submitted to *Geophys. J. Int.*

A testbed for geomagnetic data assimilation

K. Gwirtz*, M. Morzfeld*, W. Kuang°, A. Tangborn#

* *Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego;*

° *Geodesy and Geophysics Laboratory, NASA Goddard Space Flight Center;*

Joint Center for Earth Systems Technology, University of Maryland Baltimore County.

SUMMARY

Geomagnetic data assimilation merges past and present-day observations of the Earth's magnetic field with numerical geodynamo models and the results are used to initialize forecasts. We present a new "proxy model" that can be used to test, or rapidly prototype, numerical techniques for geomagnetic data assimilation. The basic idea for constructing a proxy is to capture the conceptual difficulties one encounters when assimilating observations into high-resolution, 3D geodynamo simulations, but at a much lower computational cost. The framework of using proxy models as "gate-keepers" for numerical methods that could/should be considered for more extensive testing on operational models has proven useful in numerical weather prediction, where advances in data assimilation and, hence, improved forecast skill, are at least in part enabled by the common use of a wide range of proxy models. We also present a large set of systematic data assimilation experiments with the proxy to reveal the importance of localization and inflation in geomagnetic data assimilation.

Key words: Dynamo: theories and simulations; Numerical modelling; Numerical approximations and analysis; Inverse theory; Statistical methods.

1 INTRODUCTION

Data assimilation (DA) combines observations of a geophysical process with numerical models to improve the model's predictions. Data assimilation has originated in numerical weather prediction (NWP) and has been used there with remarkable success over the past few decades (Bauer et al. 2015). Progress in DA in NWP can be attributed, at least in part, to the availability and wide use of "proxy models." A proxy model represents the numerical difficulties one encounters in the "real" problem, but the computational cost of a simulation with the proxy is much less than that of an actual simulation. The Lorenz problems (Lorenz 1963, 1995), which are simple ordinary differential equations, are prominent examples of proxy models in NWP. Other examples include the Kuramoto-Sivashinsky equation (Kuramoto & Tsuzuki 1975; Sivashinsky 1977), quasi-geostrophic models (see, e.g., Evensen 1994) the primitive equations (see, e.g., Ades & van Leeuwen 2015) the shallow water equations (see, e.g., Holm et al. 2020), and a barotropic vorticity model (see, e.g., Browne 2016). While these models may not accurately describe an atmospheric flow, they have been extremely useful for prototyping and testing new schemes for DA, which has in turn led to more accurate forecasts. Perhaps it is fair to say that new ideas for DA schemes can be tested (quickly) on proxy models, and only those that pass the test should be considered further. Thus, to maximize the usefulness of the proxy, it is important that it captures some of the major challenges of the "real" model and overall DA system.

The main contribution of this paper is to describe a proxy model relevant to geomagnetic DA. In geomagnetic DA, measurements of Earth's magnetic field are assimilated into dynamic models of the magnetic field and related systems. We focus on the scenario where observations are assimilated into self-consistent dynamo models, which couple the magnetic field to fluid motion in the Earth's liquid outer core. Alternatives to this approach, which rely on simpler models have recently begun to be explored (see, e.g., Barrois et al. 2018; Bärenzung et al. 2020; Ropp et al. 2020). While geomagnetic DA is currently a very active field, many questions regarding the details of DA are not fully understood. On a fundamental level, nearly all of geomagnetic DA uses an ensemble method (see, e.g., Sanchez et al. 2019, 2020; Sun & Kuang 2015; Fournier et al. 2013), but it has never been established that this approach is appropriate or more effective than, e.g., a variational method (Li et al. 2011, 2014). Indeed, only recently was the first "hybrid" ensemble-based variational geomagnetic DA system developed (Minami et al. 2020).

On a more nuanced, technical level, ensemble DA is known to be feasible only when *localization* and *inflation* are used (Hamill et al. 2009). Here, feasible means that the ensemble

size is "small:" each ensemble member requires a simulation, so that a large ensemble size accrues a large computational cost. With a small ensemble size, sampling error in ensemble estimates of means and variances are large. In short, localization and inflation are techniques that reduce sampling error arising from a small ensemble size. Localization in particular reduces sampling error by enforcing an assumed correlation structure onto ensemble covariances and is well-understood and widely used in NWP and oceanography, where spatially decaying correlations are prevalent. Correlation structures in geomagnetic DA, however, are more complicated and many fundamental questions regarding effective localization and inflation are indeed unanswered (see Section 3 for more details). We believe that a proxy model can be useful for finding partial answers to these important questions. Surprisingly, some of the first papers on geomagnetic DA in fact used proxy models (Sun et al. 2007; Fournier et al. 2007), but this train of thought was discontinued, perhaps because this "first generation" of proxy models were overly simplified. Finally, the proxy we derive has pedagogical value and can be used as a (computationally) simple tool to teach geomagnetic DA to the next generation of students.

The rest of this paper is organized as follows. In Section 2, we first present a set of characteristics that define a useful proxy model for geomagnetic DA and then construct a proxy by coupling a chaotic flow to an induction equation. For computational reasons, we consider the model only in two-dimensional geometries (square and sphere *). We then provide background on data assimilation, localization and inflation in Section 3, before we showcase how to use the proxy model to study the effects of localization techniques in geomagnetic DA in Section 4. We finish the paper with a summary of our conclusions.

2 A PROXY MODEL FOR GEOMAGNETIC DA

2.1 Background on geodynamo models and geomagnetic DA

The geodynamo is the process in which the Earth's intrinsic magnetic field is generated and maintained by convective fluid motion in the Earth's liquid outer core. In particular, the system exhibits *dynamo action* if the fluid flow maintains the magnetic field in the absence of an external magnetic source. The fluid motion, or core convection, is driven by strong thermal and compositional buoyancy arising from the secular cooling and differentiation of

* Throughout, we use the mathematical definition of a sphere, which refers a two-dimensional geometrical object, embedded in a three-dimensional space, i.e., the sphere is defined by a set of points that all have equal distance from a specified point in three-dimensional space, which we take to be the origin.

the Earth during its entire evolutionary history. Geomagnetic observations and simulations have demonstrated that the geomagnetic field varies on vastly different time scales from several years, to millions of years. These variations, called the secular variation (SV), occur to the overall field intensity and its detailed morphology. Moreover, we know that the strength of the geomagnetic field is such that the Lorentz force is a major contributor to non-geostrophic core flow. The resulting fluid motion in Earth's outer core is turbulent and strongly coupled to the geomagnetic field. The latter two characteristics of the geodynamo – coupling between velocity and magnetic fields and chaotic behavior – must be properly represented in any useful proxy model.

A typical, magnetohydrodynamic (MHD), three-dimensional, self-consistent geodynamo *model* features the Navier-Stokes equation to describe the momentum balance of the electrically conducting fluid, a magnetic induction equation, and a thermodynamic equation which describes the energy source for driving the core convection. Using the Boussinesq approximation in a reference frame rotating at angular velocity $\mathbf{\Omega}$, the model is given by the set of partial differential equations (PDE)

$$\rho_0\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla\right)\mathbf{v}+2\rho_0\mathbf{\Omega}\times\mathbf{v} = -\rho_0\nabla p+\rho\mathbf{g}+\rho_0\nu\nabla^2\mathbf{v}+\mathbf{J}\times\mathbf{B}, \quad (1)$$

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

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla\right)\rho = \kappa\nabla^2\rho, \quad (3)$$

where ρ is the fluid density with mean ρ_0, p is the pressure, ν is the kinematic viscosity, η is the magnetic diffusivity, $\mathbf{v}$ is the fluid velocity in the outer core, $\mathbf{B}$ is the magnetic field, $\mathbf{g}$ is gravity, κ is the thermal diffusivity, and $\mathbf{J}=(\nabla\times\mathbf{B})/\mu_0$ is the current density (Kuang & Bloxham 1999).

In geodynamo models, the magnetic field is typically decomposed into poloidal and toroidal components

$$\mathbf{B}(\mathbf{r})=\mathbf{B}_T(\mathbf{r})+\mathbf{B}_P(\mathbf{r})=\nabla\times T(\mathbf{r})\hat{\mathbf{r}}+\nabla\times(\nabla\times P(\mathbf{r})\hat{\mathbf{r}}). \quad (4)$$

The scalar fields $T(\mathbf{r})$ and $P(\mathbf{r})$ can be further decomposed into the coefficients $T_\ell^m(r)$ and $P_\ell^m(r)$ for spherical harmonic functions of degree ℓ and order m. The divergence free velocity field of the outer core is similarly described in spherical harmonics along with the scalar field representing the temperature perturbation. Upon discretization via spherical harmonics, the total model state can then be represented as a vector of spherical harmonic coefficients defining the magnetic field, velocity field and temperature perturbation at each radial level on the grid space of the numerical model.

In geomagnetic DA, geodynamo models are informed by observations of the poloidal magnetic field at, or near, the core-mantle boundary (CMB) (see, e.g., Sanchez et al. 2019, 2020; Tangborn & Kuang 2018; Fournier et al. 2013). The observations come from geomagnetic field models (see, e.g., Sabaka et al. 2020) in the form of spherical harmonic coefficients. Assimilating spherical harmonics derived from field models, rather than the "raw" data, is useful for isolating the field of the geodynamo from sources external to the outer core. Indeed, it is generally agreed that large scale features, up to degree $\ell = 13$, are primarily generated in the outer core (Langel & Estes 1982). Modern, satellite-based measurements allow for regular observations that resolve the poloidal field near the outer core to this level. Geomagnetic field models covering earlier periods, provide a lower spatial and temporal resolution. For example, DA using archaeomagnetic-based representations of the field may only assimilate coefficients up to degree $\ell = 3$ (see, e.g., Sanchez et al. 2016).

2.2 A wishlist of properties for a useful proxy model

Based on the dynamical properties of the core state, the mathematical/numerical implementation of geodynamo simulations and of geomagnetic DA, we synthesize the following list of characteristics that a useful proxy model should exhibit:

(i) The proxy should consist of two coupled fields (velocity and magnetic fields);
(ii) the proxy should be amenable to a spectral discretization (spherical harmonics);
(iii) observations should be spherical harmonic coefficients of the (proxy) magnetic field;
(iv) the proxy should exhibit chaotic behavior.

Regarding the last point, we note that chaotic behavior of geodynamo models is implied by the Navier-Stokes equations and that chaos is strong in the geodynamo because the predictability of the Earth's magnetic field is known to be at most a century (Hulot et al. 2010).

Finally, we emphasize that a proxy model should be computationally less demanding than a geodynamo model, or else there is nothing to be gained by using the proxy. Given the complexity of a geodynamo model, which involves several millions of unknown state variables, a single simulation is already computationally demanding. Ensemble DA, which requires multiple simulations of the model per assimilation cycle, thus easily leads to peta-flop (floating-point operations) computations. This is perhaps not a critical issue for performing *one* data assimilation in itself, but systematic studies of DA require *repeated* DA experiments. For example, there are several competing DA techniques in the (geophysical) literature, including ensemble DA (see, e.g., Evensen 2006) and variational methods (see, e.g., Courtier 1997). Computational limitations make it (nearly) impossible to test and compare both classes of DA methods

in geomagnetic DA, but it is known that the type of DA algorithm used has a large impact on forecast skill.

On a more technical level, localization and inflation are critical for any useful ensemble DA system (Hamill et al. 2009). In short, localization and inflation are techniques that reduce sampling error, that arise from a small ensemble size (hundreds of ensemble members), but a small ensemble size is critical to keep the computational budget reasonable. Recently, the ideas of localization have appeared in geomagnetic DA (Sanchez et al. 2019, 2020). Yet, a systematic study of *how* these methods should be used in geomagnetic DA is missing because such studies require repeated and systematic DA experiments which cause the computational budget to explode. A proxy model that runs easily on a simple laptop computer, can be useful to answer these fundamental questions because, with access to high-performance computing (HPC), repeated and systematic DA experiments are easily within reach. In this context, we remind the reader again that some of the rapid progress in the application of DA in NWP can be attributed to the wide availability (and use) of proxy models such as the Lorenz models (Lorenz 1963, 1995), or quasi geostrophic models (see, e.g., Evensen 1994).

2.3 Mathematical derivation of the proxy model

We construct a proxy model that consists of a chaotic velocity field, coupled to a magnetic field via an induction equation. To keep computational requirements reasonable, we make the following two choices from the very beginning:

(i) The proxy is defined on a two-dimensional domain (square or sphere);
(ii) we substitute the Kuramoto-Sivashinsky (KS) equation for the Navier-Stokes equation.

We begin our description of the proxy within a generic 2D geometry and discuss the implications of this choice: self-sustained dynamo action is impossible in 2D so that the proxy requires an external magnetic source. We then describe the modified KS equation and the coupling of the resulting velocity and magnetic fields, and finally specify the proxy to square and spherical domains.

2.3.1 The proxy magnetic field

The magnetic field of the Earth is sustained because the magnetic induction arising from the convective flow compensates for the Ohmic dissipation resulting from the finite electrical conductivity of the core fluid (dynamo action). The 2D geometry we impose on the proxy model implies that the proxy cannot exhibit self sustained dynamo action (Cowling 1933). For this reason, the proxy is a 2D *magnetoconvection* system with a nontrivial, but steady,

"background" field that sustains the magnetic field. Specifically, we write the total magnetic field as

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{b} \tag{5}$$

where $\mathbf{B}_0$ is the steady background field ($\partial \mathbf{B}_0/\partial t = \mathbf{0}$) and $\mathbf{b}$ is a two-dimensional, time-varying perturbation. The background field $\mathbf{B}_0$ will be chosen such that the perturbation field does not decay (see below). The perturbation magnetic field $\mathbf{b}$ is divergence free within the 2D domain and, thus, is defined by a magnetic vector potential that is everywhere normal to the 2D surface. Thus,

$$\mathbf{b} = \nabla \times A\hat{\mathbf{n}}, \tag{6}$$

where A is a scalar field and $\hat{\mathbf{n}}$ is the unit vector normal to the domain. Taking the normal component of the curl of equation (2) and using equations (5) and (6), along with the fact that $\partial \mathbf{B}_0/\partial t = \mathbf{0}$ gives

$$\left[\nabla \times (\nabla \times \frac{\partial A}{\partial t}\hat{\mathbf{n}})\right]_\perp = \left[\nabla \times (\nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B})\right]_\perp, \tag{7}$$

where the subscript $\perp$ indicates the normal component of the bracketed vector quantity and where $\eta > 0$ is a constant.

2.3.2 The proxy velocity field and the modified Kuramoto-Sivashinsky equation

In a geodynamo model, the turbulent flow field of the outer core is described by the Navier-Stokes equation (1). We require that the velocity field of the proxy model be similarly chaotic, but need to avoid the computational difficulties that result from the Navier-Stokes equation. For this reason, we "replace" the Navier-Stokes equation by the Kuramoto-Sivashinsky (KS) equation

$$\frac{\partial \omega}{\partial t} = -\alpha \nabla^2 \omega - \beta \nabla^4 \omega - \gamma |\nabla \omega|^2, \tag{8}$$

where α, β and γ are positive constants (with appropriate units). The KS equation has its origins in the study of chaotic physical processes (Kuramoto & Tsuzuki 1975; Sivashinsky 1977) and has since been used as a proxy model for exploring chaotic behavior, particularly in fluid dynamics, (see, e.g., Hooper & Grimshaw (1985); Papageorgiou et al. (1990)) and data assimilation (see, e.g., Jardak et al. 2010; Chorin et al. 2013; Morzfeld et al. 2018). It can be shown that the mean value of solutions to the KS equation in this form increase (in magnitude) without bound but exert no influence on the evolution of perturbations about the mean. For this reason it is common to remove the mean so that solutions are centered at zero at all times (for further details see, e.g., Kalogirou et al. 2015; Armbruster et al. 1989).

We interpret the solution of the KS equation as the normal component of vorticity:

$$\omega = [\nabla \times \mathbf{v}]_\perp. \tag{9}$$

Defining the velocity $\mathbf{v}$ by a streamfunction ψ

$$\mathbf{v} = \nabla \times \psi \hat{\mathbf{n}}, \tag{10}$$

gives

$$\omega = [\nabla \times (\nabla \times \psi \hat{\mathbf{n}})]_\perp. \tag{11}$$

Thus, the vorticity ω, governed by the KS equation, defines the velocity field.

Finally, we need to couple the velocity to the magnetic field. In a geodynamo, this coupling occurs via the Lorentz force, which influences the fluid flow by acting on induced currents. The Lorentz force coupling is expressed through the term $\mathbf{J} \times \mathbf{B}$ in the Navier-Stokes equation (1) and we construct the coupling term of the proxy model in a similar way. Recall that the solution to the KS equation is related to the velocity through equation (9). This motivates the additional term $\lambda[\nabla \times (\mathbf{J} \times \mathbf{B})]_\perp$, where λ is a positive constant and $\mathbf{J} = (\nabla \times \mathbf{B})/\mu_0$ as before. The proxy then consists of the induction equation coupled to a modified KS equation,

$$\frac{\partial \omega}{\partial t} = -\alpha \nabla^2 \omega - \beta \nabla^4 \omega - \gamma |\nabla \omega|^2 + \lambda \left[\nabla \times (\mathbf{J} \times \mathbf{B})\right]_\perp \tag{12}$$

$$\left[\nabla \times (\nabla \times \frac{\partial A}{\partial t} \hat{\mathbf{n}})\right]_\perp = \left[\nabla \times (\nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B})\right]_\perp \tag{13}$$

To derive a non-dimensional form of the equations, we scale time by the diffusion timescale associated with the fourth-order term of the KS equation, i.e., $\bar{t} = (\beta/L^4)t$, we scale the magnetic field by a characteristic magnetic field intensity B, and distances by a typical length scale L. The dimensionless equations of (12) then become

$$\frac{\partial \bar{\omega}}{\partial \bar{t}} = -R_a \bar{\nabla}^2 \bar{\omega} - \bar{\nabla}^4 \bar{\omega} - \Gamma |\bar{\nabla} \bar{\omega}|^2 + \Lambda [\bar{\nabla} \times ((\bar{\nabla} \times \bar{\mathbf{B}}) \times \bar{\mathbf{B}})]_\perp \tag{14}$$

$$\left[\bar{\nabla} \times (\bar{\nabla} \times \frac{\partial \bar{A}}{\partial \bar{t}} \hat{\mathbf{n}})\right]_\perp = [\bar{\nabla} \times (\bar{\nabla} \times (\bar{\mathbf{v}} \times \bar{\mathbf{B}}) + \frac{1}{R_m} \bar{\nabla}^2 \bar{\mathbf{B}})]_\perp, \tag{15}$$

where $R_a = \alpha L^2/\beta$, $\Gamma = \gamma/L^2$, $\Lambda = \lambda L^2 B^2/\mu_0 \beta^2$ and $R_m = \beta/\eta L^2$; and horizontal bars indicate dimensionless quantities and operators, e.g., $\bar{\mathbf{B}} = \mathbf{B}/B$.

After non-dimensionalization, the proxy model is characterized by four dimensionless parameters (R_a, Γ, Λ and R_m) which have physical interpretations. In this context, we first recall that the diffusive term in the KS equation is of fourth order, while the second order term acts as a forcing (due to the negative sign). For this reason, R_a can be viewed as the ratio of the diffusive and the forcing/convective time scales, which is why we call this quantity the *modified Rayleigh number* of the proxy model. Moreover, R_a controls the length scales in

the vorticity $\bar{\omega}$: the larger R_a is, the smaller are the features present in the solutions of $\bar{\omega}$ (and thus the velocity field $\mathbf{v}$ also exhibits smaller features). The nondimensional parameters Γ and Λ control the relative influence of the original nonlinear term of the KS equation and the Lorentz force, respectively. In the induction equation, the *magnetic Reynolds number* (R_m) is the ratio between the characteristic velocity (β/L^3), multiplied by the length scale, and the magnetic diffusivity. The value of R_m thus determines the influence of induction of the magnetic field relative to magnetic diffusion. For the remainder of this paper we work with the nondimensional proxy and for convenience drop the bar notation for dimensionless quantities.

2.3.3 Proxy model on the square

We now specify the generic proxy to a square domain with periodic boundary conditions. In this geometry, the normal vector is the z-direction ($\hat{\mathbf{n}} = \hat{\mathbf{z}}$) and the perturbation magnetic field $\mathbf{b}$ and the velocity field $\mathbf{v}$ have components only in the xy-plane. We chose the steady background field $\mathbf{B}_0 = B\hat{\mathbf{x}}$ to avoid uninteresting solutions that decay to zero. In general, any background field for which $\mathbf{B}_0 \neq \nabla \times A_0\hat{\mathbf{n}}$ for periodic A_0 prevents the decay (this can be derived by revisiting the definition of $\mathbf{b}$ by a vector potential in (6)). With $\mathbf{B}_0 = B\hat{\mathbf{x}}$, the proxy on a square geometry is described by the following set of PDEs:

$$\frac{\partial \omega}{\partial t} = -R_a\nabla^2\omega - \nabla^4\omega - \Gamma|\nabla\omega|^2 + \Lambda[\nabla \times ((\nabla \times \mathbf{B}) \times \mathbf{B})]_z, \tag{16}$$

$$\omega = -\nabla^2\psi, \tag{17}$$

$$v_x = \frac{\partial \psi}{\partial y}, \quad v_y = -\frac{\partial \psi}{\partial x}, \tag{18}$$

$$\frac{\partial A}{\partial t} = -(\mathbf{v}\cdot\nabla)A + \frac{1}{R_m}\nabla^2 A - v_y, \tag{19}$$

$$b_x = \frac{\partial A}{\partial y}, \quad b_y = -\frac{\partial A}{\partial x}, \tag{20}$$

where v_x and v_y are the components of $\mathbf{v}$, b_x and b_y the components of $\mathbf{b}$, and $\mathbf{B} = \mathbf{b} + \hat{\mathbf{x}}$ is the total magnetic field. Here, the del operator is $\nabla = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y}$ (recall that derivatives normal to the square vanish), and the Laplacian is $\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$.

2.3.4 Proxy model on the sphere

If the domain of the generic proxy model is a sphere, the vector normal to the domain is in the radial direction ($\hat{\mathbf{n}} = \hat{\mathbf{r}}$) and the magnetic and velocity fields $\mathbf{b}$ and $\mathbf{v}$ have components in the polar and azimuthal directions $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$. The background field should be chosen to prevent uninteresting solutions that decay to zero. In general, any $\mathbf{B}_0 \neq \nabla \times A_0\hat{\mathbf{n}}$ for which A_0 is smooth and continuous over the sphere can be used, and we pick a dipole field $\mathbf{B}_0 =$

$B[2\cos\theta\hat{\mathbf{r}} + \sin\theta\hat{\boldsymbol{\theta}}]$, where B is the characteristic magnetic field intensity. With a dipole background field, the proxy model on the sphere is defined by the set of PDEs

$$\frac{\partial\omega}{\partial t} = -R_a\nabla^2\omega - \nabla^4\omega - \Gamma|\nabla\omega|^2 + \Lambda[\nabla\times((\nabla\times\mathbf{B})\times\mathbf{B})]]_r, \tag{21}$$

$$\omega = -\nabla^2\psi, \tag{22}$$

$$v_\theta = \frac{1}{\sin\theta}\frac{\partial\psi}{\partial\varphi}, \quad v_\varphi = -\frac{\partial\psi}{\partial\theta}, \tag{23}$$

$$\begin{aligned}-\nabla^2\frac{\partial A}{\partial t} &= -\nabla^2[-(\mathbf{v}\cdot\nabla)A - v_\varphi\sin\theta + \frac{1}{R_m}(\nabla^2A + A)] \\ &\quad + \frac{2}{\sin\theta}\left[\frac{\partial}{\partial\theta}(v_\varphi\sin\theta\cos\theta) - \frac{\partial}{\partial\varphi}(v_\theta\cos\theta)\right],\end{aligned} \tag{24}$$

$$b_\theta = \frac{1}{\sin\theta}\frac{\partial A}{\partial\varphi}, \quad b_\varphi = -\frac{\partial A}{\partial\theta}, \tag{25}$$

where v_θ and v_φ are the components of $\mathbf{v}$, b_θ and b_φ the components of $\mathbf{b}$, and $\mathbf{B} = \mathbf{b} + 2\cos\theta\hat{\mathbf{r}} + \sin\theta\hat{\boldsymbol{\theta}}$. The Laplacian operator on the unit sphere takes the form $\nabla^2 f = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial f}{\partial\theta}) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\varphi^2}$ (and derivatives in the radial direction vanish). The above are the nondimensional form of the proxy model and the radius is used as the characteristic length scale L during nondimensionalization.

2.4 Properties of the proxy

Recall the list of properties a useful proxy model should have (see Section 2): coupled fields, spectral discretization, spectral observations, and chaotic behavior. By construction, the proxy is composed of two interacting fields, the square and spherical geometries allow for spectral discretization (via fast Fourier transform or spherical harmonic transform) and spectral observations, and it can be anticipated that the proxy is chaotic because the KS equation is well understood to produce chaotic solutions. The coupling of the fields and the chaotic behavior, however, warrant more investigation: if the nonlinearity/chaos is relatively mild, and/or if the coupling between magnetic and velocity fields is weak, then the proxy is not relevant to geodynamo models.

It may be possible to study these questions analytically, but here we discuss these properties in the context of concrete proxies (on the square and sphere) with non-dimensional parameters listed in Table 1. These parameters were determined in an iterative process (trying a few options and evaluating the results), and with a limited computational budget in mind. There are good reasons, however, for why the modified Rayleigh number should be large. A linearized analysis of the KS equation (no coupling with a magnetic field) shows that all modes decay if $R_a < 4\pi^2$, on the unit square domain, and if $R_a < 2$ on the unit sphere.

Table 1. Parameter values for the proxy model on a square and sphere.

	R_a	R_m	Γ	Λ
Square	10^3	10^{-3}	10^{-3}	4×10^7
Sphere	10^2	10^{-2}	10^{-2}	2×10^4

One can thus anticipate that a large modified Rayleigh number will lead to chaotic dynamics (which we need), and we study the chaotic behavior of the proxy in more detail below.

2.4.1 Discretization, spectra and timescales

For computations, we discretize the proxy by a pseudo-spectral method (see, e.g., Fornberg 1996), using Fourier series (up to wave number 15) on the square, and spherical harmonics to degree and order 25 on the sphere. Time integration is implemented with exponential time differencing (see, e.g., Cox & Matthews 2002). The maximum time steps are $\Delta t = 5 \times 10^{-9}$ (square) and $\Delta t = 5 \times 10^{-7}$ (sphere). We also implement dealiasing. On the square, this is done by treating non-linear terms on a 48×48 grid in physical space and truncating solutions at directional wave-number 15. On the spherical surface, dealiasing is implemented by treating non-linear terms on a grid of $N_\theta = 39$ polar angles and $N_\varphi = 77$ azimuthal angles in physical space and truncating solutions at spherical harmonic degree and order 25. We emphasize that the proxy on the square is faster to compute with than the proxy in the sphere because the fast Fourier transform (FFT) is faster than the spherical harmonic transform.

Figure 1 shows a snapshot of the (numerical) solution of the proxy. The plots in the left column illustrate the proxy on the square, and the right column, the proxy on the sphere. The coloring of the top row shows the scalar A which defines the vector potential for the perturbation magnetic field, $\mathbf{b}$, depicted by the overlaid vector field. The coloring of the bottom row shows the scalar field ω, which is related to the vorticity of the overlaid velocity field. In both configurations, we note features occurring over a range of different scales, especially in the vorticity field.

We further study the power spectral density (PSD) of the proxy by generating a long simulation, and then averaging the energy in each mode (Fourier on the square, spherical harmonic on the sphere). Figure 2 shows the log base ten of the PSDs, scaled by the largest energy in the spectrum. We also highlight the modes that, on average, carry around 95% of the energy and note that energy is transferred between several modes, in particular in the vorticity. Nonetheless, the number of modes that carry significant energy is much less than in

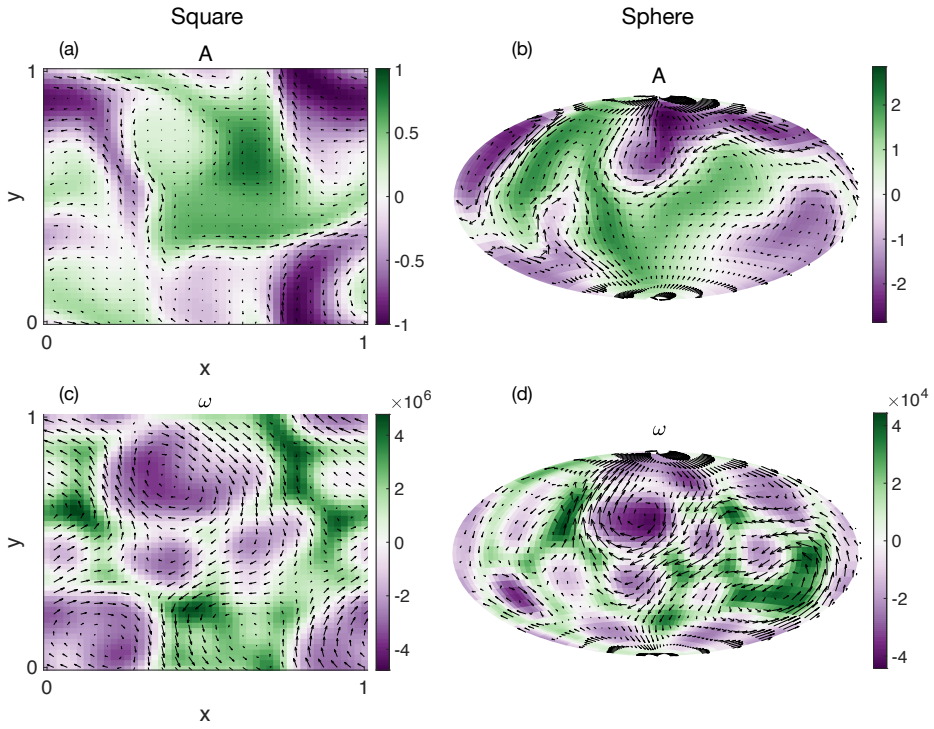

Figure 1. Snapshot of a solution of the proxy on the square and sphere *Top row:* Scalar defining the vector potential (color) for the perturbation magnetic field (vectors) on the (a) square and (b) sphere. *Bottom row:* Normal component of vorticity (color) and associated velocity (vectors) on the (c) square and (d) sphere.

the geodynamo or 3D geodynamo models. This is to be expected, however, because the proxy is designed to be computationally cheaper than the "real" model.

Finally, we compute typical timescales of the modes describing the fields of A and ω. For a long run on the square domain, we record for both fields, the average values of

$$\tau_{|k_x|,|k_y|} = \sqrt{\frac{\sum_{m=|k_x|}\sum_{n=|k_y|} c_{m,n}^2}{\sum_{m=|k_x|}\sum_{n=|k_y|} \dot{c}_{m,n}^2}} \tag{26}$$

where $c_{m,n}$ is the Fourier coefficient for wave numbers m and n in the x and y directions, respectively, and $\dot{c}_{m,n}$ is the coefficient's derivative with respect to time. Similarly, for a long run on the spherical surface domain, we record for both A and ω, the average values of

$$\tau_{\ell,m} = \sqrt{\frac{\sum_{|n|=m}(c_\ell^n)^2}{\sum_{|n|=m}(\dot{c}_\ell^n)^2}} \tag{27}$$

where c_ℓ^n is the spherical harmonic coefficient for degree ℓ and order n and $\dot{c}_\ell^n$ is the coefficient's derivative with respect to time. The log base ten of the resulting values are shown in Figure 3. We see that in both settings, the timescales vary by around an order of magnitude with the

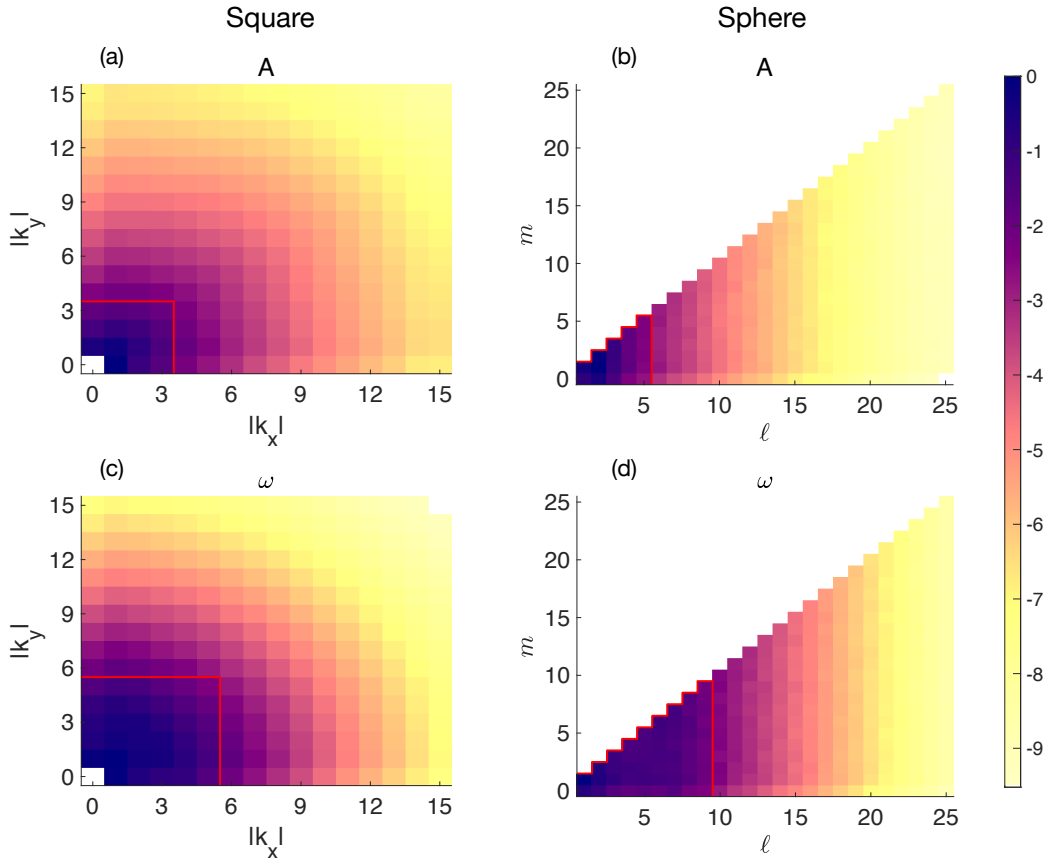

Figure 2. The logarithm base ten of the PSDs (scaled by the highest energy in the spectrum) of (a) A on the square, (b) A on the sphere, (c) ω on the square, (d) ω on the sphere. Red lines highlight the parts of the spectrums that carry approximately 95% of the energy, on average.

longer timescales occurring for large wavelengths. The shortest timescales on the square are a little less than 10^{-6} dimensionless time units, while the shortest timescales on the spherical surface are a little less than 10^{-4} dimensionless time units. In both settings, these timescales are over two orders of magnitude greater than the maximum allowed timesteps of $\Delta t = 5\times10^{-9}$ (square) and $\Delta t = 5 \times 10^{-7}$ (sphere) indicating that the time steps we chose are small enough to resolve all relevant time scales.

2.4.2 Chaotic behavior and e-folding time

We examine the sensitivity of the proxy model to initial conditions and compute the e-folding time to confirm that the proxy is indeed chaotic. To do this, we choose a set of initial conditions and generate a long run, recording the solutions at regular intervals. The same initial conditions, with a small perturbation of A, are used to generate another set of runs. In the square, the perturbation of A results from multiplying Fourier modes with directional wave

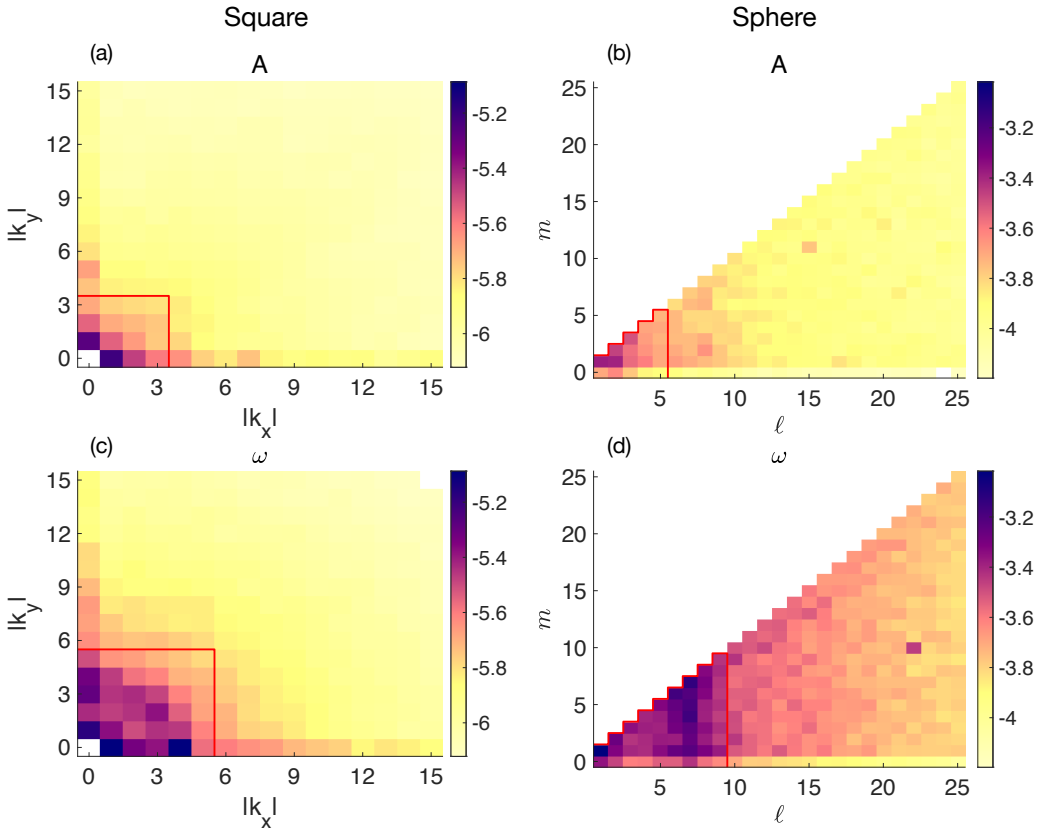

Figure 3. Log base ten of the typical timescales by mode of (a) A on the square, (b) A on the sphere, (c) ω on the square, (d) ω on the sphere. Red lines highlight the parts of the spectrums that carry approximately 95% of the energy, on average.

numbers less than or equal to three ($|k_x|, |k_y| \leq 3$) by factors drawn from $\mathcal{N}(1, 10^{-3})$. Similarly, the initial condition of the proxy on a sphere is perturbed by multiplying spherical harmonics of A, of degree less than or equal to five ($\ell \leq 5$) by factors drawn from $\mathcal{N}(1, 10^{-3})$. For both A and ω the error is computed as the L^2 norm of the difference in the solutions, divided by the L^2 norm of the solution using unperturbed initial conditions. That is, if A and ω represent the solution at a given time using the original initial conditions and A' and ω', a solution resulting from perturbed initial conditions, we compute the errors in A and ω to be

$$\varepsilon_A = \frac{||A - A'||_2}{||A||_2}, \quad \varepsilon_\omega = \frac{||\omega - \omega'||_2}{||\omega||_2}. \tag{28}$$

We repeat this process for ten independent perturbations. Figure 4 shows the average error as a function of time; the dashed lines illustrate the log-linear fit to the average error during the period of exponential error growth. Since error growth is exponential, we conclude that the proxy is chaotic. The e-folding times—the time need for an error to grow by a factor of e—are around 6×10^{-6} (square) and 5×10^{-4} (sphere).

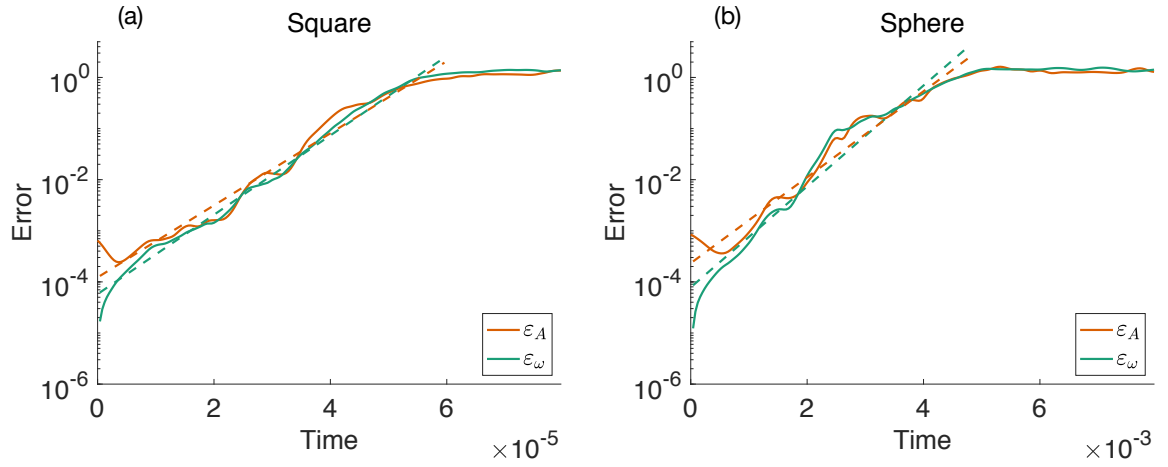

Figure 4. Average error vs. time for A (orange) and ω (green) on the (a) square and (b) sphere. Dashed lines show the log-linear fit to the average errors during periods of exponential growth. The slope of the dashed lines are inversely proportional to the e-folding times of the proxies on the square and sphere.

2.4.3 Coupling of magnetic field and velocity field

The equations of the proxy model are coupled, in the sense that the magnetic field $\mathbf{B}$ appears in the equation for ω (which determines the velocity field), and, similarly, the velocity $\mathbf{v}$ appears in the induction equation which governs the evolution of the magnetic field. For a useful proxy, however, this coupling needs to be strong in the sense that the coupling terms exert a non-negligible influence on the dynamics relative to other components of the proxy. If, on the other hand, the proxy sits in a regime were the Lorentz force largely dictates the behavior of the velocity field, we risk making the challenge of estimating the velocity from magnetic field observations during data assimilation, too easy. Specifically, even if correlations between the velocity and magnetic fields are poorly understood, accurate estimates of the magnetic field could, over time, force estimates of the velocity towards its true state if the coupling is too strong.

We first consider the influence of the Lorentz force term on the velocity field. For that purpose, we compute the ratios of the L^2 norms of the forcing term $(-R_a\nabla^2\omega)$ and the Lorentz force term $(\Lambda[\bar{\nabla}\times((\bar{\nabla}\times\bar{\mathbf{B}})\times\bar{\mathbf{B}})]_\perp)$, over the course of a long run. Table 2 shows the average and standard deviation of this ratio. The L^2 norm of the Lorentz force is typically around 25% of that of the linear forcing of the KS equation in the square geometry and around 11% on the spherical surface. Second, we consider the relative influence of the velocity field on the magnetic field through induction. Similar to the comparison of forces in the modified KS

Table 2. The ratios of the L^2 norms of the Lorentz force and linear forcing, and the induction and magnetic diffusion terms. Standard deviations in the ratios are show in brackets.

	$\frac{\|\|\text{Lorentz force}\|\|_2}{\|\|R_a\nabla^2\omega\|\|_2}$	$\frac{\|\|\text{Induction}\|\|_2}{\|\|\text{Magnetic diffusion}\|\|_2}$
Square	0.25 (0.20)	2.13 (0.45)
Sphere	0.11 (0.06)	3.54 (1.22)

equation, we compute the ratio of the L^2 norms of the induction terms and diffusion terms in the evolution equation of A. The mean and standard deviation of this ratio is listed in Table 2 We find that the norm of the induction terms is typically around 2-3 times that of the magnetic diffusion, implying that the velocity field exerts meaningful influence on the magnetic field.

2.5 Summary of the proxy model

With geomagnetic DA in mind, we constructed a proxy model that couples a chaotic flow to an induction equation. The model exhibits chaotic behavior and meaningful coupling between the proxy magnetic and velocity fields, is amenable to spectral discretization (spherical harmonics), and runs easily on a 2021 laptop. The proxy has no natural mechanism for dipole reversals and, for that reason, should be used in the context of geomagnetic DA systems for decadal forecasts.

3 USING THE PROXY TO STUDY GEOMAGNETIC ENSEMBLE DA

By construction, the proxy is not as complex as a 3D dynamo model, but is designed to represent the major challenges in geomagnetic DA. In this section, we describe the required background on DA to use the proxy for informative numerical experiments (with results in Section 4). We focus on ensemble DA, because this approach is most common thus far. Within ensemble DA, localization and inflation are required numerical "tricks" that have recently been shown to have a great impact in geomagnetic DA (Sanchez et al. 2019, 2020). Localization is indeed the focus of the rest of this paper, because it is a necessary – but understudied – requirement of computationally efficient ensemble DA (in any application of DA, including geomagnetic DA). After providing some background, we present a variety of localization schemes that are reasonable in the context of geomagnetic DA. We then put these schemes to the test in numerical experiments (Section 4). Being able to perform comprehensive and controlled numerical experiments in this way was a major motivation for us to create the proxy because,

besides the proxy's deficiencies induced by simplifications, a detailed comparative study is out of reach with (realistic) 3D geodynamo models. Nonetheless, we emphasize that our results are qualitative and should be interpreted with the simplifications of the proxy mind.

Finally, we understand that it is possible to use the proxy and similar numerical experiments to study if a variational approach can be more effective than ensemble DA in geomagnetic applications. We leave these ideas for the future, in part because a variational technique requires a tangent-linear and adjoint model (for efficient gradient computations), which is not currently available for any of the existing geomagnetic DA codes.

The rest of this section presents a rapid review of DA to set up notation and to introduce and explain the importance of localization and inflation. We then introduce five localization schemes that are intuitive in the context of geomagnetic DA (and in the absence of a spatial decay of correlations).

3.1 Review of data assimilation and the ensemble Kalman filter

DA is typically formulated within a Bayesian framework as follows. Let $\mathbf{x}$ be the vector representing the state of the system (spherical harmonic coefficients of the proxy or geodynamo), and collect all observations into a vector $\mathbf{y}$ (spherical harmonics of field models). Throughout, we use N with a subscript to denote dimensions, e.g., N_x is the dimension of the vector $\mathbf{x}$. Since the number of observations is typically much smaller than the number of state variables, we have $N_y \ll N_x$. Because the observations and model state are spherical harmonic coefficients, one can construct an $N_y \times N_x$ matrix $\mathbf{H}$, consisting of rows of the identity matrix, such that $\mathbf{Hx}$ are the observed coefficients in $\mathbf{y}$. The state $\mathbf{x}$ and observations $\mathbf{y}$ are thus related by a linear equation

$$\mathbf{y} = \mathbf{Hx} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{29}$$

where $\boldsymbol{\varepsilon}$ is an N_y-dimensional random vector (Gaussian with mean zero and covariance $\mathbf{R}$) that models uncertainties in the model and observations. Equation (29) defines the likelihood $p_l(\mathbf{y}|\mathbf{x})$ which describes the probability distribution of the observations given a system state. In a sequential DA system, numerical simulations and previous assimilation results are used to define a prior distribution $p_0(\mathbf{x})$ (see below for details). By Bayes' rule, the likelihood and prior define the posterior distribution

$$p(\mathbf{x}|\mathbf{y}) \propto p_0(\mathbf{x}) p_l(\mathbf{y}|\mathbf{x}). \tag{30}$$

The ultimate objective of DA is the approximation of this posterior distribution. The various numerical methods used in DA differ in how approximations to the posterior distribution are

found. Here, one distinguishes broadly between variational methods and ensemble methods. In variational methods, one finds the state of highest posterior probability via optimization (see, e.g., Courtier 1997). Ensemble methods, in particular the ensemble Kalman filter (EnKF) and its many variants, rely on Monte Carlo (see, e.g., Evensen 2006). Recently, hybrid techniques have been created to combine varational ideas with Monte Carlo and the ensemble approach (EDA; see, e.g., Bonavita et al. 2012; Zhang & Zhang 2012), but we do not pursue these ideas here further.

For the rest of this paper, we consider and study the EnKF, as an example of an ensemble DA technique which is the most widely used approach in geomagnetic DA. The EnKF approximates the posterior distribution by combining a Monte Carlo approach with the Kalman filter. Specifically, a *forecast ensemble* of N_e unique forecasts $X^f = \{\mathbf{x}_1^f, ..., \mathbf{x}_{N_e}^f\}$ represents a set of samples of the prior distribution $p_0(\mathbf{x})$. The aim of the EnKF is to adjust these forecasts by merging them with information contained in the observations $\mathbf{y}$. This collection of "adjusted" forecasts forms an *analysis ensemble* which is approximately distributed according to the posterior distribution. Typically, the mean of the analysis ensemble is used as an estimate of the true state of the system, with the ensemble variance indicating the estimate's uncertainty.

One assimilation cycle of the EnKF can be implemented as follows (for other implementations, see, e.g., Tippett et al. (2003); Hunt et al. (2007); Buehner et al. (2017)). The forecast ensemble is generated by running repeatedly the (proxy) model with varying initial conditions (usually starting from the analysis ensemble of the previous assimilation cycle). The forecast ensemble defines the *forecast covariance*

$$\mathbf{P}^f = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\mathbf{x}_i^f - \bar{\mathbf{x}})(\mathbf{x}_i^f - \bar{\mathbf{x}})^T, \tag{31}$$

where $\bar{\mathbf{x}} = (1/N_e)\Sigma_{i=1}^{N_e}\mathbf{x}_i^f$ is the ensemble mean (or forecast mean). The *analysis ensemble* is obtained by

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}(\mathbf{y} + \boldsymbol{\varepsilon}_i - \mathbf{H}\mathbf{x}_i^f), \tag{32}$$

for $i = 1, ...N_e$, where $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}, \tag{33}$$

is a Monte Carlo estimate of the Kalman gain.

So far, we have not specified what an appropriate ensemble size, N_e, could be. We know that a "large" ensemble size leads to small error, because Monte Carlo error reduces as $1/\sqrt{N_e}$, but it is unclear what constitutes a sufficiently "large" ensemble. Indeed, the required ensemble

size depends, among other things, on the dimension of the state space $\mathbf{x}$ and the quality and extent of the observations $\mathbf{y}$ (Chorin & Morzfeld 2013). Perhaps most importantly, the computational expense of the numerical model limits the ensemble size. In geomagnetic DA, a practical ensemble size is a few hundred, which is small compared to the state dimension, which is in the millions.

3.2 Keeping the ensemble size manageable: Localization and inflation

The small ensemble size used in ensemble DA implies that sampling error in the Monte Carlo estimates of the EnKF is large. *Localization* and *inflation* are techniques that reduce sampling error and therefore make ensemble DA feasible at reasonable ensemble sizes. Indeed, it is widely accepted that localization and inflation are required in all but the simplest cases (Hamill et al. 2009; Morzfeld et al. 2017; Harty et al. 2021).

3.2.1 Localization

The idea of localization originated in NWP, where correlations are known to decay with spatial distance (hence the name *localization*). The basic idea is that sampling error causes spurious long-range correlations which should be dampened. Localization is the process in which spurious correlations and, therefore, sampling error are reduced. In geomagnetic DA, where observations are not spatial, but spectral, similar ideas can be used. The reference to a spatial correlation structure, however, must be replaced by a spectral correlation structure. Nonetheless, we will use the common terminology of localization, with the understanding that this is a misnomer and we really mean to reduce sampling error in spectral (forecast) covariances.

One common way of implementing localization is to define a localization matrix $\mathbf{L}$, and to define a *localized* forecast covariance by the element-wise product

$$\mathbf{P}^f_{\text{loc}} = \mathbf{L} \circ \mathbf{P}^f. \tag{34}$$

The localized forecast covariance then replaces the forecast covariance in (33) for the computation of the Kalman gain within an EnKF. A common choice for the localization matrix is

$$L_{i,j} = \exp(-(\mathrm{d}_{i,j}/\rho)^2), \tag{35}$$

where $L_{i,j}$ are the elements of $\mathbf{L}$, $\mathrm{d}_{i,j}$ is the spatial distance between the ith and jth state variables and ρ is a length scale which is tuned (see below). Recall that the forecast covariance

has (at most) rank $N_e - 1$, but if the localization matrix is high/full rank, the rank of the localized matrix is much larger than N_e (by the Schur product theorem). For this reason, a localized EnKF with a small ensemble size achieves an accuracy that far exceeds what one should expect, given the small ensemble size.

Most localization techniques in use in NWP or oceanography rely in one way or another on a measure of spatial distance between states and/or observations (see, e.g., Shlyaeva et al. 2019). All these approaches are problematic in geomagnetic DA, where no natural length scales of spatial distance exist due to the spectral nature of the state and observations. Put differently, the typical motivation for localization is to keep the effect of observations "local" (each observations only informs a subset of nearby state variables), but observations in geomagnetic DA are global (spherical harmonic coefficients). In Section 3.3, we describe a few localization schemes that are applicable to gobal/spectral observations, as they appear in geomagnetic DA.

3.2.2 Inflation

Sampling error due to small ensemble size results in underestimating forecast errors. A collection of methods known as *inflation* (see, e.g., Kotsuki et al. 2017) is used to compensate for this and, like localization, improve ensemble-based DA performance when ensemble size is small. In the numerical experiments below we use *multiplicative covariance inflation* (Anderson & Anderson 1999), where all entries of the forecast covariance matrix are increased by a fixed factor $\alpha \geq 0$. This is implemented by inflating the forecast ensemble

$$\mathbf{x}^f_{i,\text{infl}} = \sqrt{1+\alpha}(\mathbf{x}^f_i - \bar{\mathbf{x}}) + \bar{\mathbf{x}}, \tag{36}$$

for $i = 1, ...N_e$. The inflated ensemble has a (sample) covariance that is larger than the original sample covariance and subsequently replaces the forecast ensemble in the computation of $\mathbf{P}^f$, the Kalman gain, and the analysis ensemble. Notice that this approach uniformly increases the entries of the forecast covariance by a factor of $1 + \alpha$ and therefore, unlike localization, the adjustment has no impact on the estimated correlations of $\mathbf{P}^f$. Instead it increases the estimated uncertainty in the forecast prior with the effect that the forecast is deemed largely uncertain and, hence, more weight is given to the observations.

3.2.3 Tuning localization and inflation

Once one has decided on particular localization and inflation schemes, one has to determine the (hyper) parameters that define the schemes, e.g., the correlation length scale and the inflation

level. This is done via a tuning process as follows. The extent of localization influences the required inflation and vice versa so that the parameters that define localization and inflation must be tuned *simultaneously*. A simple way to tune localization and inflation is to do a "grid search": A DA is performed for a pre-defined set of localization/inflation parameters and the combination that leads to the smallest errors, or best forecast skill, is deemed to be "optimal." For already computationally intensive DA systems, the computational cost of this tuning is significant and an increasing body of work is emerging that discusses approaches which require little to no experimentation to determine useful localization (see, e.g., Anderson 2012, 2016; Zhen & Zhang 2014), and inflation (see, e.g., Li et al. 2009; Gharamti et al. 2019), or both (Lunderman et al. 2021).

3.3 Localization schemes for geomagnetic DA

We present five localization schemes that are reasonable to try in geomagnetic DA because they do not rely on a spatial correlation structure. The first four schemes rely on a localization matrix $\mathbf{L}$ and depend on a single parameter which is tuned (see above). The first scheme, a *shrinkage* scheme, reduces cross-covariances by a uniform amount. The second scheme is based on *climatological correlations* and attempts to enforce a correlation structure on the forecast covariance that is similar to the climatological covariance – the covariance computed from snapshots of a long run of the model. The climatological localization is most similar to the "checkerboard" localization (Sanchez et al. 2019, 2020), where the localization matrix is also the result of studying climatological correlations. The third and fourth schemes localize by *wave-vectors* or *wave-numbers*. We describe these two schemes only for the proxy on the square (Fourier modes), because it will turn out that these schemes do not perform well (see Section 4). For that reason, we do not perform additional testing with the spherical proxy. Nonetheless, we wish to bring up these schemes because they are somewhat natural possibilities to try, and being able to try new ideas *without* incurring a huge computational cost is one of the advantages of a proxy model. Moreover, studying how the wave-number and wave-vector localization schemes "fail" actually provides useful insights into localization in geomagnetic DA. The fifth localization scheme is the *sample error correction scheme* (SEC) (SEC, Anderson 2012). SEC differs from the above localization schemes in that it modifies the Kalman gain, rather than the forecast covariance, and that it does not require tuning because it is adaptive.

Shrinkage localization. "Shrinkage" techniques are routinely used for estimating high-

dimensional covariances based on a small sample (see, e.g., Touloumis 2015). The idea is to estimate the covariance $\mathbf{P}^f_{\text{loc}}$ by a weighted sum of the sample covariance $\mathbf{P}^f$ and a target covariance $\mathbf{T}$, i.e., $\mathbf{P}^f_{\text{loc}} = (1-\rho_1)\mathbf{P}^f + \rho_1\mathbf{T}$ where $0 \leq \rho_1 \leq 1$. We use the diagonal matrix of the sample variances (the diagonal of $\mathbf{P}^f$) as the target matrix. With this choice, shrinkage localization uses a localization matrix $\mathbf{L}$ consisting of all ones along the diagonal and values of $1-\rho_1$ on all off-diagonal entries, and effectively reduces all correlations by a factor of ρ_1. Shrinkage localization is useful if there is a net benefit to reducing *all* sample-based correlations, to avoid the damaging effects of spurious correlations that arise from a small ensemble size.

Climatological localization. The rationale for climatological localization is that the forecast covariance matrix may have a similar correlation structure as the climatological covariance. One thus first performs a long model run, with initial conditions from an already well developed solution, and collects "snapshots" of the model, taken at regular (or even irregular) time intervals. The snapshots are used to generate a correlation matrix that represents the climatological correlations. Subsets of these matrices are shown in figure 5. In the top row we see that climatological correlations between modes of the magnetic field (A) are nearly zero with the exception of some weak correlations between low-order spherical harmonics for the model on the sphere. The bottom row of figure 5 indicates that for both geometries there exist strong climatological correlations between modes of the magnetic field (A) and velocity field (ω). A localization matrix is obtained from the climatology, by raising the absolute values of the entries of the correlation matrix to the power $1/\rho_2$, where $\rho_2 > 0$ is a localization parameter (to be tuned). In short, the effect of a climatological localization is that the stronger the correlation between two elements of the state space over time (according to the offline, long, free model run) the more a sample correlation in $\mathbf{P}^f$ is trusted. Climatological localization is similar to the "checkerboard" localization (Sanchez et al. 2019, 2020), where a localization is constructed from climatological correlations.

Wave-vector localization (square). We base the localization on a "distance" defined by wave-vectors. Specifically, we retain correlations between Fourier modes of similar total wave number and orientation, but dampen correlations between modes of different length-scales or orientations. We implement wave-vector localization via the localization matrix $L_{i.j} = \exp(-(\mathrm{d}_{i,j}/\rho_3)^2)$ with $d_{i,j} = ||\mathbf{k}^i - \mathbf{k}^j||$ where $\mathbf{k}^i$ and $\mathbf{k}^j$ are the wave-number vectors of the Fourier modes corresponding to the coefficients in the ith and jth elements of the state space, and $||\cdot||$ denotes the Euclidean norm. Thus, modes of similar total wave number but

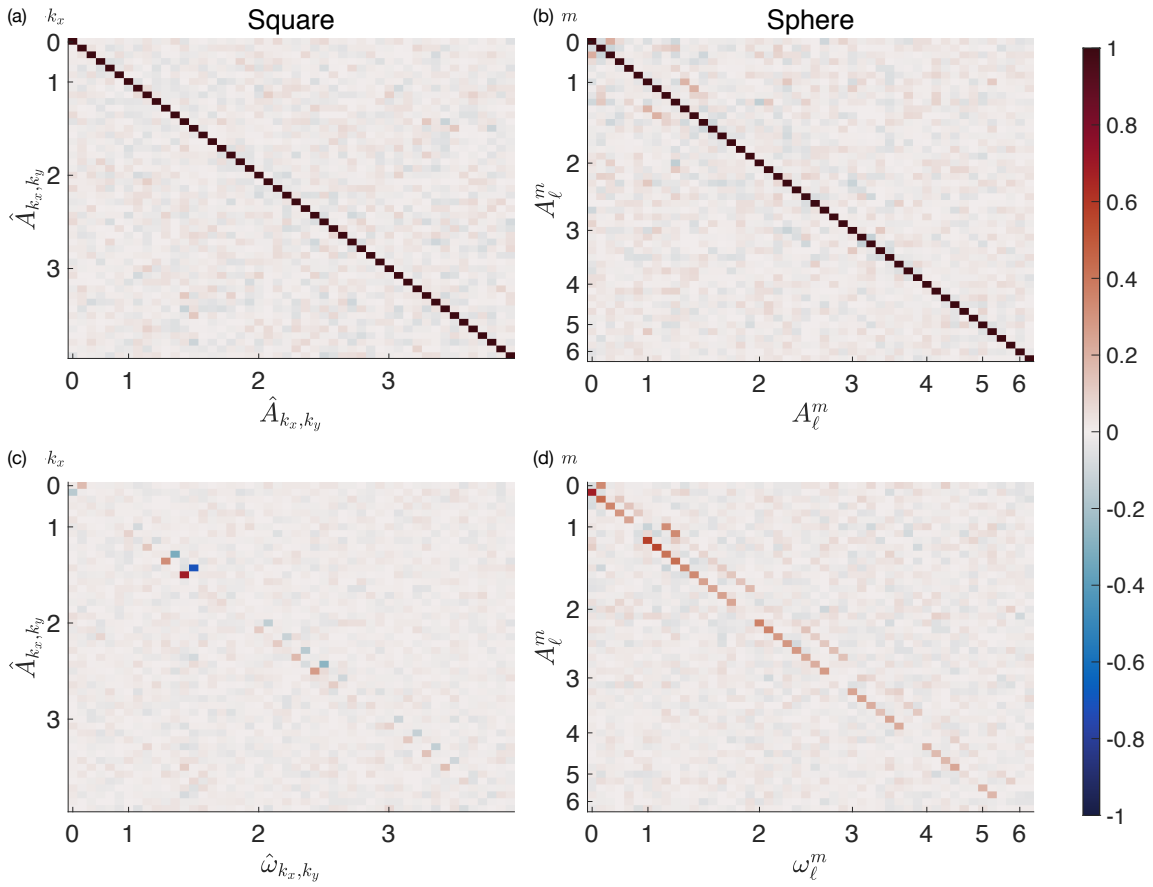

Figure 5. Subsets of the climatological correlation matrices of the proxy on the square (left column, grouped by k_x and truncated at $|k_x|, |k_y| = 3$) and the proxy on the sphere (right column, grouped by m and truncated at $m = 6$). (a) Correlations of the proxy on the square between modes of A. (b) Correlations of the proxy on the sphere between modes of A. (c) Correlations of the proxy on the square between modes of A and ω. (d) Correlations of the proxy on the sphere between modes of A and ω.

different orientation are "far apart" and sample correlations between them are suppressed. For example, if $\mathbf{k}^i = 3\hat{\mathbf{x}} + 3\hat{\mathbf{y}}$ and $\mathbf{k}^j = 3\hat{\mathbf{x}} - 4\hat{\mathbf{y}}$, then the modes have a similar total wave number but a different orientation and we get $d_{i,j} = 7$ (strong localization). If, on the other hand, $\mathbf{k}^i = 3\hat{\mathbf{x}} + 3\hat{\mathbf{y}}$ and $\mathbf{k}^j = 3\hat{\mathbf{x}} + 4\hat{\mathbf{y}}$, then the modes have a similar total wave number *and* orientation and, thus, $d_{i,j} = 1$ (mild localization).

Wave-number localization (square). Sample correlations between modes corresponding to similar total wave numbers are largely retained while sample correlations between modes of different length-scales are reduced. The localization matrix takes the form $L_{i.j} = \exp(-(\mathrm{d}_{i,j}/\rho_4)^2)$ with $d_{i,j} = |k^i - k^j|$ where k^i and k^j are the total wave numbers of the Fourier modes corresponding to the coefficients in the ith and jth elements of the state space. With this approach,

only the difference in length scales of Fourier modes are relevant. For example, by this measure, elements of the state space corresponding to modes with wave-number vectors of $\mathbf{k}^i = 3\hat{\mathbf{x}} - 4\hat{\mathbf{y}}$ *or* $\mathbf{k}^j = 3\hat{\mathbf{x}} + 4\hat{\mathbf{y}}$ will be "equidistant" from other parts of the space, since for either mode the total wave number is $k = \sqrt{3^2 + 4^2} = 5$.

Sampling error correction (SEC). In the sample error correction scheme (SEC, Anderson 2012), the larger the magnitude of a sample correlation, and the larger the ensemble from which it is computed, the more it is trusted. The SEC scheme does not operate through covariance localization as the four schemes described above. Instead, each element of the Kalman gain (see (33)) is multiplied by a factor between zero and one, which depends only on the ensemble size and the sample correlation between the observed state and the element of the state space being adjusted. In general, the larger the ensemble and the larger the magnitude of the sample correlation, the closer the factor is to one. An offline Monte Carlo simulation is used to construct a lookup table of factors as a function of ensemble size and sample correlation.

3.4 Observing system simulation experiments (OSSEs)

One can use the proxy to perform a systematic numerical study via observing system simulation experiments (OSSEs). In OSSEs (sometimes also called twin experiments), one generates observations using a model and subsequently assimilates the synthetic observations back into a model. The simulation with the model that generates the data is called the *nature run*. OSSEs have the advantage that one has full control over the observation network, observation errors, and one also has full knowledge of the system states, which allows the study of (forecast) errors in observed *and* unobserved quantities. For this reason, OSSEs are a typical first step when testing the impact of new observations, or for testing DA technology in many Earth science problems. In fact, OSSEs are routinely used at NASA and NOAA (see, e.g., Zeng et al. 2020; Hoffman & Atlas 2016; Errico et al. 2013).

The proxy models can be used for OSSEs. A long run of the models is performed using initial conditions taken from already well-developed solutions (the end of long runs starting from random perturbations). A portion of these solutions are recorded as the nature runs and from the remaining part, we randomly draw initial ensemble members as needed. Emulating the situation in geomagnetic DA, it is reasonable to only observe the low-frequency spectral coefficients of the proxy magnetic field. In our numerical experiments below, we use Gaussian additive errors with a diagonal observation error covariance $\mathbf{R}$ (uncorrelated errors) to gen-

erate the synthetic observations. We then supply the DA system with knowledge of the error statistics, but one can easily consider different choices. In the experiments below, the nature run is generated with the proxy that is used during the assimilation, but future studies may use different configurations. For example, one can use a higher resolution proxy in the nature run, and a lower resolution proxy for the DA. One can also envision using one set of parameters (e.g., modified Rayleigh or magnetic Reynolds numbers) in the model that generates the synthetic data, and another set of parameters in the model that assimilates these data. One can further supply the DA with false information about observation errors, e.g., one can vary the observation error covariance or one could generate synthetic observations that violate the usual Gaussian error assumptions. Such a settings may in fact more realistically represent geomagnetic DA, where coarse models are known to be biased, and where many assumptions or approximations are known to be inadequate. Nonetheless, the experiments with idealized configurations we present below is a typical and necessary first step towards understanding how localization can function (or not) in geomagnetic DA.

4 RESULTS OF DATA ASSIMILATION EXPERIMENTS

We carry out a series of DA experiments (OSSEs) with the proxy model. These experiments are designed to reveal if localization can help reduce the required ensemble size in EnKF and, therefore, make EnKF more effective in geomagnetic DA. We begin by discussing the general framework of the numerical experiments and metrics for evaluation of DA performance. Because the proxy on the square is easier to compute with (FFT vs. spherical harmonic transforms), we perform a large number of experiments with the proxy on the square, learn our lessons, and present a condensed set of numerical experiments with the spherical proxy. This hierarchical approach is effective because the proxies on the square and sphere share many characteristic features, but the proxy on the sphere, is dynamically more complex (more modes carry energy and more energy is transferred across spatial and temporal scales).

4.1 Observation network and metrics of success

All numerical experiments are OSSEs as explained above. We use the same observation network in all numerical experiments, i.e., the observed coefficients and the time interval between observations is the same for all experiments. Specifically, on the square, we observe Fourier modes of A with wave numbers in the x and y directions that are less than or equal to three. On the sphere, we observe coefficients of A corresponding to spherical harmonics of degree five

or less. In both geometries, the observed modes represent around 95% of the mean energy in the magnetic perturbation field. The time interval between observations, the *analysis time*, is 20% of the typical timescale of the magnetic perturbation field (the mean of $||A||_2/||\partial A/\partial t||_2$) or around 7% of the e-folding time. We assume (and use) Gaussian observation errors with a diagonal observation error covariance ($\mathbf{R}$). The standard deviations, which are the square roots of the diagonal elements of $\mathbf{R}$, are set to 1% of the respective coefficient's mean magnitude over the course of the nature run.

To evaluate the performance of the DA we consider the *observation-minus-forecast* (OmF) residual, the *truth-minus-forecast* (TmF) error and the *ensemble spread*. The OmF of one DA cycle is defined as

$$\mathrm{OmF} = \sqrt{(\mathbf{H}\bar{\mathbf{x}} - \mathbf{y})^T(\mathbf{H}\bar{\mathbf{x}} - \mathbf{y})}, \tag{37}$$

where $\bar{\mathbf{x}}$ is the mean of the forecast ensemble (the ensemble prior to assimilating the observations $\mathbf{y}$). In other words, we take the square root of the sum of the squares of the differences in the observations and their forecasted values. Note that for the normalized Fourier and spherical harmonic coefficients, this amounts to transforming the difference into physical space and computing the L^2 norm over the domain. For this reason, we do not average the errors (divide by the number of observations) as in the root mean squared error (RMSE). Our main motivation for OmF is that is a measure of performance in an operational DA system.

TmFs (truth minus forecast) are very useful for diagnosing how well a DA system performs on observed and unobserved quantities, but since TmF requires knowledge of a "true" state, TmFs can only be used in OSSEs. We compute TmFs separately for the magnetic field and velocity field, because of their differing orders of magnitude. We thus define

$$\mathrm{TmF}_A = \sqrt{(\bar{\mathbf{x}}_A - \mathbf{x}_A)^T(\bar{\mathbf{x}}_A - \mathbf{x}_A)}, \quad \mathrm{TmF}_\omega = \sqrt{(\bar{\mathbf{x}}_\omega - \mathbf{x}_\omega)^T(\bar{\mathbf{x}}_\omega - \mathbf{x}_\omega)}, \tag{38}$$

where $\mathbf{x}_A$, $\mathbf{x}_\omega$ are from the nature run ("truth"), and $\bar{\mathbf{x}}_A$, $\bar{\mathbf{x}}_\omega$ are the forecast means of the EnKF. We note that this amounts to computing, in the physical domain, the L^2 norm of the difference between the forecast and nature run, for the fields A and ω. As with OmF, the defintion of the TmFs is, therefore, different from the usual RMSE.

The ensemble spread (spread, for short) typically indicates an average variance in the forecast and is defined by $\sqrt{\mathrm{tr}(\mathbf{P}^f)/n}$, where $\mathbf{P}^f$ is the forecast ensemble covariance matrix, $\mathrm{tr}(\mathbf{P})$ denotes the trace of a matrix $\mathbf{P}$ (the sum of the diagonal elements), and where n is the overall state dimension. The spread is compared to the forecast root mean square error (RMSE) and in a well-tuned and well-working DA system, the spread should be approximately equal to the RMSE, so that expected errors are comparable to actual errors. In geomagnetic

DA, using non-averaged errors such as TmF and OmF is more reasonable because of their physical interpretation (see above). For this reason, we modify the usual definition of spread to be in line with the definitions of TmF and OmF as follows. We compute spread separately for the observed magnetic field, the overall magnetic field, and the velocity field as

$$\text{Spread}_{\text{obs}} = \sqrt{\text{tr}(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})}, \quad \text{Spread}_A = \sqrt{\text{tr}(\mathbf{P}_A^f)}, \quad \text{Spread}_\omega = \sqrt{\text{tr}(\mathbf{P}_\omega^f)}, \tag{39}$$

where $\mathbf{P}_A^f$ and $\mathbf{P}_\omega^f$ are covariances of the ensemble of forecasts for the spectral coefficients of the fields A and ω. With these definitions, an indicator for a well-tuned DA system is that

$$\text{OmF} \approx \text{Spread}_{\text{obs}}, \quad \text{TmF}_A \approx \text{Spread}_A, \quad \text{TmF}_\omega \approx \text{Spread}_\omega. \tag{40}$$

We note the addition of the observation error covariance ($\mathbf{R}$) in (39), which is necessary if observation errors are comparable to estimation errors to avoid overfitting the data.

Throughout the rest of this paper, we scale OmFs, TmFs and corresponding spreads by the macroscopic errors. This has the effect that an error of approximately one means that the DA is ineffective – it produces forecast errors about as large as what one would get with an independent simulation. The macroscopic errors are determined from the same runs used to estimate the e-folding time (see Section 2.4.2). Specifically, we average OmF and TmFs, in time and over the perturbations, once these errors stopped to increase exponentially.

4.2 Results: proxy on the square

We present and discuss results of DA experiments with the proxy model on the square. As mentioned above, we perform a large number of experiments with this proxy because of its numerical simplicity and computational affordability compared to the spherical proxy (FFT is faster than spherical harmonic transform).

4.2.1 What error reduction can we expect: Experiments with large ensembles

Since computations with the proxy on the square are easy, we can use EnKFs with large ensemble sizes and without localization or inflation. In fact, we can determine the smallest ensemble size that leads to asymptotically small errors (large ensemble size limit), i.e., forecast errors have stabilized, and no improvements result from further increasing the ensemble size. Forecast errors in the large ensemble size limit will serve as a baseline for the localization and inflation techniques. If localization and inflation are appropriate, one should be able to obtain similarly small errors, but with a much smaller ensemble size.

The nature run for these experiments is 601 analysis times long (over 40 e-folding times)

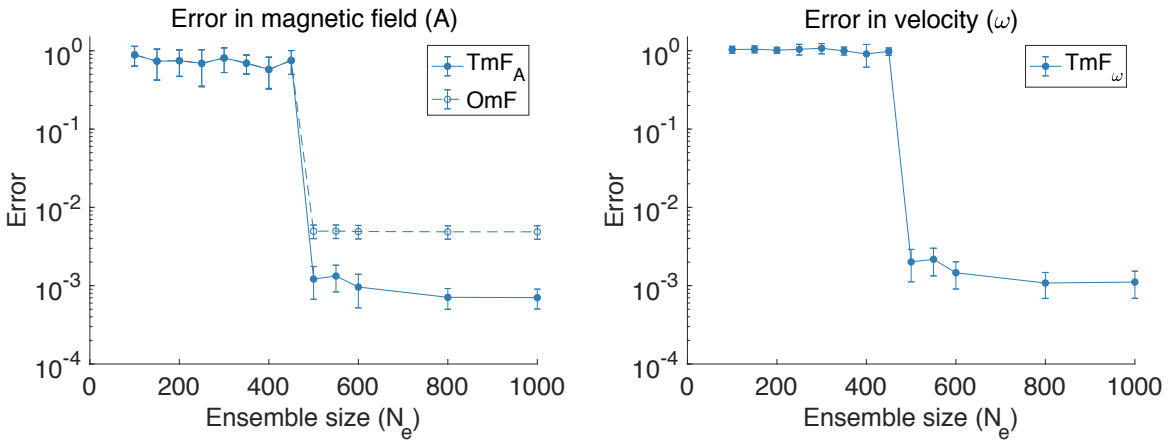

Figure 6. Forecast errors as a function of ensemble size (no localization or inflation). (a) Errors in magnetic field (OmF and TmF). (b) Errors in velocity (OmF). All errors are scaled by corresponding macroscopic errors, so that an error of about one means that the DA is not reducing forecast errors. The open/filled dots are average errors and the error bars indicate one standard deviation (all averages are over time).

and we apply the EnKF with ensemble sizes from $N_e = 300$ to $N_e = 1000$. The mean and standard deviation of OmFs and TmFs are computed over the last 300 cycles to avoid contamination of the errors by the EnKF's spin-up period (see Section 4.2.3). We apply no localization or inflation. The results are summarized in Figure 6, which shows OmF and TmFs as a function of the ensemble size. The OmFs drop sharply (two orders of magnitude) and become relatively stable around $N_e = 500$, with the TmFs in the magnetic field and the unobserved velocity leveling off in the same way for ensemble sizes around $N_e = 800$. With smaller ensemble sizes, we observe macroscopic error, and the DA has no effect. We thus conclude that

(i) an ensemble size of at least $N_e = 500$ is required for stable, reliable and accurate DA;
(ii) the DA is able to reduce errors by more than two orders of magnitude.

It is thus clear that DA can have a huge effect on forecast accuracy, but this accuracy comes at the cost of a large ensemble size. Moreover, if the proxy we created is indeed simpler than realistic geodynamo models, then our experiments suggest that ensemble DA *without* localization or inflation requires ensemble sizes much larger than 500. For example, an ensemble size of $N_e = 1000$ is not sufficient for effective DA on the spherical proxy (see section 4.3), which is only mildly more dynamically rich than the proxy on the square. The remaining numerical experiments, however, are designed to investigate if localization and inflation can boost the

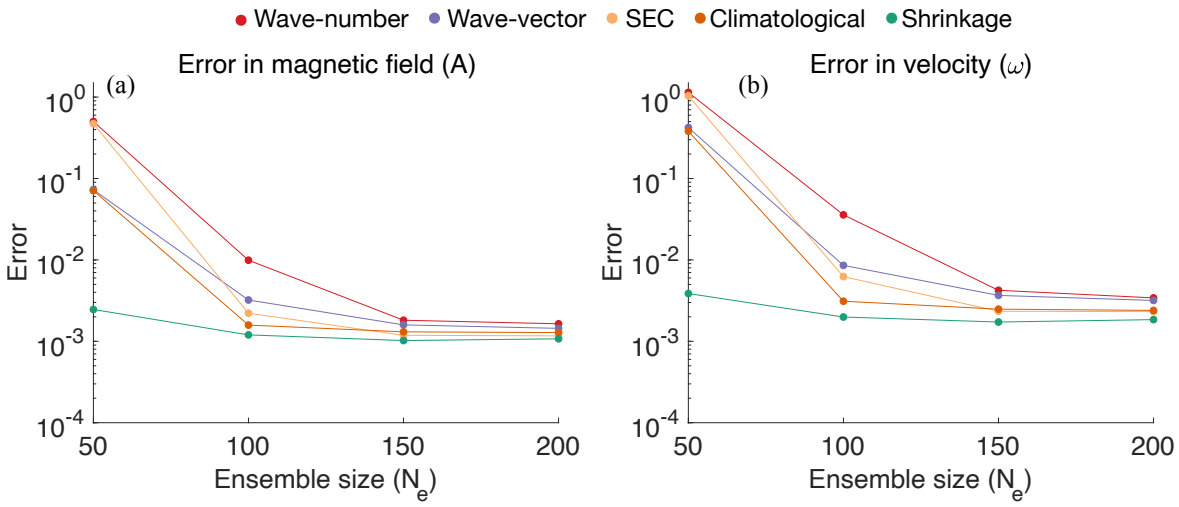

Figure 7. Forecast errors (TmF) as a function of ensemble size (N_e) for EnKFs when using optimal levels of localization and inflation (for each particular combination of scheme and ensemble size). (a) Errors in magnetic field. (b) Errors in velocity. All errors are scaled by corresponding macroscopic errors.

computational efficiency of geomagnetic DA by achieving the accuracy of a large ensemble EnKF at a much smaller ensemble size.

Finally, we note that OmF can be larger than TmF. The reason is that observation errors (the additive Gaussian noise) comes into play for OmF, but not for TmF. Because the observation errors are larger in magnitude than the unobserved components of A we have that OmF>TmF$_A$.

4.2.2 Computational efficiency of DA with (tuned) localization and inflation

We now apply localization and inflation. This requires tuning via a grid-search: We chose a number of localization and inflation parameters, perform a DA experiment with each combination, and then compute the forecast error. This results in a gridded error, and we base an "optimal" localization on these errors (see Section 4.2.4 for more details). The optimal localization and inflation further depend on the ensemble size, so we repeat this tuning step at each ensemble size we consider.

With tuned localization and inflation, we carry out an OSSE with 601 analysis times and neglect the first 300 DA cycles as spin-up. The errors (TmFs) one obtains with a localized and inflated EnKF are illustrated in Figure 7, where we show TmFs as a function of ensemble size. The numerical experiments reveal three interesting facts:

Table 3. Average errors (OmF and TmFs), scaled by the respective macroscopic errors and multiplied by 10^{-3} (proxy on the square). The large ensemble EnKF ($N_e = 1000$) does not use localization or inflation. All other EnKFs are run with ensemble size $N_e = 100$ and use the optimal levels of localization and inflation (for the particular scheme and $N_e = 100$).

		Magnetic field			Velocity	
	OmF	$\text{Spread}^f_{\text{obs}}$	TmF^f_A	Spread^f_A	TmF^f_ω	Spread^f_ω
Large Ens.	4.9	4.9	0.7	0.7	1.1	1.1
Shrinkage	5.0	5.1	1.2	1.7	2.0	3.3
Climatological	5.1	5.2	1.6	1.8	3.1	3.8
Wave-vector	5.7	5.7	3.2	3.4	8.6	9.1
Wave-number	9.4	11.8	9.9	14.0	35.8	55.3
SEC	5.3	5.4	2.2	2.4	6.2	7.4

(i) All localization schemes we tried lead to dramatically reduced errors, but at small ensemble sizes.
(ii) How quickly errors converge with ensemble size depends on the localization scheme used, with the shrinkage scheme exhibiting the fastest convergence.
(iii) The localization scheme has an effect on the smallest error one can achieve with a localized EnKF.

Thus, while every localization leads to dramatic improvements over the "vanilla" EnKF without localization and inflation, the details of the localization are indeed very important. This is perhaps intuitive because the localization scheme that most accurately reflects the "true" underlying correlation structure should lead to the smallest errors.

We now focus on errors at ensemble size $N_e = 100$ to investigate how the details of the localization affect forecast error. The errors for all five localization schemes and for the large ensemble EnKF ($N_e = 1000$) are summarized in Table 3. As indicated above, all localization schemes drastically reduce OmFs (by a factor of 2 or more). Indeed, the OmFs of the localized EnKFs with $N_e = 100$ are comparable to the OmFs of the EnKF with $N_e = 1000$ *without* localization and inflation (perhaps with the exception of wave-number localization). Moreover, the ensemble spread is comparable to the forecast error, indicating that the DA works as it is supposed to. Thus, when considering OmF only, it is difficult to say which localization scheme is most appropriate, because the differences between the shrinkage, climatological, SEC and wave vector schemes are minor. Perhaps the wave-number localization can be ruled out at this stage because the OmFs are much larger than for the other schemes.

The various differences between the localization schemes are more dramatic when one

Table 4. Longe-range forecast error (four e-folding times). Listed are average forecast errors (OmFs and TmFs), scaled by the respective macroscopic errors. The large enssemble EnKF uses $N_e = 1000$, all localized/inflated EnKFs use $N_e = 100$.

	Large Ens.	Shrinkage	Climatological	Wave-vector	Wave-number	SEC
OmF	0.17	0.29	0.41	0.83	0.74	0.69
TmF_A	0.18	0.30	0.43	0.83	0.75	0.69
TmF_ω	0.18	0.29	0.48	0.80	0.93	0.65

considers TmFs. In fact, none of the localization schemes we tried achieves TmFs that are comparable to those of the large ensemble EnKF. This means that none of the schemes we applied accurately capture the underlying correlation structure. Nonetheless, the shrinkage scheme overall gives the smallest errors, followed by climatological localization, SEC, and wave-vector localization. Wave-number localization again leads to much larger errors. We also note that the ensemble spreads for TmFs tends to overestimate the uncertainty (Spread > TmF). What this implies is that while different localization schemes may perform similarly when considering OmF, errors in unobserved system components can be dramatically different. This is important in geomagnetic DA, where one of the main motivations is to infer the core state from measurements of only the magnetic field. Our numerical experiments suggest that it is difficult to determine a "truly appropriate" localization scheme based on only short-term forecasts of observed quantities.

One way forward is to consider longer range forecast, i.e., forecast that stretch over more than one assimilation window. The reasons are as follows. Errors in unobserved quantities become more important as time evolves, due to nonlinear couplings. More appropriate localization schemes improve estimates of unobserved quantities, because a more appropriate correlation structure is enforced. Thus, the better the localization, the smaller the forecasts errors. This is true at any scale, but when considering observed quantities only (OmF), the differences between localization schemes become more apparent with longer range forecasts.

We now test this idea with the proxy by performing forecasts over four e-folding times. Specifically, we compute forecast errors for five independent forecasts and average the results, shown in Table 4. The ranking of (average) long-range forecast errors generally reflects that of the short-range forecasts (see Table 3), but with OmFs now clearly distinguishing between the better performing localizations. We can now identify the shrinkage scheme as the most appropriate localization scheme, based only on observed quantities (OmFs) – we no longer rely on TmFs for that conclusion. In other words, our earlier experiments, with short range

forecasts, suggest that shrinkage and climatological localization perform similarly in terms of errors in observed quantities, but when taking TmFs into account, the shrinkage scheme has clear advantages. The longer range forecasts reveal the advantages *without* considering TmFs. This suggests that long range forecasts might indeed be an effective tool to identify useful localization schemes in geomagnetic DA.

4.2.3 Shortened spin-up with localized DA

In an effective EnKF system, forecast errors from an uninformed, initial ensemble are typically reduced over several assimiliation cycles before settling around a minimum. This period of transition from large to small forecast errors is referred to as the *spin-up time*. We previously examined average forecast errors only after the spin-up time. This is typical in many applications of DA, e.g., NWP, where one may spin-up only once, because new observations are routinely collected and their assimilation maintains accuracy. In geomagnetic DA, fundamental aspects of, e.g., the assimilation schemes, model parameters, and observational records being used, are often the focus of new studies (see, e.g., Tangborn & Kuang 2015; Sanchez et al. 2020). For this reason, assimilations frequently begin with relatively uninformative priors and must undergo a spin-up period. Given the limited observational record of the geomagnetic field, the rate of error reduction through initial assimilations is relevant and a short spin-up time is essential: If the spin-up time is long, the observational record may be insufficient for achieving the sought-after error levels and/or carrying out desired validation experiments. We now consider forecast errors during the spin-up time, for the purpose of determining how rapidly forecast errors are reduced from an initial ensemble.

Specifically, we compare the spin-up times of EnKFs with and without localization/ inflation, using a large ensemble size ($N_e = 1000$) for the unlocalized EnKF and $N_e = 100$ for the localized EnKFs. The TmFs in the magnetic and velocity fields over the first 200 assimilation cycles are shown in Figure 8. The average rate of decrease in forecast error over the initial assimilations varies among the runs with the large ensemble (blue) clearly, and unsurprisingly, being the most rapid. Generally, the shrinkage scheme (green) produces the lowest, or among the lowest errors, during spin-up time. The climatological scheme (orange) decreases slowly but steadily while the SEC (yellow) initially struggles along with the all-around, poor performing wave-number scheme (red).

For a more objective analysis, we record the number of assimilation cycles required to reduce forecast error to 1% of the macroscopic error in each of the two fields. The results are shown in Table 5 and serve to summarize and synthesize the time series in Figure 8. We find

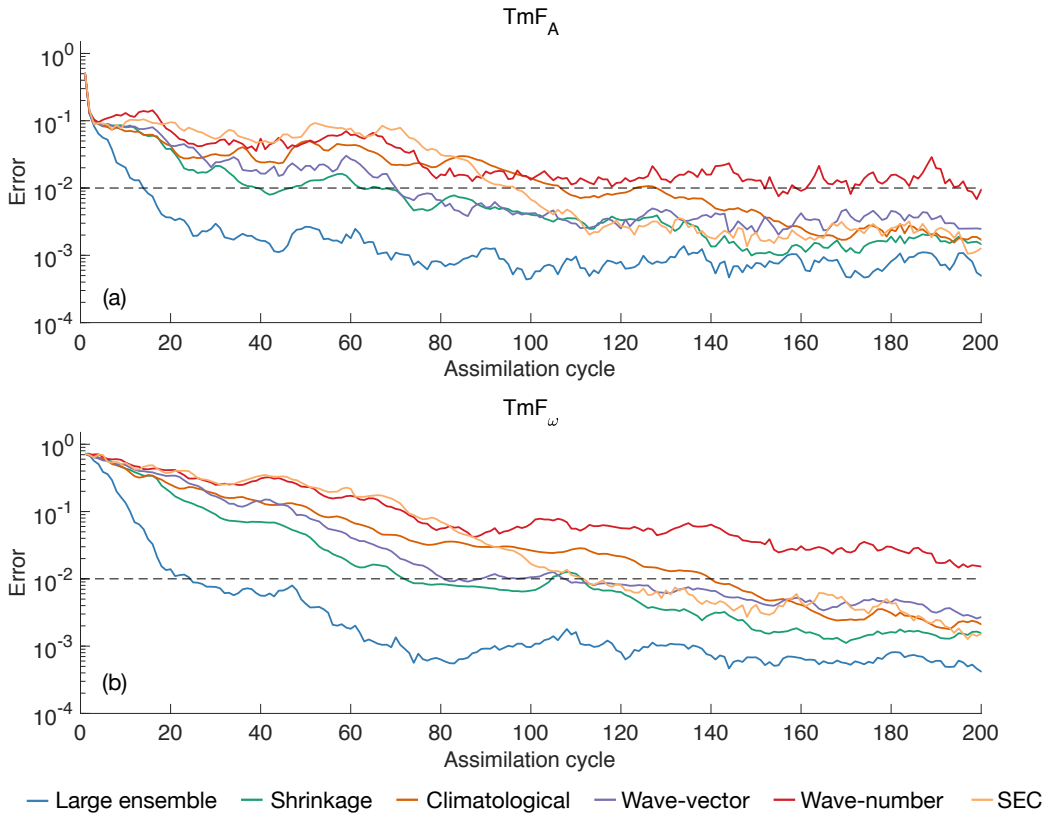

Figure 8. Forecast errors (TmF) as a function of the first 200 assimilation cycles for various EnKFs with and without localization/inflation. (a) Error in magnetic field. (b) Error in the velocity field. In both panels the blue line results from a run using a large ensemble of $N_e = 1000$ with no localization or inflation. The localized and inflated EnKFs (scheme indicated by color) use a smaller ensemble size of $N_e = 100$. All errors are scaled by the macroscopic errors. The black dashed line indicates an error level of 1% of the macroscopic error.

that an EnKF with a large ensemble requires 14 assimilation cycles to reduce error to 1% in the magnetic field, and 25 to similarly reduce errors in the velocity. The localized EnKFs require more assimilations to achieve similarly small errors. The shortest spin-up with $N_e = 100$ is achieved by an EnKF with shrinkage localization, followed by wave-vector localization, SEC, and climatological localization. The wave-number scheme takes the longest to reduce error to 1% in the magnetic field and never really stabilizes in the velocity field. We thus see that the localization schemes that lead to the smallest errors also reduce the overall spin-up time of an EnKF. Localization and inflation are thus not only useful to bring down the "stationary" errors, but they are also critical to reduce the number of assimilation cycles one has to do before one reaches acceptable error levels.

Table 5. Assimilation cycles required to reduce forecast errors in the magnetic field (first row) and velocity field (second row) to 1% of the macroscopic error. The large ensemble EnKF uses an ensemble size $N_e = 1000$ and no localization or inflation. The remaining EnKFs use localization and inflation (scheme indicated) and an ensemble size $N_e = 100$.

	Large Ens.	Shrinkage	Climatological	Wave-vector	Wave-number	SEC
Magnetic field	14	40	107	71	152	97
Velocity	25	72	140	81	-	112

4.2.4 Localization vs. inflation and details of the tuning process

We discuss some details of the tuning of localization and inflation, and how to determine an "optimal" localization scheme by resolving a trade-off between localization and inflation. First, a typical gridded error, as obtained during the tuning of localization and inflation, is shown in Figure 9. Red regions contain localization and inflation parameters that result in macroscopic errors (OmF ≈ 1); green region contains the localization and inflation parameters that result in small OmF (OmF $\approx 10^{-3}$–10^{-2}). We notice a pattern that less aggressive localization (smaller ρ) leads to more aggressive inflation (larger α). For example, at $\rho = 90\%$, inflation is $\alpha = 3\%$, but at $\rho_1 = 50\%$ the inflation is $\alpha = 12\%$. We observe this pattern in all four schemes we tried. The fifth scheme, the SEC, is adaptive and does not have a tunable localization parameter. We thus only tuned inflation and found that only a large inflation could reduce the OmF.

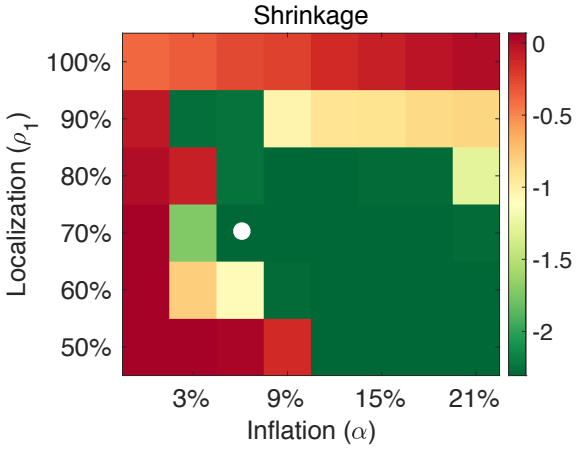

Figure 9. Base 10 logarithm of OmFs, scaled by macroscopic error, of an EnKF with ensemble size $N_e = 100$ and different localization (ρ) and inflation (α) parameters (shrinkage localization). Red indicates ineffective DA (macroscopic error); green indicates an effective DA with OmFs comparable to those of an EnKF with a much larger ensemble. A white dot indicates the parameters chosen for further experiments.

This is indeed in line with what we find with the non-adaptive localization schemes: The SEC is not a very "aggressive" localization scheme and, for that reason, requires relatively large inflation.

The pattern, or trade-off, between localization and inflation is in fact an important detail: Larger values of inflation mean that less weight is given to the prior in each assimilation and that the ensemble is pushed more towards the observations than with lower inflation. Thus, small OmFs in the presence of increasingly large inflation can simply be the result of pulling the observed part of a model closer to the observations, while estimates of unobserved quantities become poorer (over-fitting). For this reason, in an operational assimilation system where one only has access to OmFs as a measure of forecast accuracy, it is desirable to minimize the level of inflation used. Following this line of thinking, we base our notion of optimal localization and inflation on successfully reducing OmF at low inflation.

We further observed that localization and inflation combinations near regions of large OmF may not be useful. For example, in Figure 9, values of $\rho_1 = 90\%, \alpha = 3\%$ are successful in reducing OmFs, but when applied to a different nature run this combination failed to reduce OmF. To avoid localization and inflation which may be unreliable in this way, we define optimality by the localization with lowest inflation, that successfully reduces OmF, while not being in the immediate vicinity of macroscopic error. For example, from Figure 9 we define the combination of values indicated by the white dot at $\rho_1 = 70\%, \alpha = 6\%$ to be "optimal."

4.2.5 Summary and conclusions drawn from the proxy on the square

We summarize the main conclusions supported by the numerical experiments with the proxy on the square.

(i) The EnKF can drastically reduce forecast errors (by two orders of magnitude).
(ii) Localization and inflation dramatically reduce the computational cost of the EnKF because a small ensemble size is sufficient (in the proxy localization reduced the ensemble size from $N_e \approx 800$ to $N_e \approx 100$).
(iii) Localization and inflation reduce the spin-up time of an EnKF (the number of assimilation cycles required to reach a low error level).
(iv) The details of the localization are marginally important when one considers observed quantities (OmF) over short forecast horizons, but are critical when one also takes unobserved quantities (TmF) into consideration.
(v) Long range forecasts can be used to clarify differences between localization schemes that lead to similar errors in observed quantities, but very different errors in unobserved

quantities. More specifically, longer-range forecast are an effective tool to identify useful localization schemes based on observed quantities only.

The reasons for (iii) and (iv) are that during localization, one assumes an underlying correlation structure, and this assumed structure may be appropriate or not. We found that four out of five of the localization schemes we tried lead to a clear reduction in errors, but none of the schemes leads to errors that are comparable to those of a large ensemble EnKF. Thus, the schemes do not capture all relevant aspects of the underlying correlation structure. Many localization schemes may indeed exhibit similar performance when considering errors in observed quantities; longer range forecasts can be used to identify appropriate localization schemes.

Overall, we found that shrinkage localization is most effective. This is perhaps surprising because shrinkage localization draws on no motivation from the physics or statistics of the proxy model, but rather broadly suppresses correlations at a global level. Indeed, we find that when we rank the localization schemes by increasing errors, the schemes dampen fewer correlations. For example, the climatological localization suppresses correlations except for a few between certain modes of A and ω with the same wave-number vector. Thus, climatological localization performs similarly to shrinkage localization, because both schemes significantly suppress the vast majority of correlations. Wave-vector localization largely allows correlations between modes with similar wave-number vectors, whether they represent A or ω. Thus, the correlations suppressed by wave-vector localization are a subset of those suppressed by the climatological or shrinkage schemes, and we observe that forecast errors are slighlty larger for wave-vector localization than for shrinkage or climatological localization. Wave-number localization allows correlations between modes of similar total wave number, and, therefore, permits correlations similar to those of the wave-vector scheme *and* correlations between modes of similar length scales, but different orientations. For this reason, wave-number localization suppresses the least amount of correlation and indeed exhibits the largest errors. Thus, the success of a localization technique is, at least in part, related to the extent to which it provides a blanket suppression of all correlations and, as a consequence, limits the damage of spurious correlations.

To gain further insights, we consider the EnKF with a large ensemble size ($N_e = 1000$, no localization or inflation), and compute the ensemble correlation between an observed mode of A with wave-vector $\mathbf{k} = \hat{\mathbf{x}}$ and an unobserved, high-energy mode of ω, with wave vector $\mathbf{k} = 2\hat{\mathbf{x}} - 4\hat{\mathbf{y}}$. We take this large ensemble as an indicator of what the relevant correlation structure may be. The time series of correlation, along with a histogram of the correlation, is

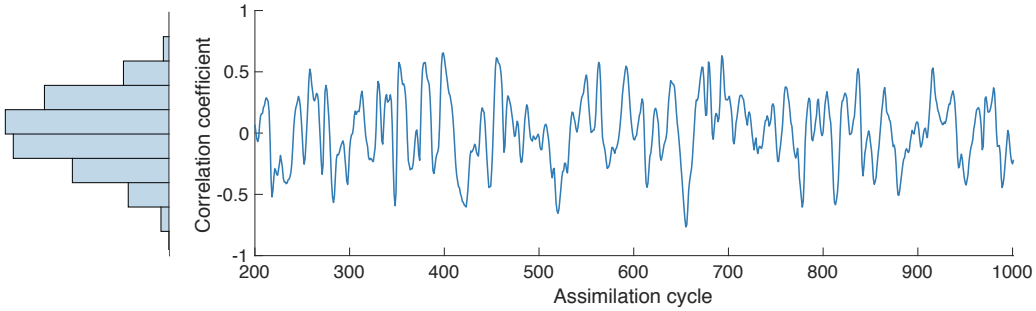

Figure 10. Forecast ensemble correlation between an observed mode of A and an unobserved mode of ω for 800 forecasts during a DA run using an ensemble of $N_e = 1000$ and no localization or inflation. The left portion of the figure provides a histogram illustrating the distribution of the correlations. Despite the significant correlations seen above, the modes have shown no strong climatological correlation and differ in wave-number/vector.

shown in Figure 10. We note that while the wave vectors are certainly of different direction and magnitude, the large ensemble regularly indicates strong correlations between these modes. Moreover, the climatological correlation of these two modes is -0.02, which differs quite a bit from the actual correlation at many instances. At time instances when the climatological correlation differs from the instantaneous correlation, we expect that climatological localization is ineffective because it dampens a relevant correlation. Indeed, we observe this behavior not only in the two modes shown here. The large ensemble indicates strong correlations between many modes which are dissimilar in wave-vector/number and show no climatological correlation. This occurs because over short-term forecasts, perturbations propagate through the spectrum in complicated ways which can not necessarily be captured by the similarity of modes, or the long-term correlations within the system. For example, consider the Fourier expansions of A and ω

$$A = \sum_{k_x,k_y} \hat{A}_{k_x,k_y} e^{2\pi i(k_x+k_y)}, \quad \omega = \sum_{k_x,k_y} \hat{\omega}_{k_x,k_y} e^{2\pi i(k_x+k_y)}, \tag{41}$$

where $\hat{A}_{k_x,k_y}$ and $\hat{\omega}_{k_x,k_y}$ are the respective Fourier coefficients. Substituting into equations (17)-(19) it can be seen that the evolution of the Fourier coefficient $\hat{A}_{1,0}$ is described by

$$\frac{d\hat{A}_{1,0}}{dt} = \sum_{k_x^2+k_y^2\neq 0} \left[\frac{k_y}{k_x^2+k_y^2}\hat{\omega}_{k_x,k_y}\hat{A}_{1-k_x,-k_y}\right] - \frac{4\pi^2}{R_m}\hat{A}_{1,0} + \frac{\hat{\omega}_{1,0}}{2\pi}i. \tag{42}$$

The last two terms of the right hand side of (42) come from the linear terms of (19). We note that in the climatological covariance, the real and imaginary parts of $\hat{A}_{1,0}$ only exhibit strong correlations with the imaginary and real parts of $\hat{\omega}_{1,0}$, respectively. This is a natural

Table 6. Average errors (OmF and TmFs), scaled by the respective macroscopic errors and multiplied by 10^{-3} (proxy on the sphere). Both EnKFs are run with ensemble size $N_e = 100$.

		Magnetic field			Velocity	
	OmF	$\text{Spread}^f_{\text{obs}}$	TmF^f_A	Spread^f_A	TmF^f_ω	Spread^f_ω
Shrinkage	4.9	5.6	2.1	3.2	3.3	9.5
Climatological	5.5	6.3	3.2	4.9	8.6	18.4

consequence of the last term being proportional to $i\hat{\omega}_{1,0}$. The summation is the result of the non-linear term $-(\mathbf{v}\cdot\nabla)A$, and it is here that we see the wide-ranging mixture of modes which can influence $\hat{A}_{1,0}$. While the long-time variations in two modes may be uncorrelated, it is clear from (42) that perturbations in, e.g., $\hat{\omega}_{2,-4}$ can propagate to $\hat{A}_{1,0}$ over short-term forecasts (see figure 10). Moreover, the influence of the perturbation on $\hat{A}_{1,0}$ is dependent not only on the sign and magnitude of $\hat{\omega}_{2,-4}$, but also that of $\hat{A}_{-1,4}$. The result is that the structure of the forecast covariances vary with the states of the ensemble members. Indeed, since the correlation structure seems to vary, adaptive localization schemes might indeed be a constructive way forward (with hopes that adaptive strategies can be devised that capture the relevant correlations, which is *not* an easy task, as indicated by the relatively poor performance of the adaptive SEC scheme).

The above findings and discussion are immediately applicable to geomagnetic DA, where, thus far, only a variant of climatological localization has been applied to an ensemble-based system (Sanchez et al. 2019, 2020). Our results suggest that it may be well worthwhile to test a simple shrinkage scheme and, if one uses OSSEs (not actual geomagnetic data), to check errors in unobserved components of the geodynamo. In an operational scheme, where one is restricted to observed quantities, experiments with longer range forecasts may turn out to be important.

4.3 Results: proxy on the sphere

The numerical experiments with the proxy on the square suggest that the shrinkage and climatological localization schemes are most promising and we now test these on the spherical proxy. We tune localization and inflation as before and consider an ensemble of size $N_e = 100$. The nature run is 301 DA cycles long and we compute average errors (OmF and TmF) over the last 50 cycles to account for spin-up. Table 6 lists the errors (OmF and TmF) of the tuned EnKFs. We note that we tried EnKF without localization or inflation with ensemble size $N_e = 1000$, but this configuration still leads to macroscopic error. This means that the

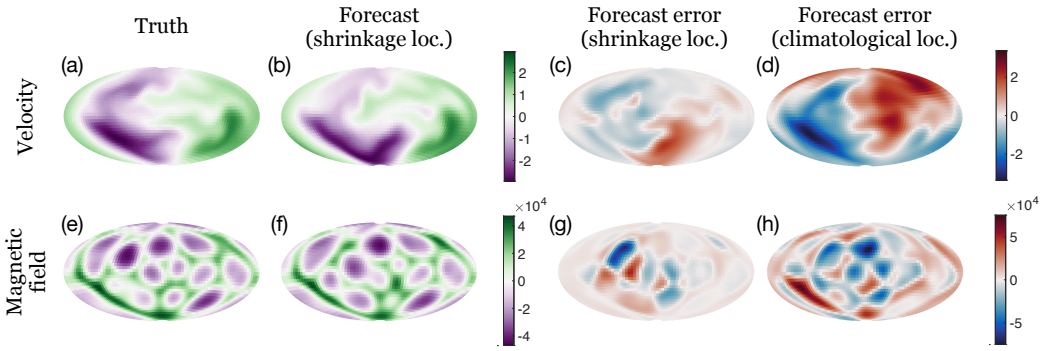

Figure 11. A comparison of forecast skill when using initial conditions estimated using the shrinkage and climatological localization schemes. (a) true state of A, (b) forecast of A from shrinkage localization (c) error in forecast of A from shrinkage localization, (d) error in forecast of A from climatological localization, (e) true state of ω, (f) forecast of ω from shrinkage localization (g) error in forecast of ω from shrinkage localization, (h) error in forecast of ω from climatological localization.

required ensemble size for reliable DA *without* localization or inflation is larger than 1000 (but our experiments are not sufficient for determining what the minimum required ensemble size really is, we only found a lower bound). The fact that the localized and inflated EnKFs lead to drastically reduced errors thus, again, speaks for the relevance of localization and inflation to keep the computational budget in check.

As observed in the experiments with the proxy on the square, we find that both localization schemes are effective and reduce the errors by more than two orders of magnitude and that ensemble spread is comparable to the errors (suggesting a well-tuned DA system). Overall, we find the smallest errors in the shrinkage scheme, which again is consistent with what we found for the proxy on the square. We note, however, that the errors in the unobserved velocity field is much larger for the climatological localization than for the shrinkage method. On the spherical proxy, shrinkage localization seems to have a clear-cut advantage over climatological localization. This provides further support to the notion that there exists a structure to correlations between modes, in short-term forecasts of the model, which is more complicated than indicated by the climatological statistics.

The advantages of shrinkage localization can further be highlighted by considering longer-range forecasts (see experiments with the proxy on the square). We initialize forecasts of four e-folding times with the ensemble mean of the two Kalman filters (with climatological and shrinkage localization). The results are illustrated in Figure 11. We see that the forecast error of the shrinkage scheme is much smaller than the forecast error of climatological localization. The reason is that climatological localization leads to larger errors in the unobserved state

Table 7. Assimilation cycles required to reduce forecast errors in the magnetic field (first row) and velocity field (second row) to 1% of the macroscopic error using shrinkage (second column) and climatological localization (third column).

	Shrinkage	Climatological
Magnetic field	137	170
Velocity	179	212

(the velocity), which are amplified relatively quickly by the dynamics. Indeed, on longer range forecasts, we note a clear reduction in forecast errors in observed *and* unobserved quantities. This corroborates the idea that longer range forecast errors can be used as a tool to diagnose the applicability of localization methods: An appropriate localization scheme should lead to smaller longer-range forecast errors.

Finally, we compare the spin-up times of the EnKFs using the shrinkage and climatological localization schemes on the sphere. Beginning with an uniformed, initial ensemble, we record the number of assimilations needed to reduce the error to 1% of the macroscopic error in each field. The results are shown in Table 7 where, as observed with the proxy on the square, the shrinkage scheme requires fewer assimilations to reduced errors. Thus, we again find that the choice of localization scheme not only determines the average forecast errors over a long DA run, it also impacts the the number of assimilations necessary to reduce initial errors to a desired level.

5 SUMMARY AND CONCLUSIONS

We designed a simplified proxy model to represent some of the numerical and computational challenges in geomagnetic DA. The proxy swaps the Navier-Stokes equation for a simpler, but chaotic flow (the Kuramoto-Sivashinsky equation), and couples the flow variables to a magnetic field via an induction equation and a Lorentz-like coupling term. The proxy model is computationally simple enough to enable systematic numerical studies and data assimilation experiments.

The strategy of using proxy models to rapidly prototype and test numerical DA schemes has proven useful and effective in NWP. Specifically, we designed the proxy to replicate the following challenges in geomagnetic DA:

(i) The proxy is a chaotic dynamic system, consisting of two coupled PDEs for magnetic and velocity fields;
(ii) the proxy is amenable to spectral discretization;

(iii) magnetic and velocity fields are coupled, but only the spectrum of the magnetic field is partially observed;
(iv) the proxy model is dynamically rich enough to require localization and inflation for efficient numerical DA.

Due to the simplifications, results obtained with the proxy model are qualitative, not quantitative and one must carefully evaluate how results obtained with the proxy model may extend to realistic, or real, geomagnetic DA systems.

The proxy model has a 2D geometry to keep computations simple and we present versions of the proxy on square and spherical geometries. The proxy on the square makes use of FFT and, for that reason, is computationally simpler than the proxy on the sphere which requires spherical harmonic transforms. This suggests using the proxies hierarchically: The proxy on the square can be use to screen for possibly useful techniques, and one can then perform a selected set of experiments on the spherical proxy. Using the model hierarchy, we ran a large number of systematic DA experiments with the goal to investigate the role of localization and inflation within geomagnetic DA. The numerical experiments suggest the following key findings.

(a) Localization and inflation are invaluable for efficient geomagnetic DA because localized DA can keep the errors small, even if the ensemble size is also small;
(b) localization and inflation can reduce the spin-up period, i.e., the number of DA cycles required to reach a steady and low error level;
(c) a simple shrinkage localization scheme may be more appropriate than popular climatological localization methods, because the correlation structure relevant for DA may be very different from the climatological correlations;
(d) many localization schemes may produce similar errors in observed quantities on short time scales, but errors in the unobserved quantities can be quite different for the different schemes;
(e) performing longer range forecasts may prove to be a useful tool for determining how appropriate a localization scheme really is.

A natural next step for continuing this work, and to determine the actual usefulness of the proxy, is to test some of our hypothesis in a more realistic or real geomagnetic DA system.

ACKNOWLEDGEMENTS

KG acknowledges that this work was supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program - Grant “80NSSC18K1351”. MM and KG are

supported by the US Office of Naval Research (ONR) grant N00014-21-1-2309. AT and WK are supported by NASA Earth Surface and Interior Program.

All authors contributed to the ideas behind the approach taken in the manuscript with KG taking the lead; all authors contributed to writing the paper, with KG taking the lead and producing the first draft; KG wrote the code and designed and performed the numerical experiments.

DATA AVAILABILITY

The code for the model on the square (implemented in Matlab) is available at github link will be inserted for publication (curated version). The code for the model on the sphere cannot be made publicly available at this time as it relies on a spherical harmonic transform owned by a third party and used with their permission. All numerical results, and the code used for the model on the square in this paper, are at a Zenodo link will be inserted for publication.

REFERENCES

Ades, M. & van Leeuwen, P. J., 2015. The effect of the equivalent-weights particle filter on dynamical balance in a primitive equation model, *Mon. Weather Rev.*, **143**(2), 581–596.

Anderson, J. L., 2012. Localization and sampling error correction in ensemble Kalman filter data assimilation, *Mon. Weather Rev.*, **140**, 2359–2371.

Anderson, J. L., 2016. Reducing correlation sampling error in ensemble Kalman filter data assimilation, *Mon. Weather Rev.*, **144**, 913–925.

Anderson, J. L. & Anderson, S. L., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Mon. Weather Rev.*, **127**(12), 2741 – 2758.

Armbruster, D., Guckenheimer, J., & Holmes, P., 1989. Kuramoto-Sivashinsky dynamics on the center-unstable manifold, *SIAM J Appl Math.*, **49**(3), 676–691.

Bärenzung, J., Wicht, M. H. J., Lesur, V., & Sanchez, S., 2020. The Kalmag model as a candidate for IGRF-13, *Earth, Planets and Space*, **72**(163).

Barrois, O., Hammer, M. D., Finlay, C. C., Martin, Y., & Gillet, N., 2018. Assimilation of ground and satellite magnetic measurements: inference of core surface magnetic and velocity field changes, *Geophys. J. Int.*, **215**(1), 695–712.

Bauer, P., Thorpe, A., & Brunet, G., 2015. The quiet revolution of numerical weather prediction, *Nature*, **252**, 45–55.

Bonavita, M., Isaksen, L., & Hólm, E., 2012. On the use of EDA background error variances in the ECMWF 4D-Var, *Q. J. R. Meteorol. Soc.*., **138**(667), 1540–1559.

Browne, P. A., 2016. A comparison of the equivalent weights particle filter and the local ensemble transform Kalman filter in application to the barotropic vorticity equation, *Tellus A*, **68**, 30466.

Buehner, M., McTaggart-Cowan, R., & Heilliette, S., 2017. An ensemble Kalman filter for numerical weather prediction based on variational data assimilation: VarEnKF, *Mon. Weather Rev.*, **145**(2), 617 – 635.

Chorin, A. J. & Morzfeld, M., 2013. Conditions for successful data assimilation, *J. Geophys. Res. Atmos.*, **118**(20), 11,522–11,533.

Chorin, A. J., Morzfeld, M., & Tu, X., 2013. Implicit sampling, with application to data assimilation, *Chin. Ann. Math. Ser. B*, **34**, 89–98.

Courtier, P., 1997. Variational methods, *J. Meteorol. Soc. Japan*, **75**(1B), 211–218.

Cowling, T. G., 1933. The Magnetic Field of Sunspots, *Mon. Not. R. Astron. Soc.*, **94**(1), 39–48.

Cox, S. & Matthews, P., 2002. Exponential time differencing for stiff systems, *J. Comput. Phys.*, **176**, 430–455.

Errico, R. M., Yang, R., Priv, N. C., Tai, K.-S., Todling, R., Sienkiewicz, M. E., & Guo, J., 2013. Development and validation of observing-system simulation experiments at NASA's global modeling and assimilation office, *Q. J. R. Meteorol. Soc.*., **139**(674), 1162–1178.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res. Oceans*, **99**(10), 10143–10162.

Evensen, G., 2006. *Data assimilation: the ensemble Kalman filter*, Springer.

Fornberg, B., 1996. *A Practical Guide to Pseudospectral Methods*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press.

Fournier, A., Eymin, C., & Alboussière, T., 2007. A case for variational geomagnetic data assimilation: insights from a one-dimensional, nonlinear, and sparsely observed MHD system, *Nonlinear Process. Geophys.*, **14**(2), 163–180.

Fournier, A., Nerger, L., & Aubert, J., 2013. An ensemble Kalman filter for the time-dependent analysis of the geomagnetic field, *Geochem. Geophys. Geosyst.*, **14**, 4035–4043.

Gharamti, M. E., Raeder, K., Anderson, J., & Wang, X., 2019. Comparing adaptive prior and posterior inflation for ensemble filters using an atmospheric general circulation model, *Mon. Weather Rev.*, **147**(7), 2535 – 2553.

Hamill, T. M., Whitaker, J. S., Anderson, J. L., & Snyder, C., 2009. Comments on "Sigma-Point Kalman Filter Data Assimilation Methods for Strongly Nonlinear Systems", *J. Atmos. Sci.*, **66**(11), 3498–3500.

Harty, T., Morzfeld, M., & Snyder, C., 2021. Eigenvector-spatial localisation, *Tellus A*, **73**(1), 1–18.

Hoffman, R. N. & Atlas, R., 2016. Future observing system simulation experiments, *Bull. Am. Meteorol. Soc.*, **97**(9), 1601 – 1616.

Holm, H. H., Sætra, M. L., & van Leeuwen, P. J., 2020. Massively parallel implicit equal-weights particle filter for ocean drift trajectory forecasting, *J. Comput. Phys.*, **6**, 100053.

Hooper, A. P. & Grimshaw, R., 1985. Nonlinear instability at the interface between two viscous fluids, *Phys. Fluids*, **28**(1), 37–45.

Hulot, G., Lhuillier, F., & Aubert, J., 2010. Earth's dynamo limit of predictability, *Geophys. Res. Lett.*, **37**, L06305.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D*, **230**(1), 112–126, Data Assimilation.

Jardak, M., Navon, I. M., & Zupanski, M., 2010. Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation, *Int. J. Numer. Methods Fluids*, **62**(4), 374–402.

Kalogirou, A., Keaveny, E. E., & Papageorgiou, D. T., 2015. An in-depth numerical study of the two-dimensional Kuramoto-Sivashinsky equation, *Proc. R. Soc. A.*, **471**.

Kotsuki, S., Ota, Y., & Miyoshi, T., 2017. Adaptive covariance relaxation methods for ensemble data assimilation: experiments in the real atmosphere, *Q. J. R. Meteorol. Soc.*., **143**(705), 2001–2015.

Kuang, W. & Bloxham, J., 1999. Numerical modeling of magnetohydrodynamic convection in a rapidly rotating spherical shell: Weak and strong field dynamo action, *J. Comput. Phys.*, **153**(1), 51–81.

Kuramoto, Y. & Tsuzuki, T., 1975. On the formation of dissipative structures in reaction-diffusion systems, *Prog. Theor. Phys.*, **54**(3), 687–699.

Langel, R. A. & Estes, R. H., 1982. A geomagnetic field spectrum, *Geophys. Res. Lett.*, **9**(4), 250–253.

Li, H., Kalnay, E., & Miyoshi, T., 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*., **135**(639), 523–533.

Li, K., Jackson, A., & Livermore, P. W., 2011. Variational data assimilation for the initial-value dynamo problem, *Phys. Rev. E*, **84**, 056321.

Li, K., Jackson, A., & Livermore, P. W., 2014. Variational data assimilation for a forced, inertia-free magnetohydrodynamic dynamo model, *Geophys. J. Int.*, **199**, 1662–1676.

Lorenz, E. N., 1963. Deterministic nonperiodic flow, *J. Atmos. Sci.*, **20**(2), 130 – 141.

Lorenz, E. N., 1995. Predictability: a problem partly solved., *ECMWF Seminar Proceedings on Predictability*, **1**, 118.

Lunderman, S., Morzfeld, M., & Posselt, D. J., 2021. Using global Bayesian optimization in ensemble data assimilation: parameter estimation, tuning localization and inflation, or all of the above, *Tellus A*.

Minami, T., Nakano, S., Lesur, V., Takahashi, F., Matsushima, M., Shimizu, H., Nakashima, R., Taniguchi, H., & Toh, H., 2020. A candidate secular variation model for IGRF-13 based on MHD dynamo simulation and 4DEnVar data assimilation, *Earth, Planets Space*, **72**(136).

Morzfeld, M., Hodyss, D., & Snyder, C., 2017. What the collapse of the ensemble Kalman filter tells us about particle filters, *Tellus A*, **69**(1), 1283809.

Morzfeld, M., Adams, J., Lunderman, S., & Orozco, R., 2018. Feature-based data assimilation in geophysics, *Nonlinear Process. Geophys.*, **25**(2), 355–374.

Papageorgiou, D., Maldarelli, C., & Rumschitzki, D., 1990. Nonlinear interfacial stability of core-annular film flows, *Phys. Fluids A*, **2**(3), 340–352.

Ropp, G., Lesur, V., Bärenzung, J., & Holschneider, M., 2020. Sequential modelling of the earth?s core magnetic field, *Earth, Planets and Space*, **72**(153).

Sabaka, T. J., Tøffner-Clausen, L., Olsen, N., & Finlay, C. C., 2020. CM6: a comprehensive geomagnetic field model derived from both CHAMP and Swarm satellite observations, *Earth, Planets and Space*, **72**(80).

Sanchez, S., Fournier, A., Aubert, J., Cosme, E., & Gallet, Y., 2016. Modelling the archaeomagnetic field under spatial constraints from dynamo simulations: a resolution analysis, *Geophys. J. Int.*, **207**(2), 983–1002.

Sanchez, S., Wicht, J., Bärenzung, J., & Holschneider, M., 2019. Sequential assimilation of geomagnetic observations: perspectives for the reconstruction and prediction of core dynamics, *Geophys. J. Int.*, **217**(2), 1434–1450.

Sanchez, S., Wicht, J., & Bärenzung, J., 2020. Predictions of the geomagnetic secular variation based on the ensemble sequential assimilation of geomagnetic field models by dynamo simulations, *Earth Planets Space*, **72**(157).

Shlyaeva, A., Whitaker, J., & Snyder, C., 2019. Model-space localization in serial ensemble filters, *J. Adv. Model. Earth Syst.*, **11**(6), 1627–1636.

Sivashinsky, G., 1977. Nonlinear analysis of hydrodynamic instability in laminar flames-I. derivation of basic equations, *Acta Astronaut.*, **4**, 1177–1206.

Sun, Z. & Kuang, W., 2015. An ensemble algorithm based component for geomagnetic data assimilation, *Terr. Atmospheric Ocean. Sci.*, **26**.

Sun, Z., Tangborn, A., & Kuang, W., 2007. Data assimilation in a sparsely observed one-dimensional modeled MHD system, *Nonlinear Process. Geophys.*, **14**, 181–192.

Tangborn, A. & Kuang, W., 2015. Geodynamo model and error parameter estimation using geomagnetic data assimilation, *Geophys. J. Int.*, **200**(1), 664–675.

Tangborn, A. & Kuang, W., 2018. Impact of archeomagnetic field model data on modern era geomagnetic forecasts, *Phys. Earth Planet. Inter.*, **276**, 2 – 9, Special Issue:15th SEDI conference.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker", J. S., 2003. Ensemble square root filters, *Mon. Weather Rev.*, **131**, 1485–1490.

Touloumis, A., 2015. Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings, *Computational Statistics & Data Analysis*, **83**, 251–261.

Zeng, X., Atlas, R., Birk, R. J., Carr, F. H., Carrier, M. J., Cucurull, L., Hooke, W. H., Kalnay, E., Murtugudde, R., Posselt, D. J., Russell, J. L., Tyndall, D. P., Weller, R. A., & Zhang, F., 2020. Use of observing system simulation experiments in the United States, *Bull. Am. Meteorol. Soc.*, **101**(8), E1427 – E1438.

Zhang, M. & Zhang, F., 2012. E4DVar: Coupling an ensemble Kalman filter with four-dimensional

variational data assimilation in a limited-area weather prediction model, *Mon. Weather Rev.*, **140**(2), 587 – 600.

Zhen, Y. & Zhang, F., 2014. A probabilistic approach to adaptive covariance localization for serial ensemble square root filters, *Mon. Weather Rev.*, **142**(12), 4499 – 4518.