Consequently, the degree $l = 13$ is considered the transition point from the core dominance to the crustal dominance. However, earlier observations could only provide the coefficients up to lower degrees [6, 10]. Therefore $L_{obs} \leq 13$ is chosen in our studies.

3 Framework architecture and computation

The geodynamo and geomagnetic field models used in MoSST_DAS are transplanted from existing models. Considering independent efforts on numerical dynamo model

and geomagnetic field model development, the framework architecture is designed with two basic requirements: easy integration (minimum re-engineering) of external models into the framework; and efficient maintenance for users and developers. Obviously a modular based structure is sufficient for the two requirements.

The computational requirement (in both CPU time and data storage) for the geomagnetic data assimilation is enormous, several orders of magnitude larger than simple dynamo simulation (which is already very computationally intensive). Exploring new technologies is important to help meet the computing demand with available systems. We discuss mainly the parallelization of the dynamo simulation on a single distributed system. The parallel computation of the ensemble simulation over networked systems is also briefly described.

3.1 Program structure

The basic building blocks of MoSST_DAS are individual modules. They are divided into three levels according to their functions. The Level_1 modules are those defining state variables and model parameters. The Level_2 modules are those defining the quantities derived from those given in the first level modules. These derived quantities are often used for simulation, and for model results analysis. The Level_3 modules are those for dedicated applications, e.g. dynamo simulation, geomagnetic field modeling and assimilation algorithms. Any module can use only the modules in the same level or in the lower levels.

An example of the Level_1 modules is that to define the truncation orders for the numerical simulation:

```

MODULE mod_dimparam
  implicit none
  integer Lmax_v,mmax_v,miner
  integer nmaxo,nmaxi,nmaxm,nmax_v
  ... ..
END MODULE mod_dimparam

```

An example of the Level_2 modules is that for defining and evaluating variables in the spherical geometries:

```

MODULE mod_sphgeoms
  use mod_dimparam
  implicit none
  real (kind=8), allocatable :: aslg(:,:,:), clm(:,:,:)
  ... ..
CONTAINS
  ... ..
  SUBROUTINE sphgeoms

```

```

      ... ..
      END SUBROUTINE sphgeoms
      ... ..
      MODULE mod_sphgeoms

```

In this example, a Level_1 module, `mod_dimparam`, is used. The Level_3 modules have similar structures, but with more complicated functions. For example, some subroutines are `private` (not accessible from external), while others are `public` and can be called by external routines. In particular, with slight modification, Level_3 modules can be easily converted to independent numerical models.

The main code becomes very simple. It comprises of the simple commands of use various modules, and various options for users to choose. Specific applications can be simply added to the framework by adding/modifying individual modules.

3.2 Parallel computing

Distributed systems have become the main tool for high-end scientific computation. Therefore, one of our major tasks is to re-engineer the MoSST core dynamics model based on MPI libraries for distributed systems.

As discussed previously, in the simulation the linear system (2.10) is updated at each time step. Therefore, the computation is divided into two parts: (i) evaluating the nonlinear forcing \mathbf{f} , and (ii) solving the set of linear equations (2.10). The details of the dynamo model are given in [15]. We focus here only on the properties necessary for the parallel computing discussion.

Computation of \mathbf{f} can be carried out independently at different radial grid points r_i , given the spectral coefficients defined at the point. The numerical procedure \mathcal{W}_i can be summarized as

$$\mathcal{W}_i \Rightarrow \left\{ \begin{array}{l} \text{Input:} \quad \mathbf{x}_i \equiv [x_1^0(r_i), x_2^0(r_i), \dots, x_l^m(r_i), \dots, x_L^M(r_i)]^T, \\ \text{Call Spherical Transforms} \\ \text{Output:} \quad \mathbf{f}_i \equiv [f_1^0(r_i), f_2^0(r_i), \dots, f_l^m(r_i), \dots, f_L^M(r_i)]^T, \end{array} \right. \quad (3.1)$$

The superscript “ T ” in the above descriptions implies the transpose of the array. Therefore, in our model, all \mathcal{W}_i are evenly distributed among the nodes allocated for communication.

The matrices \mathbf{A}_1 and \mathbf{A}_2 in (2.10) are block diagonal in the spectral space (l, m) [15],

e.g.

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{A}_1^{10} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1^{20} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_1^{lm} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{A}_1^{LM} \end{pmatrix}, \quad (3.2)$$

and similarly for \mathbf{A}_2 . Each \mathbf{A}_1^{lm} in (3.2) is an $N \times N$ sub-matrix. Therefore (2.10) can be solved via the following independent procedures \mathcal{Q}_l^m :

$$\mathcal{Q}_l^m \Rightarrow \begin{cases} \text{Input:} & \mathbf{x}_l^m(t_k) \equiv [x_l^m(r_1), x_l^m(r_2), \dots, x_l^m(r_N)]^T(t_k), \\ & \mathbf{f}_l^m(t_k) \equiv [f_l^m(r_1), f_l^m(r_2), \dots, f_l^m(r_N)]^T(t_k); \\ \\ & \text{Solve } \mathbf{A}_1^{lm} \mathbf{x}_l^m(t_{k+1}) = \mathbf{A}_2^{lm} \mathbf{x}_l^m(t_k) + \mathbf{f}_l^m(t_k); \\ \\ \text{Output:} & \mathbf{x}_l^m(t_{k+1}). \end{cases} \quad (3.3)$$

In the above expression, \mathbf{x}_l^m and \mathbf{f}_l^m are the subsets of \mathbf{x} and \mathbf{f} for given (l, m) , respectively.

Obviously \mathcal{W}_i and \mathcal{Q}_i^m require different subsets of the state vector \mathbf{x} . And data communication between the two parts of calculation is not simple.

We have tested two approaches in the data communication. One is the “master-slave” architecture (MSA), in which a single node (the “master” node) is dedicated to manage the data distribution and redistribution between \mathcal{W} and \mathcal{Q} , while the rest of the nodes (the “slave” nodes) carry out the actual computation in \mathcal{W} and \mathcal{Q} . This architecture is very simple for implementation. But the scalability is poor: the communication time depends only on the bandwidth between the master node and the slave nodes.

The other approach is based on the “divide-and-conquer” scheme (DCA). In this architecture, all nodes perform the dual functions of computation and communication. A node acquires from the rest only the portion of data it needs for computation. Synchronization among the nodes are ensured by the pair of MPI SEND and RECEIVE functions. Implementation of this architecture is more complicated. But it is scalable, and the effective bandwidth increases with the node number in simulation.

Since the ensemble runs in error covariance analysis are independent from each other, and since only the final results from the individual runs are required to obtain the covariance matrix, this part of computation can therefore be distributed among individual computing systems connected via networks. We have tested this approach successfully, using XCAT3 framework, a Common Component Architecture (CCA)-compliant framework, to manage job distribution and model output collection. The details of the experiment are given in [29]. With this parallel computing over networked systems, computation needs for the geomagnetic data assimilation could be reasonably accommodated with currently available technologies.

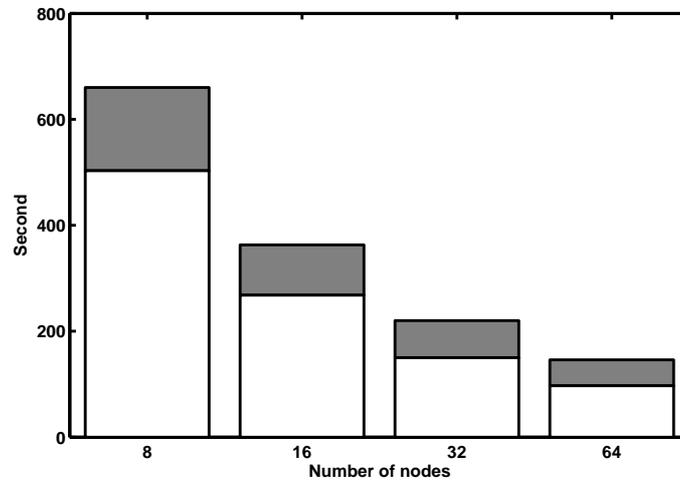


Figure 1: Scalability Test with up to 64 nodes. The total time steps $K=50$. The resolution is $62 \times 40 \times 110$. The columns are the wall clock time for 8, 16, 32 and 64 nodes. Note that the communication time (dark blocks) reduces as the node number increases.

4 Some numerical results from MoSST_DAS

This section is focused on the code parallelization, the covariance matrix from the ensemble simulations, the experiments with synthetic data (OSSEs), and 100 years of surface geomagnetic observations. Much of the work has been reported in manuscripts submitted for publication. Therefore the selected results are used to demonstrate the operation of the framework.

4.1 Scalability of the MoSST model

Two approaches are employed for data communication in the MPI-based MoSST core dynamics model, the “Master-Slave” structure and the “Divide-and-Conquer” algorithm. Both are tested for the scalability of the code.

The computing system used for this test is a 256-node, dual processor, Linux cluster. Up to 64 nodes are used for these experiments, and the numerical resolution used is $L \times M \times N = 60 \times 40 \times 110$. The overall CPU time is divided into communication time τ_{co} and the data crunching time τ_{dc} (for evaluating nonlinear terms and solving the linear system). In all cases there are 50 steps of time integration.

The overall performance of the first approach is as expected, τ_{dc} decreases as the number of nodes used in simulation increases. But τ_{co} remains approximately constant during the process. In other words, the communication in the first approach is not scalable.

The results for the second approach are shown in Fig. 1. From the figure we can observe clearly that τ_{co} (the red sectors of the columns) is approximately halved when

the number of the nodes doubles. Reduction of τ_{co} is slowing down when more nodes are used. Similar properties are also observed in τ_{dc} (the white sectors of the columns).

The scalability at large node numbers (e.g. from 32 to 64) is not as good as that at smaller node numbers (e.g. from 8 to 16) is partly due to unbalanced workload in parallelization of nonlinear forcing evaluation. For example, the velocity field and the temperature perturbation are evaluated only in the outer core, while the magnetic field must be solved additionally in the inner core and in the D'' -layer. This could be improved by proper redistribution of the calculations in the inner core and in the D'' -layer among the individual nodes.

4.2 Monte-Carlo estimation of background error covariance

In this section we discuss the use of a Monte Carlo method to estimate the multivariate forecast error covariance [3]. Because of the large number of degrees of freedom in the model, it is not yet feasible to run the system as an Ensemble Kalman filter. Instead, we make the approximation that the error statics can be approximated as constant over the assimilation time. The estimated error covariance will then have information on spatial error correlations and correlations between the observed poloidal field to the unobserved variables.

The covariance is defined in terms of the nondimensional model state vector for a given (l, m) , which is

$$\mathbf{y} = \begin{bmatrix} (\mathbf{P}_V - \boldsymbol{\mu}_{P_V}) / \sigma_{P_V} \\ (\mathbf{T}_V - \boldsymbol{\mu}_{T_V}) / \sigma_{T_V} \\ (\mathbf{P}_B - \boldsymbol{\mu}_{P_B}) / \sigma_{P_B} \\ (\mathbf{T}_B - \boldsymbol{\mu}_{T_B}) / \sigma_{T_B} \\ (\boldsymbol{\Theta} - \boldsymbol{\mu}_{\Theta}) / \sigma_{\Theta} \end{bmatrix}, \quad (4.1)$$

where

$$(\boldsymbol{\mu}_{P_V}, \boldsymbol{\mu}_{T_V}, \boldsymbol{\mu}_{P_B}, \boldsymbol{\mu}_{T_B}, \boldsymbol{\mu}_{\Theta}) = (\langle \mathbf{P}_V \rangle, \langle \mathbf{T}_V \rangle, \langle \mathbf{P}_B \rangle, \langle \mathbf{T}_B \rangle, \langle \boldsymbol{\Theta} \rangle)$$

are the ensemble means of the state vectors, and $\sigma_{[P_V, T_V, P_B, T_B, \Theta]}$ are the (radially) maximum standard deviations in the outer core of the vectors \mathbf{P}_V , \mathbf{T}_V , \mathbf{P}_B , \mathbf{T}_B and $\boldsymbol{\Theta}$, respectively.

The forecast error covariance matrix is

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}^{P_V, P_V} & \mathbf{P}^{P_V, T_V} & \mathbf{P}^{P_V, P_B} & \mathbf{P}^{P_V, T_B} & \mathbf{P}^{P_V, \Theta} \\ \mathbf{P}^{T_V, P_V} & \mathbf{P}^{T_V, T_V} & \mathbf{P}^{T_V, P_B} & \mathbf{P}^{T_V, T_B} & \mathbf{P}^{T_V, \Theta} \\ \mathbf{P}^{P_B, P_V} & \mathbf{P}^{P_B, T_V} & \mathbf{P}^{P_B, P_B} & \mathbf{P}^{P_B, T_B} & \mathbf{P}^{P_B, \Theta} \\ \mathbf{P}^{T_B, P_V} & \mathbf{P}^{T_B, T_V} & \mathbf{P}^{T_B, P_B} & \mathbf{P}^{T_B, T_B} & \mathbf{P}^{T_B, \Theta} \\ \mathbf{P}^{\Theta, P_V} & \mathbf{P}^{\Theta, T_V} & \mathbf{P}^{\Theta, P_B} & \mathbf{P}^{\Theta, T_B} & \mathbf{P}^{\Theta, \Theta} \end{bmatrix}. \quad (4.2)$$

If the matrix $\mathbf{Y}^{N \times M}$ contains an M -member ensemble of model-state vectors as columns, then \mathbf{P} can be calculated as

$$\mathbf{P}^{N \times N} = \frac{\mathbf{Y}\mathbf{Y}^T}{M-1}. \quad (4.3)$$

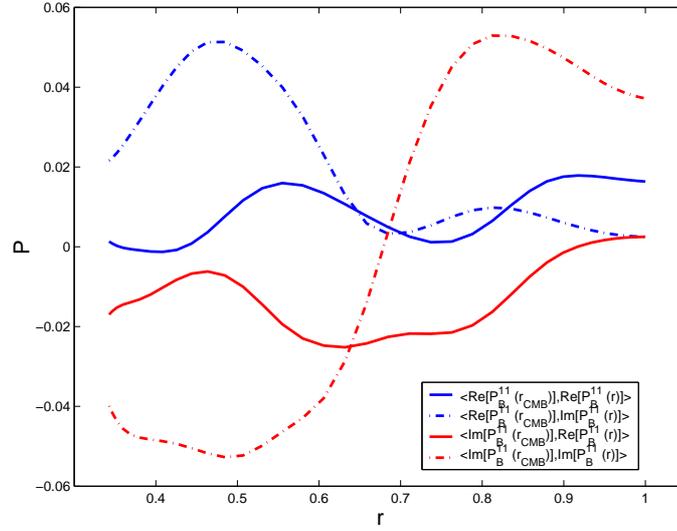


Figure 2: Original covariance between the poloidal field in the core and on the CMB, $(l,m) = (1,1)$.

An ensemble of 50 model runs with perturbed initial states is carried out in our study, with each run lasting 0.04 time units (on the magnetic diffusive timescale). This ensemble size is substantially smaller than the model dimensions, which will likely result in some spurious covariances at large distances. The covariances therefore be truncated at some correlation length c , and we use the piecewise continuous function proposed by [9]

$$C_0\left(z, \frac{1}{2}, c\right) = \begin{cases} 0, & 2c \leq |z|, \\ -\frac{1}{4}\left(\frac{|z|}{c}\right)^5 + \frac{1}{2}\left(\frac{|z|}{c}\right)^4 + \frac{5}{8}\left(\frac{|z|}{c}\right)^3 - \frac{5}{3}\left(\frac{|z|}{c}\right)^2 + 1, & 0 \leq |z| \leq c, \\ \frac{1}{12}\left(\frac{|z|}{c}\right)^5 - \frac{1}{2}\left(\frac{|z|}{c}\right)^4 + \frac{5}{8}\left(\frac{|z|}{c}\right)^3 + \frac{5}{3}\left(\frac{|z|}{c}\right)^2 - 5\left(\frac{|z|}{c}\right) + 4 - \frac{2}{3}\left(\frac{|z|}{c}\right)^{-1}, & c \leq |z| \leq 2c, \end{cases} \quad (4.4)$$

to smooth \mathbf{P} . This results in C_0 being a homogeneous and isotropic correlation function where z is the distance between two radial grid points. We compute the smoothing \mathbf{P} by multiplying each item with $C_0(z_{ij}, 1/2, c)$, given a constant c (e.g. $c = 10\%$ of the outer core thickness in this study). From this we obtain the new truncated covariance matrix \mathbf{P} . As an example, one part of the covariance and its filtered result are shown in Figs. 2 and 3.

4.3 Observing System Simulation Experiments

In order to demonstrate the overall effect as well as detailed improvement, Observing System Simulation Experiments (OSSEs) based on an Optimal Interpolation (OI) scheme, have been performed to assimilate synthetic geomagnetic observation into the geodynamo simulation system. One advantage of OSSEs lies in that it is possible to evaluate

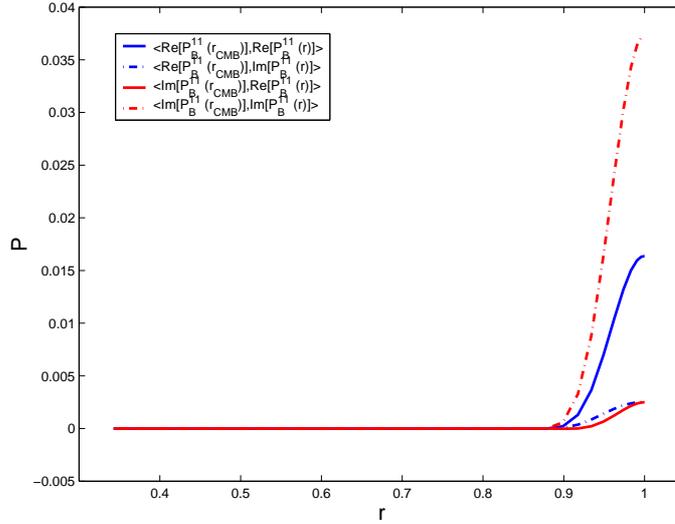


Figure 3: Smooth covariance between the poloidal field in the core and on the CMB, $(l,m)=(1,1)$.

the local as well as global RMS error in terms of the observed (\mathbf{B}_P) and unobserved (\mathbf{B}_T) variables during any computational temporal as well as spatial domain. Therefore the improvement can be visualized and analyzed numerically.

In these OSSEs, a **nature** run as well as a **free model** run, using Rayleigh number $R_{th}=15,000$ and $14,500$ respectively, have been obtained before a **forecast** run starts, so that two references are available before hand to compare with. Other parameters are specified as:

$$R_o = E = 1.25 \times 10^{-6}, \quad q_\kappa = 1.0. \quad (4.5)$$

Therefore, the **model error** here refers to the difference in the Rayleigh numbers. A number of OSSEs have been carried out using different scenarios, such as with and without model error; different observation degrees of spherical harmonic coefficients of the synthetic observation data; various forecast error correlation length scales; and varying assimilation time frames.

Through consistent data assimilation over a portion of the magnetic free decay time (τ_d), both the poloidal and toroidal magnetic fields show a positive impact - the RMS error is well bounded and in some period even consistently decreasing. A relatively long error correlation length has a slightly better effect on improving the velocity field than a shorter one.

Fig. 4(a) and (b) show the relative RMS error of 3 sets of data: free model run, forecast run (assimilating the first 8 degrees of observation data) started at $t=0.1$ and 0.95 respectively. The forecast error correlation length scale is set to be $r_c=120$ km. Fig. 4(a) has both the initial error and model error; Fig. 4(b) has only model error there. These two plots demonstrate that the overall error for \mathbf{B}_P in core mantle boundary is bounded over one time unit. At analysis times, the error is greatly reduced by assimilating the observation

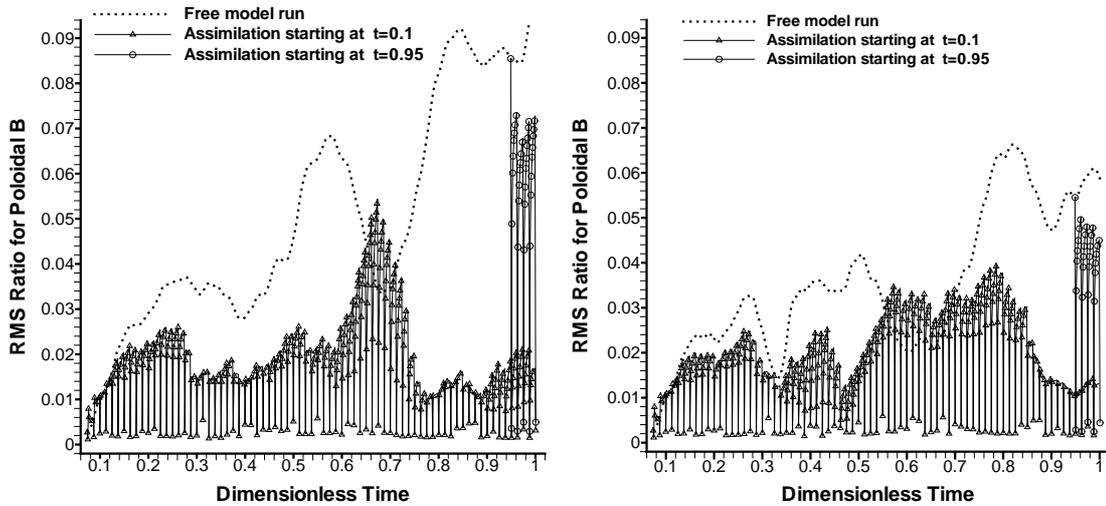


Figure 4: Relative RMS error in poloidal component of \mathbf{B} inside core mantle boundary versus time. Assimilation started at $t = 0.1$ (triangles) and 0.95 (squares) of magnetic free decay time respectively. Assimilation with both model and initial error is on the left while that with only the initial error is on the right. The error for the free model run is shown as a dotted line.

data. Between assimilation times, error forecast errors grow, but are generally lower after repeating the assimilation many times. This is a good indication that by periodically assimilating geomagnetic observation data, the geodynamo simulation is drawn closer to the true state.

Fig. 5 presents the magnetic field morphology at the top of the D'' -layer for the nature run (true solution) on the left, and the forecast run with assimilation on the right, at time $t = 0.994$, nearly in the middle between two assimilations. The dipole component of magnetic field which is about 80% of the total magnitude has been excluded to see the small structures. The assimilation started from time $t = 0.1$, using the first 8 degrees of observational data, without model error. The forecast error correlation length is $r_c = 120\text{km}$. Since the previous assimilation, although the forecast solution has developed without constraint for over 0.3% of magnetic free decay time, it still demonstrates similar small scale structure compared to the true solution on the left. This also tells us that observing the first 8 degrees of observational data is able to capture a substantial portion of the magnetic field structure. These results indicate that OSSEs is a useful tool for evaluating a geomagnetic data assimilation system.

It should be pointed out that if the molecular dissipative coefficients of the core fluid are used, the corresponding parameter values would be

$$R_0 \approx 10^{-9}, \quad E \approx 10^{-15}, \quad q_\kappa \approx 10^{-6},$$

thus orders of magnitude smaller than those (4.5) in our experiment. The effect of these parameter differences on geomagnetic data assimilation can be also analyzed similarly via carefully designed OSSEs.

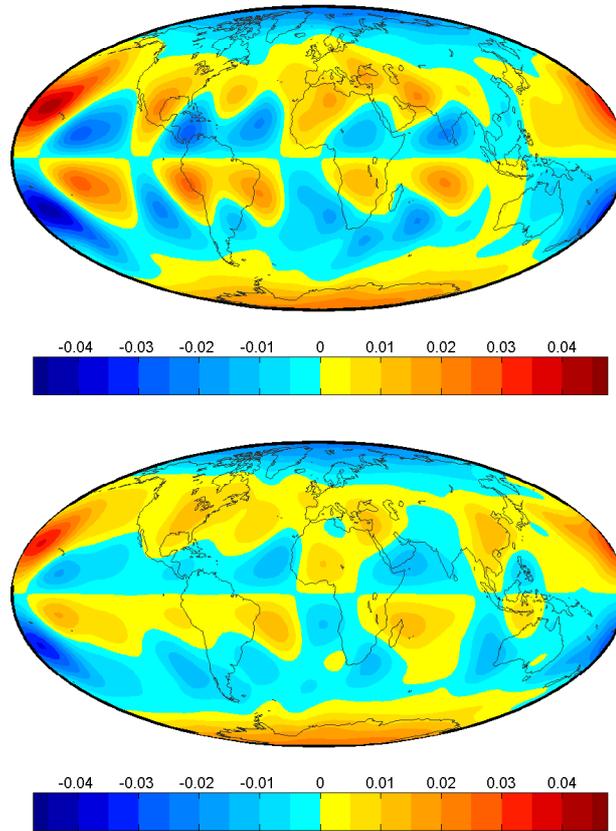


Figure 5: Surface morphology of B_r at $t=0.992$ for (a) the true solution; (b) forecast run with assimilation (assimilating the first 8 degrees of observation data $L_{obs}=8$), forecast error correlation length $r_c=120\text{km}$, without model error.

4.4 Assimilating 100 years of geomagnetic observations

We have carried out an experiment in which geomagnetic observations are assimilated into the system every 20 years, for a period of 80 years (1900 to 1980). A final forecast run is started in 1980. We used the scaled OI scheme described in Section 2.2, in which observations of the poloidal field are scaled relative to the dipole field ($l=1, m=0$). In this algorithm, corrections are only made to the poloidal component of the magnetic field, so that changes in other variables can only occur through the dynamo model itself. The goal of these experiments is to determine whether there is some improvement in the estimate of the core state resulting from this assimilation. Because the true state is not known, any improvement can only be inferred indirectly by comparing the error growth rate (or difference between observation and forecast) for the surface magnetic field at the first and last assimilation cycles. That is, if a 20 year forecast of the surface magnetic field is more accurate after 80 years of assimilation than it is after just one assimilation, then there is

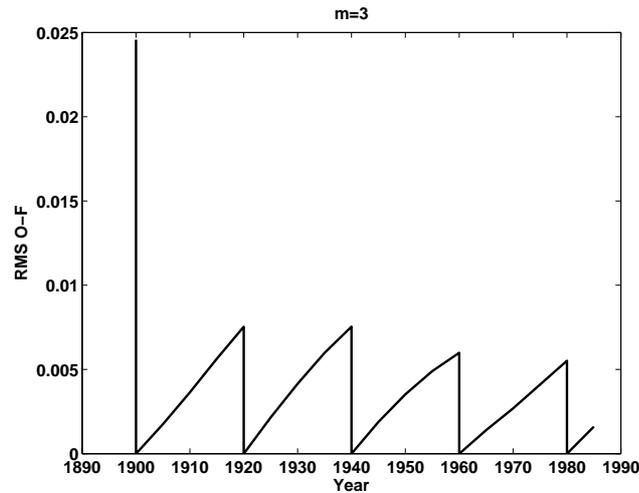


Figure 6: RMS of $(\mathcal{O}-\mathcal{F})$ for the $m=3$ wave number (all L) between 1900 and 1985. Assimilation is carried out every 20 years, and the $(\mathcal{O}-\mathcal{F})$ is calculated every 5 years. The error growth rate shows a modest decline of about 15% over the course of the assimilation run.

evidence that the initial state for the forecast has been improved.

The RMS difference between the observations and forecast for a given wave number m

$$RMS_m = \left[\frac{1}{L-m+1} \sum_{l=m}^L (b_l^{m(f)} - b_l^{m(o)}) \right]^{1/2} \quad (4.6)$$

is one possible indicator of the forecast error growth rate (similar choice can be the RMS for a given degree l). Since the observation error is assumed zero, the RMS error at analysis times is also zero. The RMS growth is plotted in Fig. 6 for $m=3$, which shows a roughly 15% drop in the RMS error growth for the 1980 forecast relative to the 1920 forecast.

While this improvement is small, it does indicate that some change in the core dynamics has taken place which has the 1980 analysis a better initial state than the 1900 analysis. While there are no other kinds of observations to compare with, we can determine the kind of impact on other state variables by comparing the model states with and without assimilation. The toroidal component of magnetic field, for example, is impacted indirectly by the observations through the induction equation. Fig. 7 shows the change in a single profile of the toroidal field as a function of radius. The maximum change to this field occurs around $r=2000\text{km}$, which is well inside the outer core ($r_{cmb}=3500\text{ km}$ in this model run). This represents roughly a 10% change to the field at this location. Although it cannot be determined whether this represents an improvement or worsening of the estimate of the toroidal component, it does show that the surface observations do begin to have some impact on non observed variables within the core, and the improvements to the magnetic field forecasts indicate that this impact is not negative.

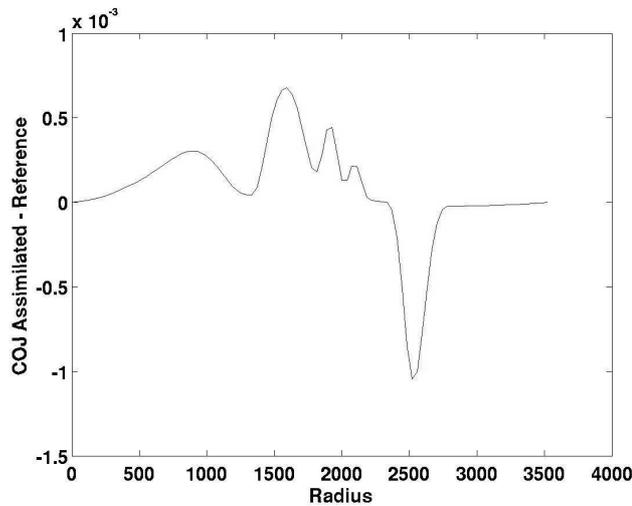


Figure 7: Change in profile of the toroidal component of magnetic field (assimilation - free running model) for the year 1980.

5 Discussion

In this paper, we have described the mathematical foundation, numerical algorithms, programming structure, and several numerical results of the geomagnetic data assimilation framework, MoSST-DAS, that we have been developing during the past several years.

The framework is numerically stable: numerical solutions from Sections 4.3 and 4.4 show that model outputs (numerical solutions) are stable over a very long period of time (one magnetic free-decay time) with the constraints imposed at the top of the D'' -layer. The numerical results also show that the model outputs do respond significantly to the imposed constraints from surface observations. However, the degree of response varies at different locations and for different fields. For example, the poloidal field near the top of the CMB responds quickly to the surface observations, as shown in Figs. 5 and 6. On the other hand, the toroidal field (not observable at the Earth's surface) changes slightly over a short (100 years) assimilation time, e.g. Fig. 7. The OSSEs also show that the improvement of the model forecast over the assimilation time period: the errors between the forecast and the "truth" decay with time, as demonstrated in Fig. 4.

These results only demonstrate the successful operation of the MoSST-DAS. Many problems still remain to be solved for a fully functional system. First, not all field models have been completely integrated into the framework. One of the key issues for implementation is the consistency of the various field models. Different field models generate the main field Gauss coefficients from different surface observations, and with different optimization algorithms. Consequently, the degrees of the available Gauss coefficients vary in time. We are investigating the impact of the varying degrees of the observed

poloidal field on geomagnetic data assimilation with carefully designed OSSEs.

In the OSSEs and the geomagnetic data assimilation experiments presented in this paper, we have assumed (as the first step) that the observations are precise, i.e. no observation error is included in the framework. In reality, observations error is non-zero, and they depend on observational techniques, field modeling algorithms, assumptions on other electromagnetic processes on and near the Earth's surface. Therefore, it is necessary to understand the impact of observational errors on the model forecast, and to estimate appropriately the errors in the Gauss coefficients.

Cross correlation between the poloidal field and the other physical quantities (e.g. toroidal magnetic field, velocity field, density perturbation) in the outer core has been implemented into our MoSST_DAS framework. But it has not been applied to either OSSEs nor geomagnetic data assimilation experiments. We are currently examining the impact of the cross-correlation on the geomagnetic data assimilation, focusing on the speed-up of the assimilation process (or, in other words, reduction of the spin-up time of the system).

Though, there are many problems to be resolved and many questions to be answered, we are beginning to shape up the study of geomagnetic data assimilation. This could enable us not only to forecast geomagnetic secular variation in the coming decades, but also to provide observational constraints to improve numerical geodynamo models.

Acknowledgments

This research is supported by NSF Mathematical Geosciences program under the grants EAR-0327875 and EAR-0327843, NASA Solid Earth and Natural Hazard Program, NASA Mars Fundamental Research Program. We also thank GSFC NPPCS and NASA NAS for computation support.

References

- [1] J. Bloxham and D. Gubbins, The secular variation of the Earth's magnetic field, *Nature*, 317 (1985), 777-781.
- [2] J. Bloxham, D. Gubbins and A. Jackson, Geomagnetic secular variation, *Philos. T. Roy. Soc. A*, 329 (1989), 415-502.
- [3] A. Borovikov, M. M. Rienecker, C. L. Keppenne and G. C. Johnson, Multivariate error covariance estimates by Monte-Carlo simulation for assimilation studies in the Pacific Ocean, *Mon. Weather Rev.*, 133 (2005), 2310-2334.
- [4] S. I. Braginsky, Magnetic waves in the Earth's core, *Geomagn. Aeronomy*, 7 (1967), 851-859.
- [5] S. I. Braginsky and P. H. Roberts, Equations governing convection in Earth's core and the geodynamo, *Geophys. Astro. Fluid Dyn.*, 79 (1995), 1-97.
- [6] C. G. Constable, C. L. Johnson and S. P. Lund, Global magnetic field models for the past 3000 years: transient or permanent flux lobes?, *Philos. T. Roy. Soc. A*, 258 (2000), 991-1008.
- [7] V. Courtillot and J. L. Le Mouél, Geomagnetic secular variation impulses, *Nature*, 311 (1984), 709-716.

- [8] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99 (1994), 10143-10162.
- [9] G. Gaspari and S. E. Cohn, Construction of correlation function in two and three dimensions, *Q. J. Roy. Meteor. Soc.*, 125 (1999), 723-757.
- [10] A. Jackson, A. R. T. Jonkers and M. R. Walker, Four centuries of geomagnetic secular variation from historical records, *Philos. T. Roy. Soc. A*, 358 (2000), 957-990.
- [11] D. Jault, C. Gire and J.-L. LeMouél, Westward drift, core motions and exchanges of angular momentum between core and mantle, *Nature*, 333 (1988), 353-356.
- [12] M. Kono and P. H. Roberts, Recent geodynamo simulations and observations of the geomagnetic field, *Rev. Geophys.*, 40 (2002), 1-53.
- [13] M. Korte and C. G. Constable, The geomagnetic dipole moment over the last 7000 years – new results from a global model, *Earth Planet. Sci. Lett.*, 236 (2005), 348-358.
- [14] W. Kuang and J. Bloxham, An Earth-like numerical dynamo model, *Nature*, 389 (1997), 371-374.
- [15] W. Kuang and J. Bloxham, Numerical modeling of magnetohydrodynamic convection in a rapidly rotating spherical shell: weak and strong field dynamo actions, *J. Comput. Phys.*, 153 (1999), 51-81.
- [16] W. Kuang and B. F. Chao, Geodynamo modeling and core-mantle interaction, in: V. Dehant, K. Creager, S. Karato and S. Zatman (Eds.), *The Core-Mantle Boundary Region*, AGU, Geodynamics Series 31, 2003, pp. 193-212.
- [17] R. A. Langel, The main field, in: J. A. Jacobs (Ed.), *Geomagnetism*, Vol. 1, Academic Press, London, 1987, pp. 249-512.
- [18] R. A. Langel and R. H. Estes, A geomagnetic field spectrum, *Geophys. Res. Lett.*, 9 (1982), 250-253.
- [19] J. Larmor, How could a rotating body such as the Sun become a magnet, *Brit. Assn. Adv. Sci. Rep.*, (1919), 159-160.
- [20] D. Liu, A. Tangborn and W. Kuang, Observing system simulation experiments in geomagnetic data assimilation, *J. Geophys. Res.*, (2007), in press.
- [21] S. Macmillan and J. M. Quinn, The 2000 revision of the joint UK/US geomagnetic field models and an IGRF 2000 candidate model, *Earth Planets Space*, 52 (2000), 1149-1162.
- [22] R. T. Merrill and M. W. McElhinny, *The Earth's Magnetic Field. Its History, Origin and Planetary Perspective*, Academic Press, London, 1983.
- [23] P. Olson and U. R. Christensen, Dipole moment scaling for convection driven planetary dynamos, *Earth Planet. Sci. Lett.*, 250 (2006), 561-571.
- [24] P. H. Roberts, *Geomagnetism*, *Ency. Earth Sys. Sci.*, 2 (1992), 277-292.
- [25] P. H. Roberts and M. Stix, α -effect dynamos, by the Bullard-Gellman formalism, *Astron. Astrophys.*, 18 (1972), 453-466.
- [26] T. J. Sabaka, N. Olson and R. A. Langel, A comprehensive model of the quiet-time near-Earth magnetic field; phase 3, *Geophys. J. Int.*, 151 (2002), 32-68.
- [27] Z. Sun, A. Tangborn and W. Kuang, Data assimilation in a sparsely observed one-dimensional modeled MHD system, *Nonlinear Proc. Geophys.*, 14 (2007), 181-192.
- [28] A. W. P. Thomson, Improving the modelling of the geomagnetic main-field: Isolating the average ionospheric field in satellite data, *Earth Planets Space*, 52 (2000), 1199-1206.
- [29] S. Zhou, W. Kuang, W. Jiang, P. Gary, J. Palencia and G. Gardner, High-speed network and grid computing for high-end computation: Application in geodynamics ensemble simulation, *Concurr. Comp.-Pract. Ex.*, 19(5) (2007), 573-582.