Ensemble Kalman filter for the reconstruction of the Earth's mantle circulation

Marie Bocher[1], Alexandre Fournier[2], and Nicolas Coltice[1]

[1]Laboratoire de Géologie de Lyon, Université Claude Bernard Lyon 1, Ecole Normale Suprieure de Lyon, France.
[2]Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot,CNRS (UMR 7154), Paris, France

Correspondence to: M. Bocher (marie.bocher@ens-lyon.fr)

Abstract. Recent advances in mantle convection modelling led to the release of a new generation of convection codes, able to generate self-consistently plate-like tectonics at their surface. Those models physically link mantle dynamics to surface tectonics. Combined with plate tectonic reconstructions, they have the potential to produce a new generation of mantle cir-

5 culation models that use data assimilation methods and where uncertainties on plate tectonic reconstructions are taken into account. We recently provided a proof of this concept by applying a suboptimal Kalman Filter to the reconstruction of mantle circulation (Bocher et al., 2016). Here, we propose to go one step further and apply the ensemble Kalman filter (EnKF) to this problem. The EnKF is a sequential Monte Carlo method particularly adapted to solve high dimensional data assimilation problems with nonlinear dynamics. We tested the EnKF using synthetic observations consisting of surface velocity and heat flow

10 measurements, on a 2D-spherical annulus model and compared it with the method developed previously. The EnKF performs on average better and is more stable than the former method. Less than 300 ensemble members are sufficient to reconstruct an evolution. We use covariance adaptive inflation and localization to correct for sampling errors. We show that the EnKF results are robust over a wide range of covariance localization parameters. The reconstruction is associated with an estimation of the error, and provides valuable information on where the reconstruction is to be trusted or not.

15 1 Introduction

Mantle circulation models are estimates of mantle flow history. They combine two sources of information: observations on the dynamics or 3D structure of the Earth's mantle and a numerical model of mantle convection. In their effort to reconcile both observations and our physical understanding of mantle dynamics, they serve a wide variety of purposes and disciplines. Hager and O'Connell (1979) originally built instantaneous mantle circulation models to understand the effect of plates on large-scale

20 mantle flow. Since then, they have been used, among other applications, to understand the dynamics and evolution of the deep earth mantle structures (Bunge et al., 1998; McNamara and Zhong, 2005; Bower et al., 2013; Davies et al., 2012), to study the evolution of mantle plumes and their relationship to hotspots (Hassan et al., 2016), to infer changes in the Earth's rotation axis (Steinberger and O'Connell, 1997), sea-level (Moucha et al., 2008) or dynamic topography (Flament et al., 2013).

The geodynamics community has developed three alternative approaches to the problem of the reconstruction of mantle

25 circulation. The first approach, backward advection, consists in starting at present by estimating the current density field of the

mantle from seismic tomography models (see Conrad and Gurnis, 2003, for a description of this method). This density field is then advected backward in time with plate tectonic reconstructions as imposed boundary condition (Steinberger and O'Connell, 1997). This method has a limited numerical cost and exploits the two most instructive constraints on mantle circulation: plate tectonic reconstructions and seismic tomography. However, this technique neglects thermal diffusion, so it is not able to

5 reconstruct past thermal structures that have completely diffused before present and it is limited to times and regions for which the effect of diffusion is thought to be small. This limits reconstructions to the last 50 to 75 Myr (Conrad and Gurnis, 2003) or even to shorter periods if we consider the uncertainties on tomographic models (Bello et al., 2014). The second approach, the semi-empirical sequential method, estimates mantle circulation by integrating plate tectonics reconstructions chronologically into a mantle convection model. Plate tectonic reconstructions are either introduced as velocity boundary conditions, as first

10 described by Bunge et al. (1998), or with a more sophisticated method, by blending a convection solution with thermal and kinematic models of plates and slabs (Bower et al., 2015). This approach allows the use of models of convection with chemical heterogeneities (McNamara and Zhong, 2005). Also, it is not anymore the reconstruction method that limits the timespan of the reconstruction, but the availability of plate tectonic reconstructions. This led to mantle circulation models integrating up to 450 Myr of plate reconstruction history (Zhang et al., 2010). However, this method considers plate tectonic reconstructions as

15 perfect estimates of surface tectonics: uncertainties affecting the reconstructions are not taken into account although they are substantial, especially as reconstructions go further in the past (for example, there is almost no information on the state of the ocean floor before 140 Myr, see e.g. Torsvik et al., 2010). This method also requires the choice of an arbitrary initial temperature field to compute the evolution. The third approach uses data assimilation methods to solve the mantle circulation problem. Data assimilation methods are inverse methods dealing with the specific problem of estimating the evolution of a dynamical system

20 from asynchronous data and a physical model (Evensen, 2009a). The full inverse problem for mantle circulation, as stated by Bunge et al. (2003), would take into account model errors, numerical approximations, errors on plate reconstructions and on the estimation of the current tomography-derived temperature field to provide the best fit given all sources of information. However, solving the full inverse problem of mantle circulation is still a great challenge given the nonlinearities in mantle convection dynamics and the computational power required to compute a realistic forward mantle convection evolution alone

25 (Stadler et al., 2010; Burstedde et al., 2013). So far, variational data assimilation dominates over other methods to estimate mantle circulation (Bunge et al., 2003; Horbach et al., 2014; Ghelichkhan and Bunge, 2016). To simplify the problem, they minimize the misfit between the final temperature field of the mantle circulation model and the one deduced from seismic tomography. These mantle circulation models impose plate tectonic reconstructions as boundary conditions, as in the first two approaches.

30 Here, we take a different view on data assimilation methods for mantle circulation models by focusing on how to take into account the uncertainties in plate tectonic reconstructions. For almost a decade, 3-D spherical mantle convection models have shown the capability to self-consistently produce plate-like tectonics at their surface (Walzer and Hendel, 2008; Van Heck and Tackley, 2008; Yoshida, 2008; Foley and Becker, 2009). These models physically link surface tectonics comparable to that of the Earth to mantle convection processes (Coltice et al., 2012; Rolf et al., 2014; Mallard et al., 2016). In Bocher et al. (2016), we

35 took advantage of this link to build a sequential data assimilation algorithm able to integrate plate reconstructions into a mantle

convection code while taking into account the uncertainties on those plate tectonic reconstructions. This technique assimilates a time series of surface observations chronologically, by repeating two stages, analysis and forecast, until all observations are taken into account. Whenever an observation is available, the analysis evaluates the most likely state of the mantle at this time, considering a prior guess (supplied by the forecast) and the new observations at hand. For this evaluation, we used the classical

5 best linear unbiased estimate (Talagrand, 1997). Then, the forward model of mantle convection forecasts the evolution of the mantle until the next observation time. We tested this algorithm on synthetic experiments. It proved to be efficient in recovering mantle circulation given constraints on the amplitude of errors affecting observations and the timespan between observations. Here we extend this work by applying a more advanced sequential data assimilation method, the ensemble Kalman filter (EnKF, described in Evensen, 1994; Burgers et al., 1998). This method is particularly suited for high dimensional nonlinear

10 dynamical models (Evensen, 2009b). Instead of estimating the most likely state of the mantle, the Ensemble Kalman filter provides at each time an approximation of the probability density function of the state of the system in the form of a finite ensemble of states. During the forecast stage, each member of the ensemble evolves independently. For the analysis, we use the second order statistics of the ensemble to correct each ensemble member with the new observations at hand. We evaluate this method with synthetic experiments in 2D-spherical annulus geometry (Hernlund and Tackley, 2008) and compare it to the

15 algorithm developed by Bocher et al. (2016). The EnKF provides more accurate estimations than the former method, and is even able to reconstruct evolutions that the former method could not. Moreover, the EnKF also estimates locally the error on the reconstruction. The optimal size of the ensemble for our test case is 300 members. Both covariance inflation and localization eliminate spurious correlations arising from the finite size of the Ensemble that is used to compute them.

This paper is organized as follows. In section 2, we present our simplifications on the general mantle circulation reconstruc-
20 tion problem and the correspondence with the notation in the EnKF algorithm. Then, in section 3, we detail the EnKF method and justify the variants chosen for the application to mantle circulation. Section 4 presents the results obtained on synthetic experiments and compares them to results obtained by the method described in Bocher et al. (2016). Section 5 is a discussion on the choice of the method and the challenges involved in the application of such a method to a realistic setting.

2   Presentation of the problem

25 We aim at reconstructing mantle circulation for the last hundreds of millions of years by combining a mantle convection model with plate tectonic reconstructions, using an ensemble Kalman filter. To study the behavior of the Ensemble Kalman filter on such problem, we consider a simplified mantle convection model. This section describes the model used to compute a mantle evolution, the data set assimilated in this evolution, and finally the backbone of ensemble Kalman filtering.

2.1   Mantle convection model

30 At the timescales and lengthscales we are interested in ($\geq$ 10 kyr, $\geq$ 1000 km), the mantle can be modelled as a continuous visco-plastic medium. To compute mantle circulation, we solve the equations of conservation of mass (Eq. (1) below), momentum (Eq. (2) below) and energy (Eq. (8) below) for an isochemical mantle under the Boussinesq approximation. The system

of equations is non-dimensionalized to the thermal diffusion time scale (see Ricard, 2015). Given the high Prandtl number of the mantle (of the order of 10^{24}), inertia is neglected. With these assumptions, the equations of conservation of mass and momentum become diagnostic equations of the form

$$\nabla \cdot \boldsymbol{u} = 0, \tag{1}$$

$$\nabla \cdot \boldsymbol{\sigma} - \nabla p + \mathrm{Ra} T \boldsymbol{e}_r = 0, \tag{2}$$

where $\boldsymbol{\sigma}$, $\boldsymbol{u}$, p, and T are the non-dimensional deviatoric stress, velocity, dynamic pressure, and temperature, respectively. The equations are written in spherical coordinates (r, θ, ϕ), using the physical convention with r the radius, θ the colatitude and ϕ the longitude. The associated unit vectors are $(\boldsymbol{e}_r, \boldsymbol{e}_\theta, \boldsymbol{e}_\phi)$.

Ra is the Rayleigh number, defined as

$$\mathrm{Ra} = \frac{\rho_0 g_0 \alpha_0 \Delta T a^3}{\mu_0 \kappa_0}, \tag{3}$$

with ρ_0 the density for $T = 0$, g_0 the gravitational acceleration, α_0 the thermal expansivity, ΔT the temperature drop, a the depth of the layer, κ_0 the thermal diffusivity, μ_0 the dynamic viscosity of the system. The Rayleigh number in our model is 10^6. It is one or two orders of magnitude lower than that of the Earth, but high enough to ensure chaotic convection. The vertical velocities and shear-stress at the surface and the base of the model are set to zero.

The deformation response of mantle material to stress is implemented as a linear relationship linking the strain rate tensor $\dot{\boldsymbol{\epsilon}}$ to the deviatoric stress tensor $\boldsymbol{\sigma}$ as

$$\boldsymbol{\sigma} = 2\mu_{\mathrm{eff}} \dot{\boldsymbol{\epsilon}} = \mu_{\mathrm{eff}} \left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T \right). \tag{4}$$

The effective viscosity μ_{eff} takes into account both a viscous Newtonian behavior with a viscosity μ_n and a pseudo-plastic behavior with an equivalent "pseudo-plastic viscosity" μ_y,

$$\mu_{\mathrm{eff}} = \min(\mu_n, \mu_y). \tag{5}$$

The Newtonian viscosity μ_n follows an Arrhenius law

$$\mu_n = \mu_0 \exp\left(\frac{E_A}{T + T_1} \right) \tag{6}$$

with $\mu_0 = \exp\left(-\frac{E_A}{2T_1} \right)$, T_1 the temperature at which $\mu_n = 1$, and E_A the nondimensional activation energy. We implement the decrease of viscosity in the asthenosphere by reducing by a factor of 10 the viscosity μ_n when the temperature is above a solidus equation $T_s = T_{s_0} + \nabla_r T_s (r_a - r)$ with r_a the surface value of r. The implementation of a weak asthenosphere tends to favor plate-like behavior (Tackley, 2000; Richards et al., 2001).

The pseudo-plastic part of the effective viscosity μ_y is defined by

$$\mu_y = \frac{\sigma_{yield}}{2\dot{\epsilon}_{\mathrm{II}}}, \tag{7}$$

where $\dot{\epsilon}_{\textrm{II}}$ is the second invariant of the strain rate tensor and $\sigma_{yield} = \sigma_Y + (r_a - r)\nabla_r\sigma_Y$, with σ_Y and $\nabla_r\sigma_Y$ the yield stress at the surface and the depth-dependence of the yield stress, respectively.

The energy conservation equation is the only prognostic equation of the system

$$\frac{\mathrm{D}T}{\mathrm{D}t} = \nabla^2 T + \mathrm{R}_h. \tag{8}$$

R_h is the non-dimensional internal heating rate defined as

$$\mathrm{R}_h = \frac{\rho_0 D^2 H}{k_0 \Delta T} \tag{9}$$

with H the dimensional heating rate and k_0 the thermal conductivity. We set isothermal top and bottom boundaries with temperatures T_a and T_b, respectively. The models presented here have 10% basal heating and 90% internal heating.

These equations are solved using the finite volume, multigrid parallel code STAGYY (Tackley et al., 1993), on a spherical annulus grid. This geometry provides results closer to the spherical grid than cylindrical geometry (Hernlund and Tackley, 2008). In the following, the longitudinal coordinate of a point is ϕ_l, with $l \in \{1, 2, ..., L\}$ and its radial coordinate is r_m with $m \in \{1, 2, ..., M\}$, r varying from r_b to r_a.

Note that this paper focuses on the methodology of ensemble data assimilation for a convecting system similar to that of the Earth's mantle. Hence, we choose a rather simple model that can reproduce plate-like tectonics at the surface. We rely on simplifications such as 2D geometry, incompressible and isochemical mantle and a rheology which does not take into account the history of the material. Although some of the complexities we ignore may play a fundamental role in the reconstruction of the Earth's mantle evolution, we choose to focus in this manuscript on the data assimilation methodology. Moreover, we choose to keep the same parameters as the test case of Bocher et al. (2016) in order to allow direct comparison between the methods. Table 1 lists the chosen parameter values.

To ease the comparison with Earth's mantle convection, we rescale the nondimensional time in the evolution, t, by the transit time of the convective system. By definition, the transit time of the Earth's mantle is $t_t^E = a^E / v_{rms}^E$, with a^E the thickness of the mantle and v_{rms}^E the root mean square of surface velocities of the Earth, as estimated by plate tectonic reconstructions (Seton et al., 2012). We compute the same value for the model $t_t^m = a / v_{rms}^m$. The scaled time t^s is then $t^s = t \frac{t_t^E}{t_t^m}$.

2.2 Observations of mantle circulation

The state of the Earth's surface is the time integrated expression of mantle circulation. At a global scale, the main source of information for the last 100 Myr is the database of the localization and identification of magnetic anomalies on the seafloor, translated into maps of seafloor ages (Müller et al., 2008; Seton et al., 2014). This information is complemented with regional geological studies giving constraints on the timing and geometry of tectonic events as well as a synthesis of paleontological, structural geology, stratigraphical, magnetic anomalies, gravity data and seismic studies. In addition, paleomagnetic data provide constraints on the paleolatitude of continental blocks (Besse and Courtillot, 2002).

Plate tectonic reconstructions use the geometric theory of plate tectonics to integrate all these observations. The result is a time series of maps of seafloor ages, plate layout and kinematics. The continuously closed plate algorithm (Gurnis et al., 2012) produces plate tectonic reconstruction maps continuous in space and time (Seton et al., 2012; Müller et al., 2016).

Although we are aware that these plate tectonic reconstruction maps are in themselves models and not direct observations, we propose to develop an assimilation method that use them as data to assimilate in our mantle convection model. This solution is generally chosen in mantle circulation reconstructions (Bunge et al., 2002; Zhang et al., 2010; Bower et al., 2015), because it provides continuous surface boundary conditions in space and time for the period of reconstruction. One advantage of the technique we develop is that it is possible to consider errors on the data that is assimilated, another is that the reconstructions do not need to be known at all times and at all points on the surface. Hence it is possible, in principle, to design a data assimilation scheme using direct observations. However, this would require further developments both on the database design and on the data assimilation algorithm. Sequential data assimilation methods for mantle circulation are still in their infancy, so we opt for a simpler structure of the data to be assimilated: a time series of maps of surface velocity and seafloor age, as given by plate tectonic reconstructions.

In this study, we limit ourselves to the test of data assimilation in synthetic experiments. In the model described in Sect. 2.1, the absence of small scale convection at the base of the boundary layer makes the surface heat flux an excellent proxy for the age of the seafloor (Coltice et al., 2012). Consequently, we consider surface heat flux and surface velocity as the data to assimilate.

To our knowledge, the amplitude of the uncertainty on global plate tectonic reconstructions has not yet been assessed. For the synthetic tests we perform in Sect. 4, we choose an arbitrary value of 10% of the root mean square value of heat flux and surface velocity, respectively. We further discuss this choice in Sect. 5.

2.3 Ensemble Kalman filtering framework: notations

Our aim is to assimilate a time series of observations (surface velocities and heat fluxes) into a mantle convection model to estimate the evolution of the state of the mantle. We introduce here the general formulation of ensemble Kalman filtering and link them to our problem. We use the notation system recommended by Ide et al. (1997).

The time series of data is defined as a set of column vectors $\{\boldsymbol{y}_1^o, \boldsymbol{y}_2^o, ..., \boldsymbol{y}_K^o\}$, where the subscripts $\{1, 2, ..., K\}$ refer to the times at which observations are available. As seen in the previous section, the data used for our experiments are surface velocity and surface heat flux. The data vector at time k is thus defined as

$$\boldsymbol{y}_k^o = \left[q_k^o(\phi_1), q_k^o(\phi_2), ..., q_k^o(\phi_L), u_{\phi k}^o(\phi_1), u_{\phi k}^o(\phi_2), ..., u_{\phi k}^o(\phi_L) \right]^T, \tag{10}$$

where $q_k^o(\phi_l)$ and $u_{\phi k}^o(\phi_l)$ are the observed values of surface heat flux and surface horizontal velocity at the k-th timestep and longitude ϕ_l, and $(\cdot)^T$ means transpose. We model errors on observations by a random vector of zero mean and covariance matrix $\mathbf{R}_k$ (we suppose unbiased observations). Although $\mathbf{R}_k$ is a diagonal matrix of constant value and size in our experiments, it is not generally the case. Correlations between errors on observations could be specified in $\mathbf{R}_k$.

The evolution of the state of the system is estimated sequentially during the period where observations are available. At each timestep $k \in \{1, 2, ..., K\}$, we define two state vectors: the a priori state, or forecast state $\boldsymbol{x}_k^f$ and the analysis state $\boldsymbol{x}_k^a$, which is the state corrected after having assimilated the observations $\boldsymbol{y}_k^o$. The system of equations developed in Sect. 2.1 shows that we can compute velocity, viscosity and pressure values at each grid point from the temperature field. Nevertheless, the relation

5   between surface velocities and the temperature field is nonlinear. We choose to include the whole temperature field and the surface velocities to form an augmented state vector. This simplifies the computations thereafter. The state of the mantle at a timestep $k \in [1, K]$ is defined as

$$\boldsymbol{x}_k = [T_k(\phi_1, r_1), T_k(\phi_1, r_2), ..., T_k(\phi_L, r_M), u_{\phi k}(\phi_1), u_{\phi k}(\phi_2), ..., u_{\phi k}(\phi_L)]^T, \tag{11}$$

where $T_k(\phi_l, r_m)$ and $u_{\phi k}(\phi_l)$ are the values of temperature at the kth timestep, longitude ϕ_l and radius r_m and surface

10   horizontal velocity at the kth timestep and longitude ϕ_l.

The forecast and analyzed states are uncertain as well. Their uncertainties are represented by two random vectors of zero expectancy and covariance matrices $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$, respectively. We do not compute explicitly these covariance matrices. Instead, we compute two ensembles of N states $\{\boldsymbol{x}_{kn}^f\}_{n \in [1,N]}$ and $\{\boldsymbol{x}_{kn}^a\}_{n \in [1,N]}$, such that their average equals $\boldsymbol{x}_k^f$ and $\boldsymbol{x}_k^a$, respectively, and their respective sample covariance matrices approximate $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$. The ensemble of states $\{\boldsymbol{x}_{kn}^f\}_{n \in [1,N]}$

15   and $\{\boldsymbol{x}_{kn}^a\}_{n \in [1,N]}$ are stored in the matrices $\mathbf{X}_k^f$ and $\mathbf{X}_k^a$, where the nth column is the state of the nth ensemble member $\boldsymbol{x}_{kn}^f$ and $\boldsymbol{x}_{kn}^a$, respectively.

Finally, we introduce the observation operator, which maps a given state vector $\boldsymbol{x}_{kn}^e$ (e being f or a) to the corresponding data $\boldsymbol{y}_{kn}^e$. If the surface heat flux is approximated by a first order discretization of Fourier's law, then the observation operator is linear, and can be represented by the matrix $\mathbf{H}$ such that

20   $$\forall k \in \{1, 2, ..., K\}, \forall n \in \{1, 2, ..., N\}, \quad \boldsymbol{y}_{kn}^e = \mathbf{H} \boldsymbol{x}_{kn}^e. \tag{12}$$

Table 2 summarizes the dimensions of the vectors and matrices for our problem.

3   Ensemble Kalman filter with localization and inflation

The ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998) is a sequential data assimilation algorithm using the same equations as the Kalman Filter for the analysis step, but Monte Carlo methods to forecast the error statistics on the state. We

25   explain here how we adapt the ensemble Kalman filter to our problem and justify the choice of the starting ensemble.

To implement the EnKF, we used the software environment Parallel Data Assimilation Framework (PDAF, Nerger et al., 2005; Nerger and Hiller, 2013).

3.1   Initialization: first analysis and generation of the starting ensemble

As in Bocher et al. (2016), we compute second order statistics from a series of 400 decorrelated snapshots of convection simula-

30   tions. We obtain the first forecast state of average $\boldsymbol{x}_1^f$ and associated covariance matrix $\mathbf{P}_1^f$. The background covariance matrix

$\mathbf{P}_1^f$ is eigendecomposed and rank reduced into $\mathbf{P}_{1r}^f = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, with $\mathbf{\Lambda}$ a diagonal matrix containing the $n_r = 1928$ largest eigenvalues of $\mathbf{P}_1^f$ (which accounts for 99.98% of its cumulative variance) and $\mathbf{V}$ a matrix of the corresponding eigenvectors.

The first set of observations $\boldsymbol{y}_1^o$ is assimilated to obtain

$$\boldsymbol{x}_1^a = \boldsymbol{x}_1^f + \mathbf{V}\mathbf{A}\mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}(\boldsymbol{y}_1^o - \mathbf{H}\boldsymbol{x}_1^f), \tag{13}$$

5 $$\mathbf{P}_1^a = \mathbf{V}\mathbf{A}\mathbf{V}^T, \tag{14}$$

with

$$\mathbf{A} = \left[\mathbf{\Lambda}^{-1} + \mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{V}\right]^{-1}. \tag{15}$$

We generate an ensemble of N initial states using the second order exact sampling method (Hoteit, 2001; Pham, 2001). First, A is eigendecomposed

10 $$\mathbf{A} = \mathbf{V}^a\mathbf{\Lambda}^a\mathbf{V}^{aT}. \tag{16}$$

The ensemble members are then computed following

$$\mathbf{X}_1^a = \begin{pmatrix} | & & | \\ \boldsymbol{x}_{11}^a & ... & \boldsymbol{x}_{1N}^a \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \boldsymbol{x}_1^a & ... & \boldsymbol{x}_1^a \\ | & & | \end{pmatrix} + \sqrt{N-1}\mathbf{V}\mathbf{V}^a\mathbf{\Lambda}^{a1/2}\begin{pmatrix} \mathbf{\Omega}_{N\times(N-1)}^T \\ \mathbf{0}_{(n_r-N)\times N} \end{pmatrix}, \tag{17}$$

where $\mathbf{\Omega}$ is a random matrix whose columns are vectors forming an orthonormal basis and each of them is orthogonal to $\mathbf{1} = [1,...,1]^T$. $\mathbf{\Omega}$ is generated through the algorithm described in the appendix of Nerger et al. (2012). The matrix $\mathbf{\Omega}$ is 15 designed so that the sample mean of the starting ensemble is equal to $\boldsymbol{x}_1^a$ and its sample covariance matrix is equal to matrix $\mathbf{P}_1^a$ reduced to its N largest eigenvalues.

This method of generating the starting ensemble takes advantage of the extensive knowledge we have on the background statistics of the model. Several other methods have been tested to generate a starting ensemble, such as starting with random decorrelated snapshots of mantle convection obtained from a very long run, second order exact sampling from $\boldsymbol{x}_1^f$ and $\mathbf{P}_1^f$, and 20 several assimilations of the first observations $\boldsymbol{y}_1^o$. None of these solutions were as efficient for our problem as the technique used here.

3.2 Forecast

Between timesteps $k-1$ and k, the forward numerical code STAGYY computes independently the evolution of each of the analyzed states $\{\boldsymbol{x}_{k-1,n}^a\}_{n\in[1,N]}$ to produce a forecast ensemble $\{\boldsymbol{x}_{k,n}^f\}_{n\in[1,N]}$.

25 The forecast state is the average of the ensemble

$$\boldsymbol{x}_k^f = \frac{1}{N}\mathbf{X}_k^f\mathbf{1}, \tag{18}$$

and the forecast error covariance matrix is given by the sample covariance matrix of the ensemble of forecast states

$$\mathbf{P}_k^f = \frac{1}{N-1} \mathbf{X}_k^f \left(\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) \left(\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right)^T \mathbf{X}_k^{fT}. \tag{19}$$

After several assimilation cycles, the finite size of the ensemble induces the underestimation of the error variance (van Leeuwen, 1999), and can lead to filter divergence. We observed this behavior in our case, and to stabilize the filter we apply covariance

5   inflation, as suggested in Anderson and Anderson (1999) and Hamill et al. (2001).

We correct the forecast ensemble variance with an inflation factor γ according to

$$\mathbf{X}_k^f \leftarrow \frac{1}{N} \mathbf{X}_k^f \mathbf{1}\mathbf{1}^T + \left[\mathbf{X}_k^f \left(\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) \right] \sqrt{\gamma}, \tag{20}$$

where $\leftarrow$ means that we replace the matrix on the left-hand side by the term on the right-hand side. γ is computed following the same principles as in the suboptimal Kalman Filter developed in Bocher et al. (2016), i.e. by comparing the error on

10   observations and the standard deviation of the innovation $\boldsymbol{d}_k$ defined as

$$\boldsymbol{d}_k = \boldsymbol{y}_k^o - \frac{1}{N} \mathbf{H} \mathbf{X}_k^f \mathbf{1}. \tag{21}$$

The inflation factor is

$$\gamma = \frac{V^d - V^o}{V^f}, \tag{22}$$

with

15   $$V^d = \mathrm{Tr} \left(\boldsymbol{d}_k \boldsymbol{d}_k^T \right), \tag{23}$$

$$V^o = \mathrm{Tr}(\mathbf{R}_k), \tag{24}$$

$$V^f = \mathrm{Tr} \left[\mathbf{H} \mathbf{X}_k^f \left(\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) \left(\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right)^T \mathbf{X}_k^{fT} \mathbf{H}^T \right], \tag{25}$$

where $\mathrm{Tr}(\cdot)$ means the trace. The inflation factor is then truncated between a minimum value of 1 (to prevent further contraction of the ensemble spread) and a maximum value of $\gamma^+ = 1.25$ (to prevent overspread). Several values of maximum inflation

20   factor have been tested, from $\gamma^+ = 1.1$ to $\gamma^+ = 2$, and showed little impact on the efficiency of the assimilation. A constant inflation factor was also tested, but the results with an adaptive inflation factor were substantially more accurate, especially for the first assimilation times.

3.3   Analysis

The analyzed state $\boldsymbol{x}_{kn}^a$ of the nth member of the ensemble is

25   $$\boldsymbol{x}_{kn}^a = \boldsymbol{x}_{kn}^f + \mathbf{K}_k \left(\boldsymbol{y}_{kn}^o - \mathbf{H} \boldsymbol{x}_{kn}^f \right) \tag{26}$$

where $\mathbf{K}_k$ is the Kalman Gain. $\boldsymbol{y}_{kn}^o$ is the observed data vector $\boldsymbol{y}_k^o$ to which a random perturbation of zero expectation and covariance matrix $\mathbf{R}_k$ is added, as is recommended in Burgers et al. (1998).

The Kalman Gain is defined as

$$\mathbf{K}_k = (\mathbf{P}_k^f \circ \mathbf{C})\mathbf{H}^T \left[\mathbf{H}(\mathbf{P}_k^f \circ \mathbf{C})\mathbf{H}^T + \mathbf{R}_k \right]^{-1}, \tag{27}$$

where the matrix $\mathbf{P}_k^f$ is the sample covariance matrix of the ensemble of forecast states $\{\boldsymbol{x}_{kn}^f\}_{n \in [1,N]}$. We use a limited ensemble size (maximum 768) to estimate $\mathbf{P}_k^f$. Spurious correlations ensue, especially between distant points. To counteract

5 this effect, we implement direct forecast error localization by Schur multiplying (symbol $\circ$) $\mathbf{P}_k^f$ by the localization matrix $\mathbf{C}$, as introduced by Hamill et al. (2001) and Houtekamer and Mitchell (2001). The matrix $\mathbf{C}$ is itself the Schur product of a vertical localization matrix $\mathbf{C}_v$ and a horizontal localization matrix $\mathbf{C}_h$. The value of $\mathbf{C}_v(i,j)$ depends on the absolute radius difference of the $i-th$ and the $j-th$ components of the state vector and on vertical correlation length ℓ_v. The value of $\mathbf{C}_h(i,j)$ depends on the absolute angle difference of the $i-th$ and the $j-th$ components of the state vector and on vertical correlation length

10 ℓ_h. Both values follow a Gaspari-Cohn compactly supported fifth-order piecewise rational function (similar to a Gaussian but with a compact support, Eq. (4.10) of Gaspari and Cohn, 1999).

We also tested the domain localization strategy as described in Janjic et al. (2011), since it is computationally more efficient and already implemented in PDAF. However, it led to a systematic failure of the assimilation. This is due to the nature of our problem: all the observations are located at the surface of the model and we aim at estimating the temperature field over the

15 whole depth of the mantle. A vertical localization is as necessary as a horizontal localization: the localization has to be done in the state space and not only in the data space.

3.4 Implementation of the ensemble Kalman filter

We used the software environment PDAF (Nerger et al., 2005; Nerger and Hiller, 2013) in combination with the mantle convection code STAGYY (Tackley, 2008) to develop an ensemble Kalman filter code for mantle convection . PDAF provides

20 a set of core routines computing in parallel the analysis steps for a range of ensemble based data assimilation techniques. It provides as well a set of standard routines to adapt the parallelization of a preexisting parallel forward numerical model and integrate the data assimilation routines. The final product is a highly scalable ensemble data assimilation code running both forecasts and analyses in parallel.

We modified the STAGYY code following the procedure recommended by PDAF (see the online documentation wiki at

25 Nerger, 2016). We also made a few modifications in PDAF routines to allow for direct forecast error localization with the Ensemble Kalman filter. Additionally, we designed a basic observation database so as to load in a single step all the observations used in the data assimilation procedure.

4 A posteriori evaluation of the ensemble Kalman filter method

We test the data assimilation scheme on twin experiments using the model described in Sect. 2.1. Throughout this section,

30 we compare the results of the ensemble Kalman filter for mantle circulation reconstructions to the results computed using the method developed in Bocher et al. (2016), hereafter referred to as method 1.

After describing the setup used for twin experiments, we test the robustness of the EnKF method and compare it to that of method 1. Then, we determine the range of data assimilation parameters which are suitable to conduct an ensemble data assimilation. Finally, we assess the ability of the scheme to actually reconstruct specific geodynamic structures.

4.1 Twin experiment setup

5 Twin experiments are a way to assess the accuracy of a data assimilation procedure in a controlled environment, where the true evolution is perfectly known.

First, we compute a reference evolution using the forward numerical model, considered as the true evolution, from which we extract the set of true states $\{\boldsymbol{x}_k^t\}_{k\in[1,K]}$. Here, the timespan of the evolution is 150 Myr and we sample true states every 10 Myr. From these states, we compute a time series of surface heat fluxes and surface velocities, following Eq. (12). We noise

10 these observations with a random Gaussian noise of standard deviation 10% of the root mean square of surface heat flux q_{rms} and surface velocities v_{rms}, to obtain the time series of observations to assimilate $\{\boldsymbol{y}_k^o\}_{k\in[1,K]}$. It follows that the observation error covariance matrix $\mathbf{R}$ is diagonal and does not change with time.

Then, we perform ensemble data assimilation for the data set $\{\boldsymbol{y}_k^o\}_{k\in[1,K]}$, with the observation error covariance matrix $\mathbf{R}$. We did not consider any model error in the filter we describe, so the parameters of the model used in the data assimilation

15 realizations are the same as those of the reference model.

We present here tests with different assimilation parameters, varying the number of members N, the vertical correlation length ℓ_v and the horizontal correlation angle ℓ_h. Table 3 details the range of parameters tested.

We compute four different evolutions to test the accuracy of the ensemble Kalman filter for different cases. Figure 3 shows the initial and final states of these evolutions, together with the result of global error evolution, and will be discussed in the

20 next section.

4.2 Robustness of the assimilation algorithm

The evolution of the global error on the estimated temperature field over the time period $\{1,...,K\}$ is

$$\left[\epsilon_T^f(1),\epsilon_T^a(1),\epsilon_T^f(2),...,\epsilon_T^f(K),\epsilon_T^a(K)\right] \tag{28}$$

where $\epsilon_T^e(k)$, e standing for a (analysis) or f (forecast) is

25 $$\epsilon_T^e(k)=\sqrt{\frac{\sum\limits_{l=1}^{L}\sum\limits_{m=1}^{M}\left(\overline{T}_k^e(\phi_l,r_m)-T_k^t(\phi_l,r_m)\right)^2\mathcal{V}(\phi_l,r_m)}{\sum\limits_{l=1}^{L}\sum\limits_{m=1}^{M}T_k^t(\phi_l,r_m)^2\mathcal{V}(\phi_l,r_m)}} \tag{29}$$

with $\mathcal{V}(\phi_l,r_m)$ the volume of the grid cell at longitude ϕ_l and radius r_m, and $\overline{T}_k^e(\phi_l,r_m)$ the average temperature of the estimated ensemble (either forecast or analysis) at longitude ϕ_l and radius r_m.

We test the EnKF on one evolution, with sizes of the ensemble $N=96$, 288 and 768 and for each combination of the following data assimilation parameters values: vertical correlation length $\ell_v=0.3$, 0.5, 0.7 and 1 and horizontal correlation
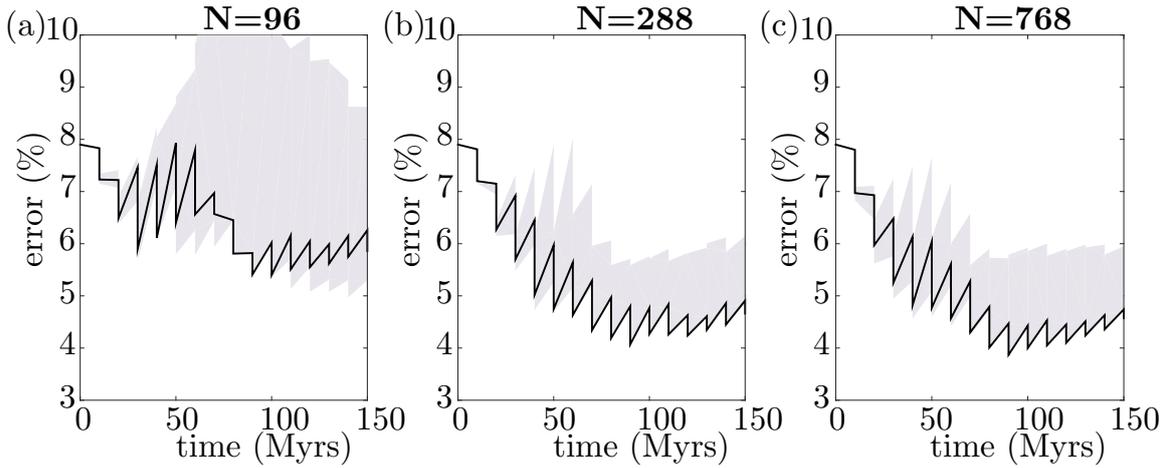
Figure 1. Time evolution of the errors on the estimated temperature field obtained from data asssimilations with the same 150 Myr dataset, but different assimilation parameters. The size of the ensemble is A) $N = 96$, B) $N = 288$ and C) $N = 768$. The assimilations are computed for any combination of data assimilation parameters: $\gamma^+ = 1.25$, $\ell_v = 0.3$, 0.5, 0.7 and 1 and $\ell_h = \pi/10$, $\pi/8$, $\pi/6$, $\pi/4$ and $\pi/2$. The black line represent the evolution of the error for the best assimilation A)$N = 96$, $\ell_z = 0.5$, $\ell_h = \pi/6$ and $\gamma^+ = 1.25$ B)$N = 288$, $\ell_z = 0.7$, $\ell_h = \pi/10$ and $\gamma^+ = 1.25$ and C)$N = 768$, $\ell_z = 0.5$, $\ell_h = \pi/4$ and $\gamma^+ = 1.25$. The grey area is delimited by the maximum and minimum values of errors at each time, for all data assimilations.

angle $\ell_h = \pi/10$, $\pi/8$, $\pi/6$, $\pi/4$ and $\pi/2$. We show in Fig. 1, for each ensemble size, the maximum and minimum values of errors obtained for all these parameters, as a function of time.

To determine the best assimilation, we compute the average error after analysis

$$\overline{\epsilon}_T^a = \frac{1}{K} \sum_{k=1}^{K} \epsilon_T^a(k). \tag{30}$$

5   For each ensemble size, the error evolution of the best assimilation (in the sense of minimum $\overline{\epsilon}_T^a$) is also shown in Fig. 1.

For any size of the ensemble, it is possible to find a set of parameters leading to a stabilization of the global error on the temperature field after a few analyses. The time after which the solution is stabilized corresponds to the transit time of the physical model (70 Myr). We can see that for $N = 288$ and $N = 768$, any combination of vertical and horizontal correlation lengths leads to errors lower than the first analysis. Although the error is decreasing through time for any combination of data

10   assimilation parameters, the difference between the maximum and the minimum errors obtained is greater than 1%, which is large given that the first analysis error is already below 8%. The best error evolutions for $N = 288$ and $N = 768$ are very similar, with a minimum error of 4.07% and 3.87% after 90 Myr, and an average global error after analysis of 5.01% and

4.85%, respectively. During the assimilation of a dataset, most of the computational time is dedicated to the forecast step, so the data assimilation with 768 members is 2.7 times longer than the assimilation with 288 members, on the account of the embarrassingly parallel nature of the forecast phase. Since we obtain very similar results for $N = 288$ and $N = 768$, we favor the assimilation with 288 members.

5   We compute the error on the estimated temperature from the true temperature field. However, in a realistic case, the true temperature is not known, and the evaluation of the data assimilation algorithm is based on the study of the statistics of the innovation vector $\boldsymbol{d}_k$ at analysis k

$$\boldsymbol{d}_k = \boldsymbol{y}_k^o - \mathbf{H}\boldsymbol{x}_k^f. \tag{31}$$

At each analysis time, we compute the Euclidean norm of the instantaneous innovation d_k and the Euclidean norm of the
10 cumulative mean innovation $\overline{\boldsymbol{d}}_k$

$$d_k = \| \boldsymbol{d}_k \| \qquad \text{and} \qquad \overline{\boldsymbol{d}}_k = \left\| \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{d}_i \right\| \tag{32}$$

Before computing these norms, we normalize the part of the innovation corresponding to surface heat flux and velocities by their respective root mean square values.

    The norm of the instantaneous innovation d_k measures the distance between the forecast data and the observation, and
15 therefore gives indications on the success or failure of the assimilation. The evolution of the cumulative mean of the innovation $\overline{\boldsymbol{d}}_k$ allows us to check the consistency of the data assimilation algorithm at first order. Indeed, the derivation of the EnKF equations assumes that the error on observations $\boldsymbol{y}^o$ and the error on the forecast data $\mathbf{H}\boldsymbol{x}^f$ are unbiased. Such hypotheses imply that the statistically expected value of $\boldsymbol{d}$ is zero, which means that the norm of the cumulative innovation should converge to zero as the number of analyses increases.

20   Figure 2 shows the evolution of d_k and $\overline{\boldsymbol{d}}_k$ as a function of the number of analyses for data assimilations with different sizes of ensemble and their respective optimum vertical and horizontal correlation lengths. The norm of the instantaneous innovation (Fig. 2(b)) first decreases during the first 8 analyses, i.e. 70 My, and then oscillates for the rest of the assimilation. We observe the same behavior for the evolution of the error on temperature $\epsilon_T^e(k)$: the norm of the instantaneous innovation for ensemble sizes of $N = 288$ and 768 are very similar, and lower than that obtained with $N = 96$. Fig. 2(a) shows the cumulative
25 innovation constantly decreasing throughout the assimilation, with comparable values for $N = 288$ and $N = 768$, and slightly higher values for $N = 96$.

    We also tested the assimilation algorithm for 4 different evolutions, with the optimal parameters for an ensemble size of $N = 288$ members ($\ell_v = 0.7$ and $\ell_h = \pi/10$). Figure 3 shows the initial and final temperature fields of the evolutions, together with the evolution of the global error, the spread of the ensemble, and the error evolution for the method 1.

30   The spread of the ensemble is an estimation of the error on the state. We compare the evolution of ϵ_T^e to the global standard deviation of the temperature field of the ensemble:

$$\left[\sigma_T^f(1), \sigma_T^a(1), \sigma_T^f(2), ..., \sigma_T^f(K), \sigma_T^a(K)\right] \tag{33}$$
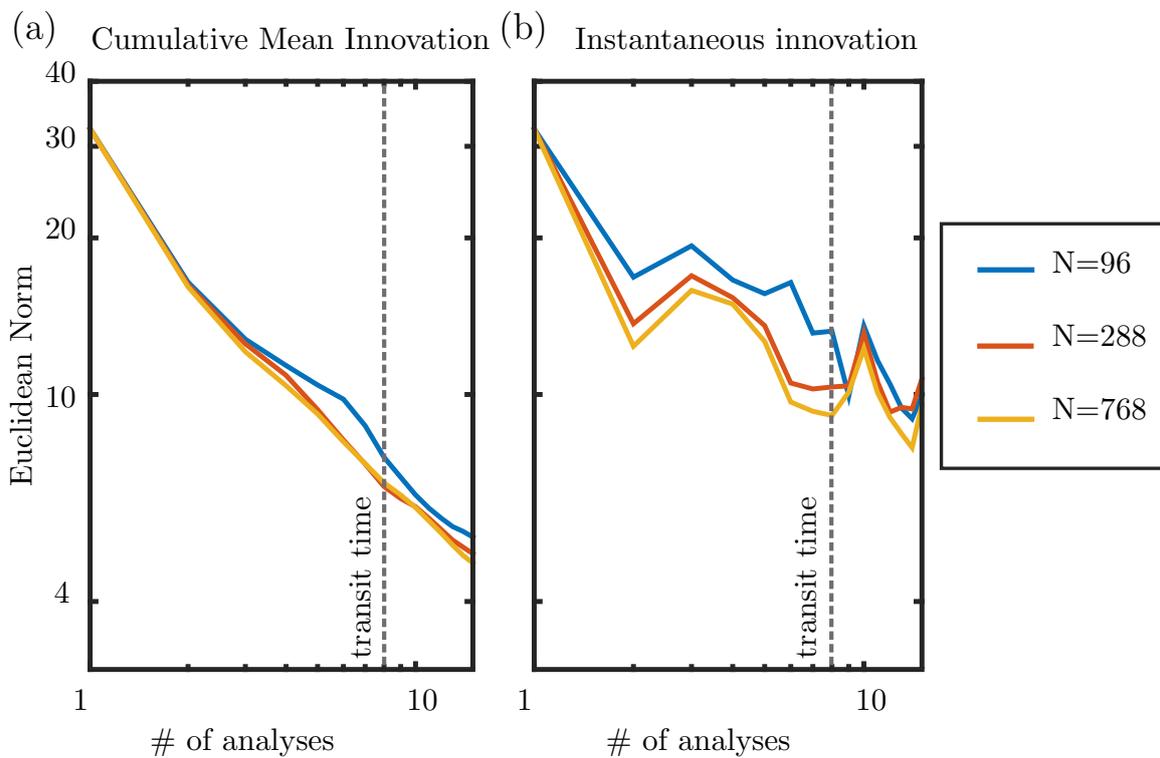
Figure 2. Evolution of (a) the cumulative mean innovation and (b) the norm of the instantaneous innovation, as a function of number of analyses and for different ensemble sizes. For each size of the ensemble, the evolutions correspond to the best combinations of correlation length parameters: $N = 96$, $\ell_z = 0.5$, $\ell_h = \pi/6$ and $\gamma^+ = 1.25$; $N = 288$, $\ell_z = 0.7$, $\ell_h = \pi/10$ and $\gamma^+ = 1.25$ and $N = 768$, $\ell_z = 0.5$, $\ell_h = \pi/4$ and $\gamma^+ = 1.25$.
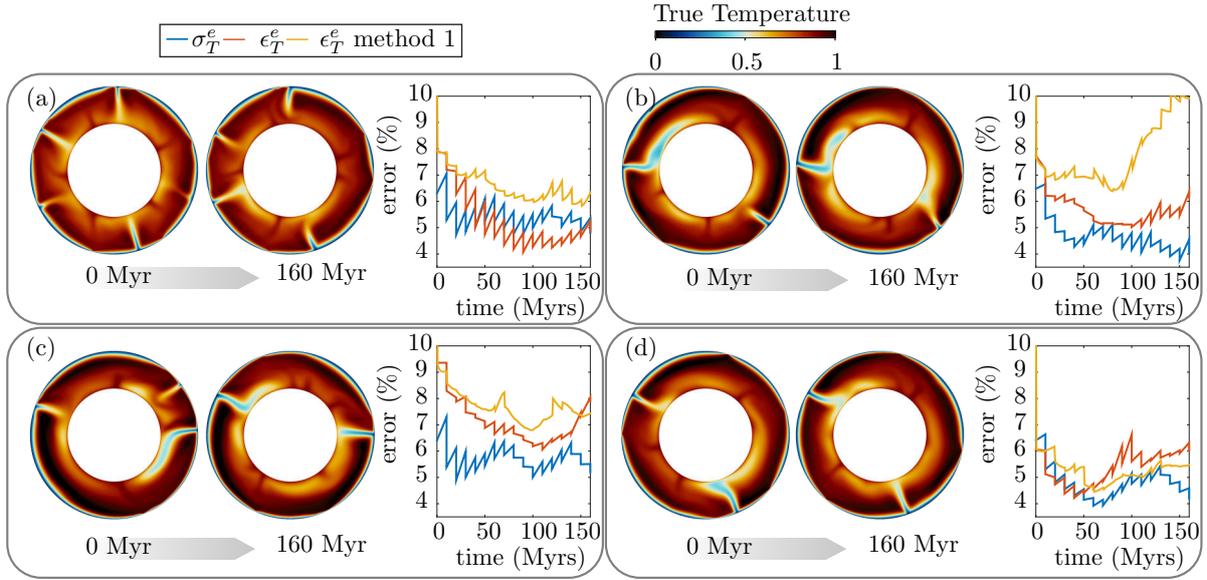
Figure 3. Evolution of the error (ϵ_T^e, red) as a function of time for 4 different evolutions with $N = 288$, $\gamma^+ = 1.25$, $\ell_v = 0.7$ and $\ell_h = \pi/10$, compared to the evolution of the spread of the ensemble (σ_T^e, blue) and the evolution of the error with the technique of Bocher et al. (2016) (ϵ_T^e method 1, yellow). The initial and final states of the true evolutions are represented on the left of each corresponding graph.

with $\sigma_T^e(k)$ defined as

$$
\sigma_T^e(k) = \sqrt{\frac{\sum\limits_{n=1}^{N}\sum\limits_{l=1}^{L}\sum\limits_{m=1}^{M}\left(T_{kn}^e(\phi_l, r_m) - \overline{T}_k^e(\phi_l, r_m)\right)^2 \mathcal{V}(\phi_l, r_m)}{(N-1)\sum\limits_{l=1}^{L}\sum\limits_{m=1}^{M}\overline{T}_k^e(\phi_l, r_m)^2 \mathcal{V}(\phi_l, r_m)}}. \tag{34}
$$

We compute the error for an estimated evolution with the method 1 using Eq. (29).

Although we ran the four evolutions with the same values of physical parameters (as described in Table 1), they show
5 different configurations: Evolution A has a shorter wavelength of convection, with the persistence of 4 subductions, 3 ridges and 5 upwellings, the death of one ridge and creation of two. Evolutions B, C and D have longer wavelengths of convection, with two major downwellings, stable throughout the evolutions. In evolution B, one of these downwellings has a very large negative temperature anomaly at the bottom of the domain. In evolution C, the remnant of a subduction merges with a larger subduction into a single downwelling.

10 In the 4 cases, the errors on the estimated temperature field systematically decrease during the analysis step for the EnKF algorithm. The errors stay below the first analysis error for evolutions A, B and C, while they reach slightly higher values for

evolution D. The error of the EnKF is always lower than that obtained with method 1 for the first 50 My. The average error is lower for the EnKF than for method 1 in 3 out of 4 cases. The average standard deviation of the ensemble (ensemble spread) is of the same order of magnitude as the true error. However, its evolution is not the same as the true error, with differences between both of more than 2% for some part of evolution C, for example. For evolutions C and D, the results of the two

5 methods are comparable whereas the assimilation with EnKF performs better than method 1 for evolutions A and B.

For evolution B, method 1 fails to reconstruct accurately the evolution, with the error reaching values greater than 10% at the end of the assimilation. This case is further investigated on Fig. 4 and 5. Figure 4 compares the true temperature field evolution with the analyzed temperature field of method 1 and of the Ensemble Kalman filter with $N = 288$, $\ell_v = 0.7$ and $\ell_h = \pi/10$. The sudden increase in the error of the estimated temperature field for method 1 seen on Fig. 3(b) happens after around 80 Myr

10 of assimilation, when the direction of bending at the bottom of the domain changes for the downwelling on the left side (see Fig. 4, second row). The analyzed temperature field of Method 1 does not predict this change of direction (see Fig. 4, first row), while the analyzed temperature field of the ENKF predicts it (see Fig. 4, third row). Method 1 computes only the evolution of the best estimate of the system. The computation of only one estimate ignores that, in this case, a slight perturbation of the estimated state could lead to a totally different dynamics. On the contrary, the EnKF method computes the evolution of

15 an ensemble of perturbed solutions and thus takes into account the nonlinearity of the solution, at least for the forecast stage. Figure 5 shows examples of the analyzed temperature fields of different ensemble members for evolution B, after 80 Myr of assimilation. Although the average temperature fields displays a downwelling bending to the right, the ensemble members show a wide variety of downwelling geometries.

4.3 Effect of the data assimilation parameters on the quality of the estimation

20 As shown in Fig. 1, the choice of N, ℓ_v and ℓ_h is critical to minimize errors in the assimilation, with errors on the estimated temperature field varying from 4% to more than 10% according to the choice of parameters. We investigate further the effect of these parameters by comparing the average global errors after analyses, $\bar{\epsilon}_T^a$, for different combinations of N, ℓ_v and ℓ_h. Figure 6 displays the values of $\bar{\epsilon}_T^a$ for sizes of ensemble $N = 96$, 288 and 768 (Fig. 6(a), (b) and (c) respectively) with ℓ_v varying between 0.3 and 1, and ℓ_h between $\pi/10$ and $\pi/2$. As in Fig. 1, we observe a dichotomy between assimilations with

25 $N = 96$ members, with higher errors, and assimilations with $N = 288$ and 768, with lower errors.

For each size of ensemble N we identify the pair (ℓ_v, ℓ_h) that leads to the assimilation with the lowest error $\bar{\epsilon}_{Tmin}^a(N)$. From this minimum value $\bar{\epsilon}_{Tmin}^a(N)$, we select all the pairs (ℓ_v, ℓ_h) that lead to data assimilation with global errors less than $\bar{\epsilon}_{Tmin}^a(N) + 0.2\%$. As the size of the ensemble increases, the optimal lengths of correlations (ℓ_v, ℓ_h) tend to increase. This is a classical effect (Houtekamer and Mitchell, 1998), observed in ensemble Kalman filters for various dynamical systems. As

30 N increases, the amplitude of noise in the sample correlation matrix $\mathbf{P}^f$ decreases, and small, yet real, correlations between distant points can be taken into account (Hamill et al., 2001). Between ensemble sizes of $N = 96$ and $N = 288$ the zone of optimal correlations is displaced towards the greater vertical correlation lengths. When we increase the size of the ensemble from $N = 288$ to $N = 768$, the zone of optimal correlations is displaced towards greater horizontal correlation angles. So the accurate estimation of correlations between points on the same vertical level needs less samples than between points on the
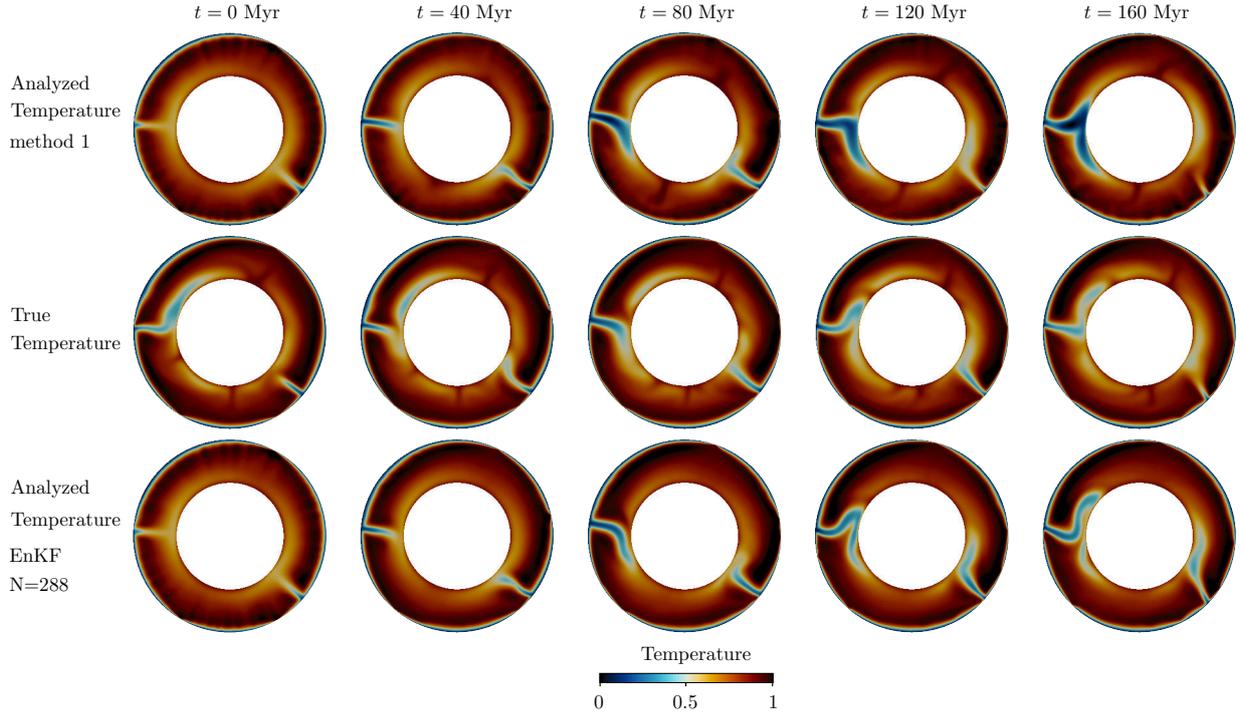
Figure 4. Comparison of temperature field evolutions for evolution B. The first row depicts the evolution of the analyzed temperature field with method 1. The second row is the true evolution of the temperature field. The third row is the evolution of the analyzed temperature field with ensemble Kalman filter, $N = 288$, $\ell_v = 0.7$ and $\ell_h = \pi/10$.

same horizontal level. This is due to the specifics of mantle convection dynamics. The highly nonlinear rheology produces plates at the surface with values of velocity and temperature that may vary substantially (by one or two orders of magnitude) on short distances in the horizontal direction, especially because of pseudoplasticity. On the contrary, highly viscous cold downwellings establish a strong continuity in the vertical direction.

5   For the ensemble size $N = 288$ and all the values of (ℓ_v, ℓ_h), we additionally evaluate the average global ensemble spread

$$\overline{\sigma}_T^a = \frac{1}{K} \sum_{k=1}^{K} \sigma_T^a(k), \tag{35}$$

the average forecast error on data

$$\overline{\boldsymbol{\epsilon}}_y^f = \frac{1}{K} \sum_{k=1}^{K} \frac{d_k}{\| \boldsymbol{y}_k^o \|} \tag{36}$$
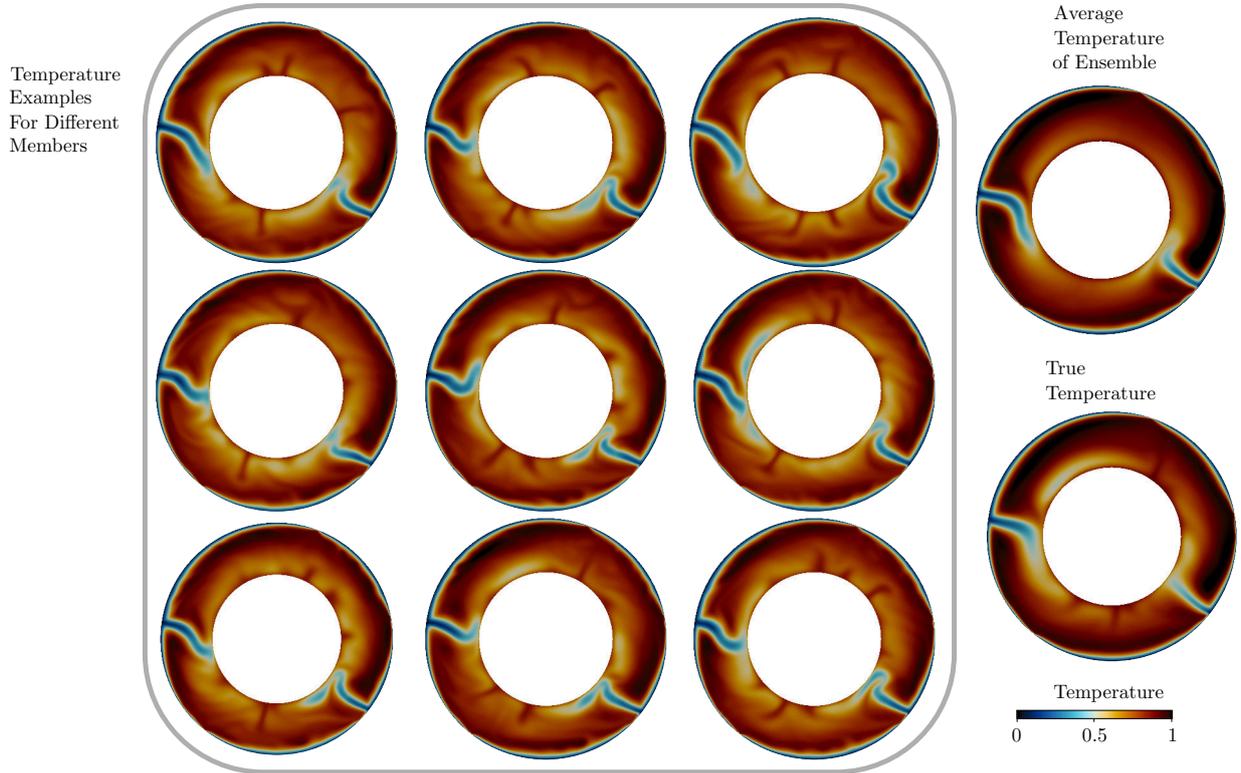
Figure 5. Example of temperature fields of the members of the ensemble. This example is taken after 80 Myr, for the assimilation of evolution B, with ENKF $N = 288$, $\ell_v = 0.7$ and $\ell_h = \pi/10$.

and the norm of the cumulative innovation for K analyses:

$$
\overline{\boldsymbol{d}}_K = \left\| \frac{1}{K} \sum_{k=1}^{K} \left(\boldsymbol{y}_k^o - \mathbf{H}\boldsymbol{x}_k^f \right) \right\|. \tag{37}
$$

These three values are indicators of the accuracy of the assimilation and can be computed in the case of an assimilation with Earth data, unlike $\overline{\epsilon}_T^a$.

5   Figure 7 represents these results along with the true error $\overline{\epsilon}_T^a$. The ensemble of optimal data assimilation parameters is also outlined ($\overline{\epsilon}_T^a < \overline{\epsilon}_{Tmin}^a(N) + 0.2\%$).

Overall, the average ensemble spread $\overline{\sigma}_T^a$ (Fig. 7(b)) decreases when ℓ_h and ℓ_v increase, with a minimum for $\ell_h = \pi/2$ and $\ell_v = 1$. The higher the correlation lengths, the more covariances will be taken into account in the analysis, and the analyzed members will be closer to each others and $\overline{\sigma}_T^a$ lower. The average ensemble spread $\overline{\sigma}_T^a$ is of the same order of magnitude as
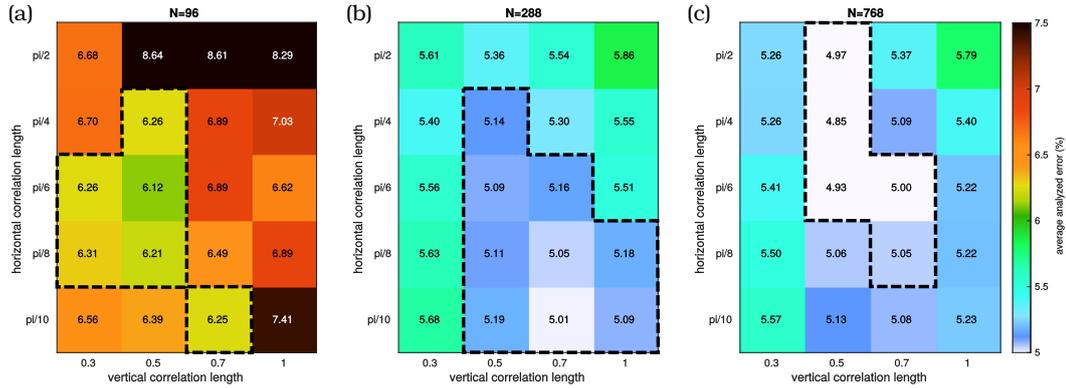
Figure 6. Value of the mean analyzed error for the same evolution and different vertical and horizontal correlation lengths. (a) for 96 ensemble members, (b) for 288 ensemble members, (c) 768 ensemble members. The dashed lines delimit the zones for which errors are less than $\overline{\epsilon}^a_{T min}(N) + 0.2\%$.

the true error $\overline{\epsilon}^a_T$. Moreover, there is a local minimum of $\overline{\sigma}^a_T$ at $\ell_v = 0.7$ and $\ell_h = \pi/10$. These parameters correspond to the minimum true error $\overline{\epsilon}^a_T$.

The average forecast errors and the norm of the cumulative innovations display the same behavior: they decrease with increasing vertical and horizontal correlation lengths. The longer the correlations lengths, the closer the forecast data are to the

5   observations, and the less biased the assimilation. This means that a better fit to the observations does not necessarily imply a better fit to the true temperature field. In a realistic context, the result of the assimilation should be checked against independent data to evaluate its accuracy. In the case of the Earth's mantle, independent data could be for example the geoid or tomographic models.

4.4   Accuracy of the reconstruction of geodynamic structures

10   We focus on three key flow structures: 1) downwelling slabs (subduction) 2) ridges, i.e. shallow structures resulting from divergent plates at the surface, 3) plumes, hot upwellings raising from the base of the model.

Figure 8 shows the final state of the assimilation after 150 My for the evolution A of Fig. 3. We selected 3 assimilations: EnKF96, an ensemble Kalman filter with $N = 96$, $\ell_v = 0.5$ and $\ell_h = \pi/6$ (first row), EnKF288 an EnKF with $N = 288$, $\ell_v = 0.7$ and $\ell_h = \pi/10$ (second row) and the assimilation with method 1 (third row). We do not show the ensemble Kalman

15   filter with 768 members since the resulting temperature field is almost indistinguishable from that of EnKF288. The first column represents the true temperature field, which is the same for all assimilations. The second column is the analyzed temperature field, i.e. the average of the temperature fields of the analyzed ensemble members. The third column is the absolute temperature
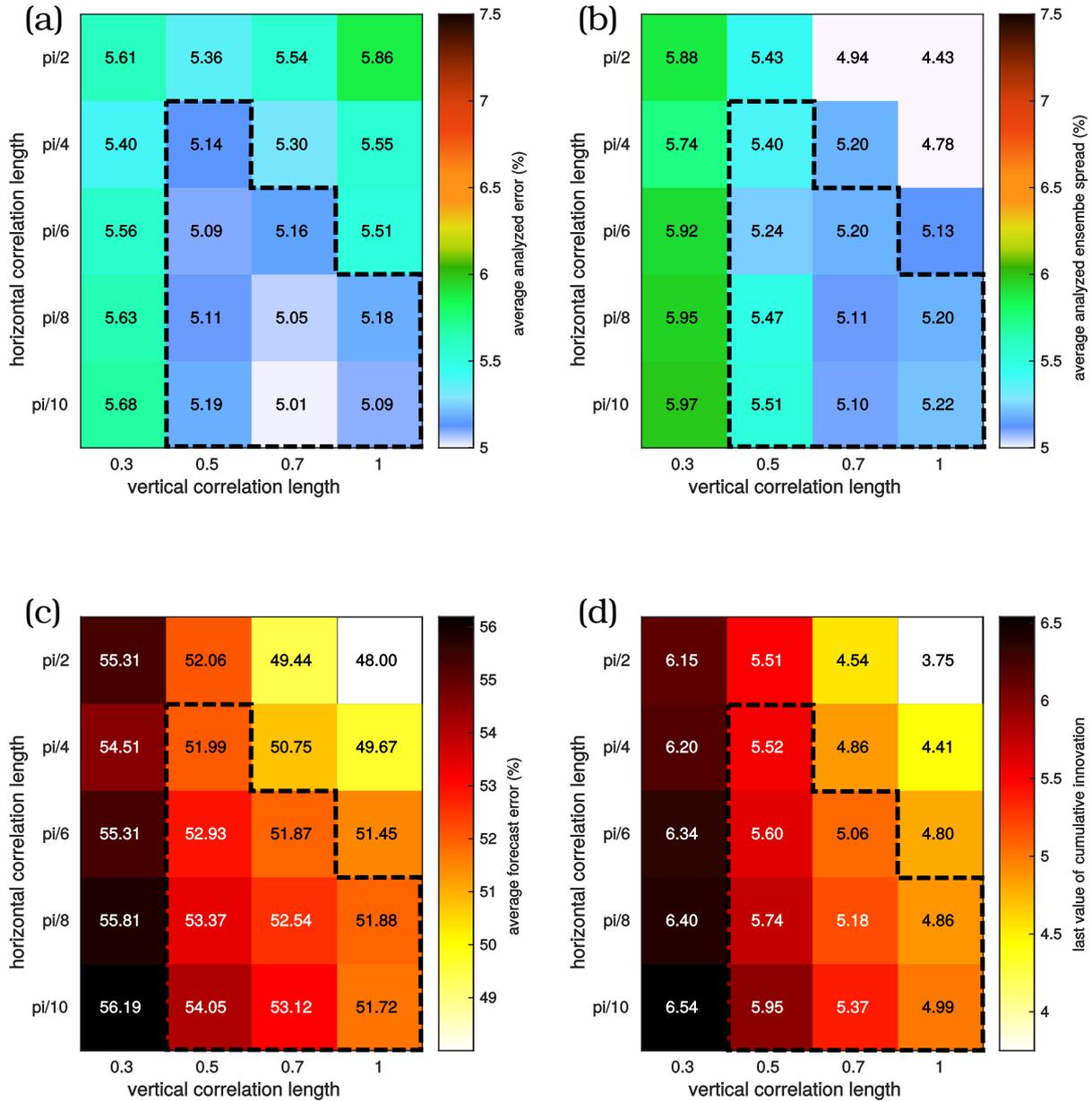
Figure 7. Value of (a) mean analyzed error, (b) mean ensemble spread, (c) average forecast error on data, (d) norm of cumulative innovation after K analyses for $N = 288$, and different vertical and horizontal correlation lengths. The dashed line delimits the zone for which errors are less than $\bar{\epsilon}^a_{Tmin}(288) + 0.2\%$.
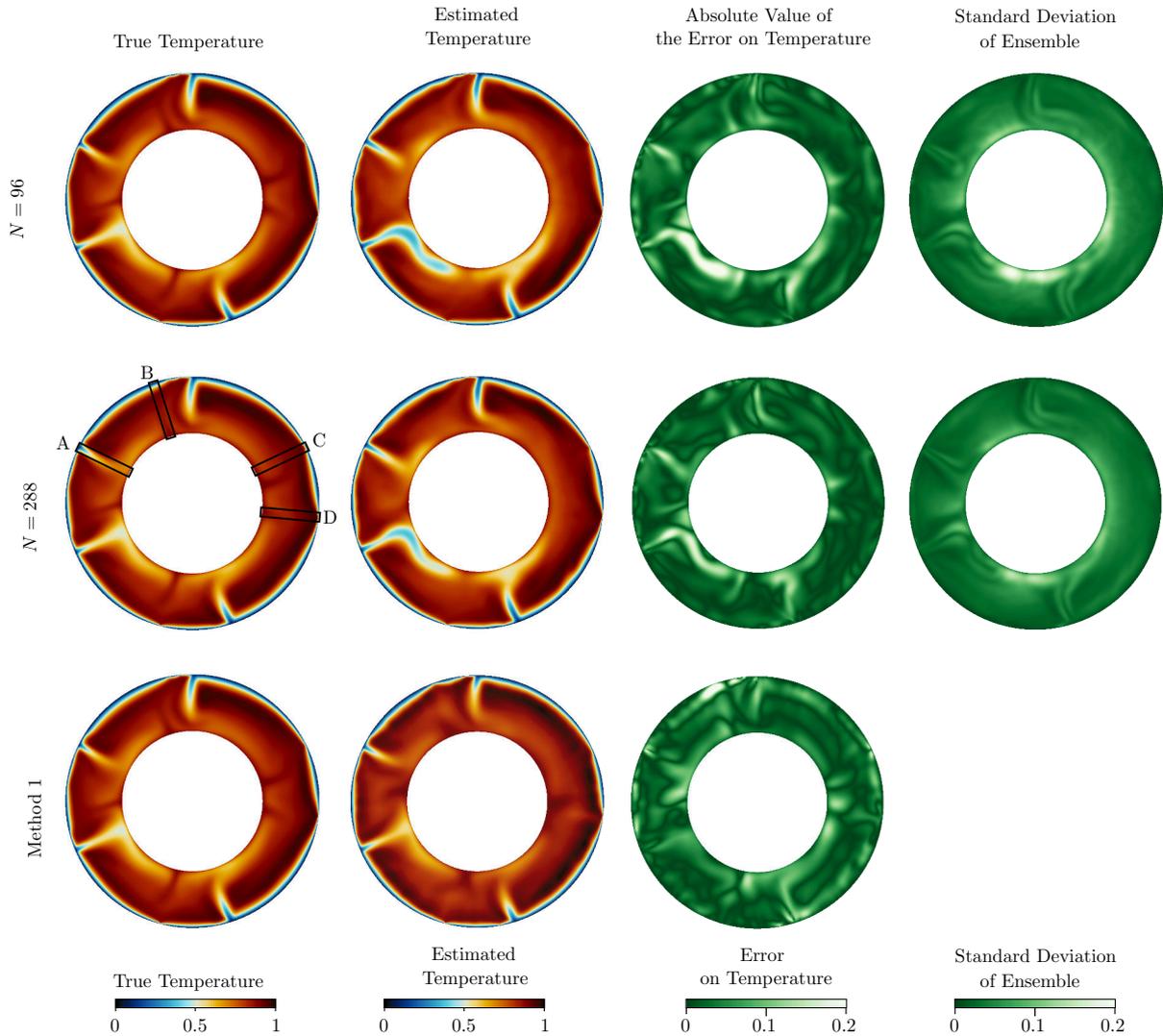
Figure 8. Comparison of estimated states after 150 Myr for the evolution A of Fig. 3. First row: ensemble Kalman filter with $N = 96$, $\ell_v = 0.5$ and $\ell_h = \pi/6$; second row: $N = 288$, $\ell_v = 0.7$ and $\ell_h = \pi/10$; third row for method 1 (third row). The first column represents the true temperature field at 150 Myr, the second column is the analyzed temperature field, the third is the absolute error on temperature value and the fourth is the estimated error on the analyzed field (spread of the ensemble). On the true temperature field of EnKF288 we framed the location of the subduction (A), plume (B), ridge initiation (C) and stable ridge (D) studied in Fig. 9.

error, and the fourth column is the standard deviation of the ensemble spread, which is an estimate of the error on the analyzed temperature field.

Globally, the EnKF288 and EnKF96 solutions for the temperature field are smoother than the solution of method 1. We observe this difference especially in the asthenosphere, the part of the mantle below the top boundary layer. For method 1, the asthenosphere shows short wavelength temperature variations. These variations are absent from the true temperature field and are inconsistent with convection solutions with the chosen parameters. They stem from the amplification of the noise in the

5 observations during the analysis. Moreover, the asthenosphere of the analyzed temperature field of method 1 is hotter than the true temperature.

Both EnKF96 and EnKF288 reconstruct successfully the ridges locations and structures, as testified by their error fields. On the contrary, method 1 fails to reconstruct the ridge on the top right of the domain. It also predicts a ridge that does not exist in the true state (in the top left quadrant). On the right of the domain, another ridge is associated with a vertical positive

10 temperature anomaly underneath. This pattern is found regularly under ridges when applying method 1. This is due to the use of a constant forecast error covariance matrix, $\mathbf{P}_0^f$ for the analysis. This constant matrix does not take into account the specifics of the dynamics under a ridge, where the positive anomaly is generally shallow. We do not observe this detrimental effect in the EnKF assimilations, where we compute the forecast error covariance matrix $\mathbf{P}_k^f$ at each analysis time from the forecast ensemble.

15 All three assimilations reconstruct the subductions and predict accurately the bending direction of slabs at the base of the model. Method 1 tends to underestimate the amplitude of the negative temperature anomalies whereas both EnKF assimilations overestimate them. This is especially noteworthy for the bottom left subduction. Moreover, the estimated slabs are wider than the true slabs. However, we note two arguments in favor of the EnKF: first, the estimation of the slab improves when the size of the ensemble increases and second, the local standard deviations of the ensemble indicates that the estimation in this part of

20 the domain is less accurate.

Both EnKF288 and EnKF96 solutions do not show any plume at the base of the mantle. However, the ensemble spread shows a greater uncertainty on the places where plumes occur. Method 1 predicts the approximate location of all plumes, but their geometry is not accurate. Method 1 provides only one estimate of the temperature field. In this evolution, the plumes are allowed to develop. EnKF96 and EnKF288 provide an ensemble of states. Each state develops plumes at different locations

25 and their averages show only a slightly hotter anomaly over a wide area of possible location for the plumes, as we showed earlier in Fig 5 for another assimilation.

To illustrate how different flow structures are reconstructed, we represent on Fig. 9 the evolution of temperature through time of the ensemble members of EnKF288 at points on the same vertical for a subduction, a plume, a ridge initiation and a stable ridge. Figure 8 shows the location of these geodynamical features on the true temperature field. We plot the temperature

30 evolutions at the surface, mid-mantle and at the bottom of the domain. Note that the surface and bottom values of temperature actually correspond to the values of the first points below the surface and above the bottom of the domain, respectively.

At the surface, the temperature is corrected accurately at each analysis, with a difference between the true temperature and the analyzed temperature of less than 0.01. The correction associated with the analysis gradually decreases with depth due to both covariance localization and the dynamics of the system. After 70 My (i.e. one transit time), the true value of temperature

35 falls within the range of values predicted by the ensemble for all geodynamical contexts and all depths.
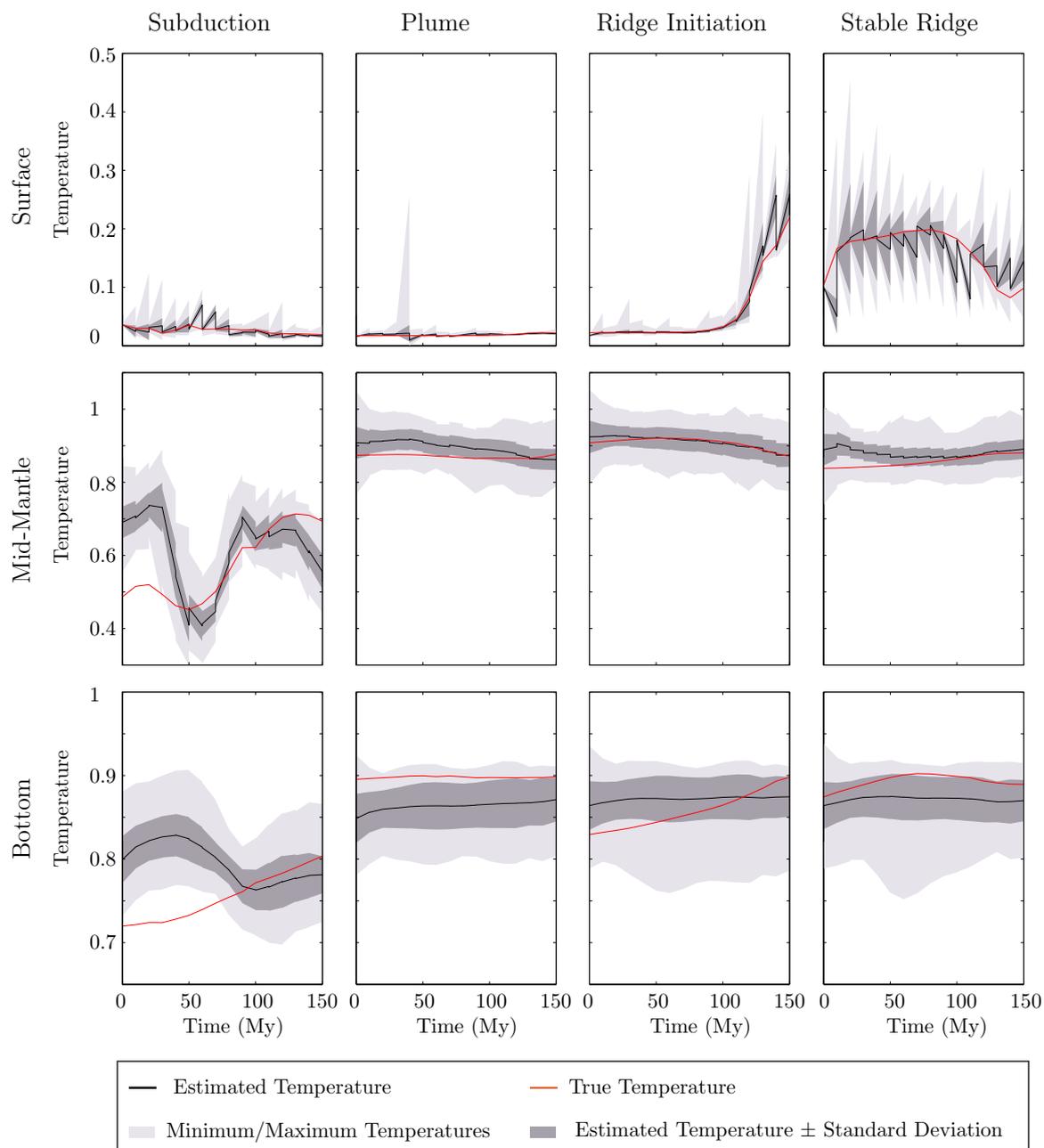
Figure 9. Evolution of the temperature value during assimilation at the surface (first row), in the mid mantle (second row) and at the bottom of the domain (third row), for a ridge initiation (first column), a subduction (second column), an upwelling (third column) and a stable ridge (fourth column).

For the subduction, the correction is first done on the surface, and then propagates gradually in depth. The reconstruction of mid-mantle temperature becomes accurate after 40 My, and at the bottom of the model after 70 My, which is the value of the transit time. At the surface, the spread of the ensemble decreases as more data are assimilated. On the contrary, the spread of the ensemble remains steady for mid-mantle depths and at the bottom of the domain. For these depths, only the average

5 temperature varies.

At the surface for the plume, the spread of the ensemble is very low except for a peak at 40 My, which corresponds to an instability, corrected after one analysis. We note that this instability affects greatly method 1 since it leads to the false prediction of the ridge seen in Fig. 8. At mid-mantle, the ensemble average is slowly converging to the true temperature. At the bottom, the estimated temperature is lower than the true temperature, although it slightly increases throughout the assimilation.

10 The ridge initiation shows how new observations affect the spread of the ensemble. At the surface, the spread of the ensemble remains low until 100 My, the time of initiation of the ridge. From then on, the estimated temperature increases and the ensemble members follow the cycle of increasing spread during forecast and dramatic decrease of spread during analysis. The temperature in the mid-mantle is estimated with a very good accuracy after 50 My. On the contrary, the assimilation does not predict the evolution of the temperature at the bottom of the domain, although the true temperature falls within the zone defined

15 by the standard deviation of the ensemble after 50 My.

For the stable ridge, the spread of the ensemble at the surface is increasing during forecast and decreasing dramatically during the analysis. At mid-mantle, the estimated temperature becomes accurate after 100 My. At the bottom of the domain the temperature is underestimated although it follows the variations of the true temperature: increase of temperature at the beginning of the assimilation and slight decrease at the end of the assimilation.

20 ## 5   Discussion

We chose the ensemble Kalman filter method for its ease of implementation and flexibility to adapt to different forward numerical models. Indeed, as long as the nature of the state and observations does not change, the computation of the analysis step remains the same regardless of the convection code used. On the contrary, the alternative method, variational data assimilation, requires the development of an adjoint code that needs further development for each additional complexity added to

25 the forward model (see Kalnay et al., 2007, for a comparison of EnKF and 4D variational methods). For the mantle circulation problem, this results in a series of derivation of the adjoint considering different approximations (Ismail-Zadeh et al., 2003; Bunge et al., 2003; Ghelichkhan and Bunge, 2016; Worthen et al., 2014). The ability of a data assimilation scheme to adapt to different numerical codes is a particularly important issue for mantle convection since models are in constant evolution, with current developments including the implementation of chemistry, nonlinear rheologies, elasticity, phase transition and

30 compressibility (see e.g. Zhong et al., 2015, for a review of recent developments of mantle convection codes). In particular, this ease of implementation allows us to work on models producing self-consistently plate-like tectonics at their surface, and hence to obtain forecasts whose data can be ultimately compared with plate reconstructions.

The application of the ensemble Kalman filter to the mantle circulation problem is the continuation of the simpler sequential filter that we developed in an earlier work (Bocher et al., 2016). The main difference between the two filters is that the EnKF evaluates the state covariance matrix with an ensemble of members. This ensemble approach allows the nonlinear evolution of errors during the forecast stage. This leads to a higher precision in the reconstruction, but also to a more robust scheme, able to reconstruct evolutions which could not be reconstructed with the former method. Moreover, the ensemble assimilation provides an estimate of the errors on the reconstruction at each point of the domain. The estimation of errors could be a valuable information for plate tectonic reconstructions for regions and times where data are scarce, by showing the possible alternative scenarios supported by the ensemble.

This gain in information and quality for reconstructions comes with a computational price. While we could perform the former assimilation method in one core hour, the method developed here requires several hundreds to several thousands of core hours. However, an efficient parallelization using the PDAF software (Nerger and Hiller, 2013) in combination with the parallel code STAGYY produces a highly parallel ensemble filter, able to perform the assimilations on 768 cores in 20 min for an ensemble of 96 members and 3 hours for an ensemble of 768 members.

The important computational cost of the EnKF limited us in the number of assimilations we could test. After checking the stability of the assimilation results on four different evolutions, we chose to focus on studying the effect of the parameters of the Ensemble data assimilation: the size of the ensemble and the vertical and horizontal correlation lengths. The optimum size of the ensemble for our problem is of the order of 288 members. Indeed, almost tripling the number of members leads to a decrease of the average error of less than 0.2%, and on the contrary, dividing the size of the ensemble by 3 leads to an increase of the average error of more than 1%. Although these differences in errors appear to be small, they affect the quality of the reconstruction of thermal structures, as is illustrated in Fig. 8. The average errors on the temperature field for the estimates shown in Fig. 8 range between 4.8 and 5.8%, so the difference in errors are less than 1%. Locally, this translates into the presence (or absence) of geodynamic structures (like ridges and upwellings) which are artefacts. Covariance localization proved to be important to minimize the error in the reconstruction of mantle structure: for 288 members, the difference in the average error is of 0.85% between the optimal correlation length and the least favorable one. We also investigated the statistics of the cumulative innovation and of the average forecast error for different ensemble sizes and correlation lengths. In a realistic case, these are the only variables available to evaluate the quality of the assimilation. The variation of both cumulative innovation and average forecast error as a function of ensemble size show the same tendency as the average error on the temperature field. On the contrary, the correlation lengths minimizing the norm of the cumulative innovation and the average forecast error were different from the ones minimizing the error on the temperature field. This shows the limits of these indicators to determine the optimal parameters for the assimilation. In a realistic case, rigorous a posteriori evaluation of a data assimilation result would require comparison of the prediction made with independent observations (Talagrand, 2014). For mantle circulation, seismic tomography, topography, true polar wander or the geoid could play this role.

By construction, sequential data assimilation methods do not propagate new information back in time. In the case of the reconstruction of mantle circulation, this is a clear disadvantage since the information on the Earth's surface tectonics tends to become more reliable as we get closer to present-day. Consequently, a natural extension of the present work would be to

implement an Ensemble Kalman Smoother (Evensen and Van Leeuwen, 2000; Van Leeuwen, 2001). In the same way as the EnKF uses sample spatial correlations of the ensemble to update the state of the system with new observations, the Ensemble Kalman smoother uses sample time and space correlations of successive ensembles to update former states with the new observations. Evensen (2003) shows how an Ensemble Kalman Smoother can be implemented with a minimal computational cost

5 alongside a preexisting EnKF. Moreover, Nerger et al. (2014) shows that such algorithm is efficient for nonlinear models, and that in their test case, optimal localization parameters for the Ensemble Kalman Smoother coincide with optimal localization parameters for the EnKF.

As a first approach to test the EnKF for mantle circulation reconstructions, we chose a fairly simple convection model. As already discussed in Sect. 2.1, a more realistic mantle model would have, among other things, a 3D-spherical shell geometry

10 and a higher Rayleigh number. This would substantially increase the size of the data assimilation problem. However, we followed the procedure as described in Nerger and Hiller (2013) to implement the EnKF. This results in a highly scalable filter, enabling the computation of the EnKF assimilation in a reasonable time. An increase in the Rayleigh number also implies thiner boundary layers, slabs and plumes. This could translate into lower optimum correlation lengths for the EnKF. A more realistic model would additionally include a viscosity increase in the lower mantle (Ricard et al., 1993), and the presence of

15 continents. This would tend to lengthen the wavelength of convection in the lower mantle and therefore might ease the mantle circulation reconstruction (see for example Ricard, 2015, sections 7.02.6.3.2 and 7.02.6.7 for a discussion of both effects on mantle convection).

In the synthetic experiments of Sect. 4, the convection model used to produce the series of data is the same as the forward model used during the assimilation. For an application with Earth data, this will not be the case. The equations solved in models

20 of mantle convection still hold some shortcomings (Ricard, 2015). Moreover, theories, observations and experiments do not yet fully constrain parameters, especially rheological ones (King, 2016), and variations in rheology affect the reconstructions of mantle circulation (Bello et al., 2015). Hence it could be fundamental to take into account model errors. A first order solution is to increase the inflation parameter γ in Eq. (20): this would overall increase the a priori uncertainty on the mantle estimation. Performing experiments where the model used to compute the observation is different from the model used for the assimilation

25 would provide us with more information on how to implement model errors. Another solution would be to consider the joint assimilation of the state and model parameters. Although it is in principle possible for the EnKF (Evensen, 2009b), it could be computationally not tractable. Indeed, the response of mantle dynamics to different rheological parametrization is highly nonlinear, and their inversion calls for the development of techniques focusing on rheology, such as adjoint based inversions of rheological parameters (Worthen et al., 2014; Ratnaswamy et al., 2015) or further applications of the recently developed

30 pattern recognition techniques for mantle convection (Atkins et al., 2016).

The choice of the synthetic experiments assimilation window of 150 Myr is a compromise between having the possibility to compute assimilations for various cases and having an assimilation window covering most part of the timespan of plate tectonic reconstructions (Seton et al., 2012; Müller et al., 2016; Torsvik et al., 2010). A real assimilation could take into account a longer timespan and therefore improve the assimilation results. However, the structure of the dataset used for the synthetic experiments

is a very idealized version of the actual plate reconstruction models. We already discuss this issue in Bocher et al. (2016). In the following, we supplement and update this discussion in the light of research that has recently come to the fore.

First, we set a time series of data covering the whole surface of the domain and regularly available, every 10 Myr. Plate tectonic reconstructions data are more complex. They are based on the estimation of finite relative rotations between individual
5 plates, structured into a hierarchy describing global relative motions and anchored in an absolute reference frame. The span of each finite relative rotation is determined depending on the amount and quality of information available for a specific context and therefore varies depending on plate pairs and times. The average span of finite rotations of recent plate models is of the order of 10 Myr (Torsvik et al., 2010) to 5 Myr (Müller et al., 2016), but varies over time with for example 1 Myr resolution for the last 20 Myr in some regions (Merkouriev and DeMets, 2014), or some gaps in the data such as during the
10 cretaceous superchron from 121 to 83 Myr ago (Granot et al., 2012). The continuously closed plate algorithm (Gurnis et al., 2012) produces plate tectonic reconstruction maps continuous in space and time which allows the creation of a series of global plate reconstructions at regular intervals. Nonetheless, creating such a regularized time series of reconstruction might miss tectonic events. Instead, we could adapt the frequency of analyses to the varying plate reconstruction resolution. Additional synthetic experiments with a time-series of data whose frequency evolves through time are necessary to explore the limits of
15 such method.

Second, the observations were perturbed independently with a Gaussian noise of 10% of the respective root mean square value of surface heat flux and surface velocities. The estimation of uncertainties on absolute plate motion models involves estimation of both uncertainties in relative plate motion and on the absolute reference frame (Müller and Wessel, 2015). The main source of information on the motion of plates comes from the map of seafloor magnetic anomalies. Hellinger (1981) developed
20 a method to compute relative motion of plates and associated uncertainties inferred from magnetic anomaly identifications. Recently, Seton et al. (2014) built an open source community database. It gathers seafloor magnetic anomaly identifications, and estimation with Hellinger (1981) method of plates relative motion and associated uncertainties. This database could be used in the future as a basis to automatically produce global plate motion histories and assess their uncertainties. To our knowledge, this has not been done so far at a global scale. On a regional scale and for recent time (5 to 20 Myr), Iaffaldano et al. (2012)
25 applied the trans-dimensional hierarchical Bayesian method to reduce noise in finite rotation data and produce time series of high resolution plate relative motions. More recently, Iaffaldano and Bunge (2015) applied this technique to the relative motion of the pacific plate with North America for the last 75 Myr. The uncertainties on relative plates velocities ranges from 5 to 40% of the root mean square surface velocity. As we go further back in time, the quantification of relative plate motion uncertainties becomes hazardous: most of the seafloor created before 150 Myr has been destroyed by subduction. These plate tectonic
30 reconstructions involve interpretation of different types of data, with a limited spatial coverage and relies heavily on human expertise. For these epochs, maintaining very high uncertainties on the regions where few data supports the reconstructions would be a solution.

6 Conclusions

We applied the ensemble Kalman filter algorithm to the reconstruction of mantle circulation through time. We chose a formulation with covariance inflation and localization to minimize the effect of sampling errors in the estimation of the forecast error covariance matrix. Synthetic "twin" experiments with different evolutions and for different parameters allowed us to assess the efficiency of the algorithm and to determine the optimal parameters for the assimilation.

This work builds on the developments of a first approach to sequential data assimilation for mantle circulation made in Bocher et al. (2016). The EnKF is more robust and on average more accurate than the former method. Additionally, the ensemble Kalman filter provides not only an estimate of mantle circulation, but also detailed maps of uncertainties on this estimation.

We evaluate the accuracy of the EnKF as a function of three main parameters: the size of the ensemble, and two covariance localization parameters, namely the vertical correlation length and horizontal correlation angle. We find that the optimal size of the ensemble is of the order of 300 members. For this ensemble size, the optimal vertical correlation length corresponds to two thirds of the domain thickness, and the optimal horizontal correlation angle is of $\pi/10$ (around 2000 km). These values should be reevaluated as the dynamical model becomes more realistic.

The EnKF was implemented using the parallel data assimilation framework PDAF in a preexisting mantle convection code, STAGYY. The resulting code is highly scalable, which means that the application of the EnKF to realistic data assimilation with plate reconstructions and a 3D spherical mantle model is within reach in a foreseeable future.

Acknowledgements. The research leading to these results has received funding from the European Research Council within the framework of the SP2-Ideas Program ERC-2013-CoG, under ERC grant agreement 617588. Calculations were performed using HPC resources from GENCI-IDRIS (grant 2016-047243). The contribution of Alexandre Fournier is IPGP contribution number XXX.

References

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Monthly Weather Review, 127, 2741–2758, 1999.

Atkins, S., Valentine, A. P., Tackley, P. J., and Trampert, J.: Using pattern recognition to infer parameters governing mantle convection, Physics of the Earth and Planetary Interiors, 257, 171–186, 2016.

Bello, L., Coltice, N., Rolf, T., and Tackley, P. J.: On the predictability limit of convection models of the Earth's mantle, Geochemistry, Geophysics, Geosystems, 15, 2319–2328, 2014.

Bello, L., Coltice, N., Tackley, P. J., Müller, R. D., and Cannon, J.: Assessing the role of slab rheology in coupled plate-mantle convection models, Earth and Planetary Science Letters, 430, 191–201, 2015.

Besse, J. and Courtillot, V.: Apparent and true polar wander and the geometry of the geomagnetic field over the last 200 Myr, Journal of Geophysical Research: Solid Earth, 107, 2002.

Bocher, M., Coltice, N., Fournier, A., and Tackley, P.: A sequential data assimilation approach for the joint reconstruction of mantle convection and surface tectonics, Geophysical Journal International, 204, 200–214, 2016.

Bower, D. J., Gurnis, M., and Seton, M.: Lower mantle structure from paleogeographically constrained dynamic Earth models, Geochemistry, Geophysics, Geosystems, 14, 44–63, 2013.

Bower, D. J., Gurnis, M., and Flament, N.: Assimilating lithosphere and slab history in 4-D Earth models, Physics of the Earth and Planetary Interiors, 238, 8–22, 2015.

Bunge, H.-P., Richards, M. A., Lithgow-Bertelloni, C., Baumgardner, J. R., Grand, S. P., and Romanowicz, B. A.: Time scales and heterogeneous structure in geodynamic Earth models, Science, 280, 91–95, 1998.

Bunge, H.-P., Richards, M., and Baumgardner, J.: Mantle–circulation models with sequential data assimilation: inferring present–day mantle structure from plate–motion histories, Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 360, 2545–2567, 2002.

Bunge, H.-P., Hagelberg, C., and Travis, B.: Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography, Geophysical Journal International, 152, 280–301, 2003.

Burgers, G., Jan van Leeuwen, P., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Monthly Weather Review, 126, 1719–1724, 1998.

Burstedde, C., Stadler, G., Alisic, L., Wilcox, L. C., Tan, E., Gurnis, M., and Ghattas, O.: Large-scale adaptive mantle convection simulation, Geophysical Journal International, 192, 889–906, 2013.

Coltice, N., Rolf, T., Tackley, P., and Labrosse, S.: Dynamic causes of the relation between area and age of the ocean floor, Science, 336, 335–338, 2012.

Conrad, C. P. and Gurnis, M.: Seismic tomography, surface uplift, and the breakup of Gondwanaland: Integrating mantle convection backwards in time, Geochemistry, Geophysics, Geosystems, 4, 2003.

Davies, D. R., Goes, S., Davies, J. H., Schuberth, B., Bunge, H.-P., and Ritsema, J.: Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity, Earth and Planetary Science Letters, 353, 253–269, 2012.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, 1994.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean dynamics, 53, 343–367, 2003.

Evensen, G.: Data assimilation: the ensemble Kalman filter, Springer Science & Business Media, 2009a.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Systems, 29, 83–104, 2009b.

Evensen, G. and Van Leeuwen, P. J.: An ensemble Kalman smoother for nonlinear dynamics, Monthly Weather Review, 128, 1852–1867, 2000.

5   Flament, N., Gurnis, M., and Müller, R. D.: A review of observations and models of dynamic topography, Lithosphere, 5, 189–210, 2013.

Foley, B. J. and Becker, T. W.: Generation of plate-like behavior and mantle heterogeneity from a spherical, viscoplastic convection model, Geochemistry, Geophysics, Geosystems, 10, 2009.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, 1999.

10   Ghelichkhan, S. and Bunge, H.-P.: The compressible adjoint equations in geodynamics: derivation and numerical assessment, GEM-International Journal on Geomathematics, 7, 1–30, 2016.

Granot, R., Dyment, J., and Gallet, Y.: Geomagnetic field variability during the Cretaceous Normal Superchron, Nature Geoscience, 5, 220–223, 2012.

Gurnis, M., Turner, M., Zahirovic, S., DiCaprio, L., Spasojevic, S., Müller, R. D., Boyden, J., Seton, M., Manea, V. C., and Bower, D. J.:
15   Plate tectonic reconstructions with continuously closing plates, Computers & Geosciences, 38, 35–42, 2012.

Hager, B. H. and O'Connell, R. J.: Kinematic models of large-scale flow in the Earth's mantle, Journal of Geophysical Research: Solid Earth (1978–2012), 84, 1031–1048, 1979.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Monthly Weather Review, 129, 2776–2790, 2001.

20   Hassan, R., Müller, R. D., Gurnis, M., Williams, S. E., and Flament, N.: A rapid burst in hotspot motion through the interaction of tectonics and deep mantle flow, Nature, 533, 239–242, 2016.

Hellinger, S.: The uncertainties of finite rotations in plate tectonics, Journal of Geophysical Research: Solid Earth, 86, 9312–9318, 1981.

Hernlund, J. W. and Tackley, P. J.: Modeling mantle convection in the spherical annulus, Physics of the Earth and Planetary Interiors, 171, 48–54, 2008.

25   Horbach, A., Bunge, H.-P., and Oeser, J.: The adjoint method in geodynamics: derivation from a general operator formulation and application to the initial condition problem in a high resolution mantle circulation model, GEM-International Journal on Geomathematics, 5, 163–194, 2014.

Hoteit, I.: Filtres de Kalman réduits et efficaces pour l'assimilation de données en océanographie, Ph.D. thesis, Université Joseph-Fourier-Grenoble I, 2001.

30   Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Monthly Weather Review, 126, 796–811, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Monthly Weather Review, 129, 123–137, 2001.

Iaffaldano, G. and Bunge, H.-P.: Rapid plate motion variations through geological time: Observations serving geodynamic interpretation,
35   Annual Review of Earth and Planetary Sciences, 43, 571–592, 2015.

Iaffaldano, G., Bodin, T., and Sambridge, M.: Reconstructing plate-motion changes in the presence of finite-rotations noise, Nature communications, 3, 1048, 2012.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.: Unified notation for data assimilation: operational, sequential and variational, Practice, 75, 181–189, 1997.

Ismail-Zadeh, A., Korotkii, A., Naimark, B., and Tsepelev, I.: Three-dimensional numerical simulation of the inverse problem of thermal convection, Computational Mathematics and Mathematical Physics, 43, 581–599, 2003.

5 Janjic, T., Nerger, L., Albertella, A., Schröter, J., and Skachko, S.: On domain localization in ensemble-based Kalman filter algorithms, Monthly Weather Review, 139, 2046–2060, 2011.

Kalnay, E., Li, H., Miyoshi, T., YANG, S.-C., and BALLABRERA-POY, J.: 4-D-Var or ensemble Kalman filter?, Tellus A, 59, 758–773, 2007.

King, S. D.: Reconciling laboratory and observational models of mantle rheology in geodynamic modelling, Journal of Geodynamics, 100, 10 33–50, 2016.

Mallard, C., Coltice, N., Seton, M., Müller, R. D., and Tackley, P. J.: Subduction controls the distribution and fragmentation of Earth's tectonic plates, Nature, 535, 140–143, 2016.

McNamara, A. K. and Zhong, S.: Thermochemical structures beneath Africa and the Pacific Ocean, Nature, 437, 1136–1139, 2005.

Merkouriev, S. and DeMets, C.: High-resolution Neogene reconstructions of Eurasia-North America Plate motion, Geophysical Journal 15 International, 198, 366–384, doi:10.1093/gji/ggu142, http://gji.oxfordjournals.org/content/198/1/366.abstract, 2014.

Moucha, R., Forte, A. M., Mitrovica, J. X., Rowley, D. B., Quéré, S., Simmons, N. A., and Grand, S. P.: Dynamic topography and long-term sea-level variations: There is no such thing as a stable continental platform, Earth and Planetary Science Letters, 271, 101–108, 2008.

Müller, D. and Wessel, P.: Pate tectonics, Treatise on Geophysics, 6, 45–93, 2015.

Müller, R. D., Sdrolias, M., Gaina, C., and Roest, W. R.: Age, spreading rates, and spreading asymmetry of the world's ocean crust, Geo- 20 chemistry, Geophysics, Geosystems, 9, 2008.

Müller, R. D., Seton, M., Zahirovic, S., Williams, S. E., Matthews, K. J., Wright, N. M., Shephard, G. E., Maloney, K., Barnett-Moore, N., Hosseinpour, M., et al.: Ocean basin evolution and global-scale plate reorganization events since Pangea breakup, Annual Review of Earth and Planetary Sciences, 44, 107–138, 2016.

Nerger, L.: *PDAF: the Parallel Data Assimilation Framework*, http://pdaf.awi.de/ [Accessed: 2016-09-01], 2016.

25 Nerger, L. and Hiller, W.: Software for ensemble-based data assimilation systems—Implementation strategies and scalability, Computers & Geosciences, 55, 110–118, 2013.

Nerger, L., Hiller, W., and Schröter, J.: The parallel data assimilation framework: experiences with Kalman filtering, in: Use of High Performance in Meteorology, Proceedings of the 11th ECMWF Workshop. World Scientific, Singapore, pp. 63–86, 2005.

Nerger, L., Janjic, T., Schröter, J., and Hiller, W.: A unification of ensemble square root Kalman filters, Monthly Weather Review, 140, 30 2335–2345, 2012.

Nerger, L., Schulte, S., and Bunse-Gerstner, A.: On the influence of model nonlinearity and localization on ensemble Kalman smoothing, Quarterly Journal of the Royal Meteorological Society, 140, 2249–2259, 2014.

Pham, D. T.: Stochastic methods for sequential data assimilation in strongly nonlinear systems, Monthly Weather Review, 129, 1194–1207, 2001.

35 Ratnaswamy, V., Stadler, G., and Gurnis, M.: Adjoint-based estimation of plate coupling in a non-linear mantle flow model: theory and examples, Geophysical Journal International, 202, 768–786, 2015.

Ricard, Y.: Physics of mantle convection, Treatise on Geophysics, 7, 23–71, 2015.

Ricard, Y., Richards, M., Lithgow-Bertelloni, C., and Le Stunff, Y.: A geodynamic model of mantle density heterogeneity, Journal of Geophysical Research: Solid Earth, 98, 21 895–21 909, 1993.

Richards, M. A., Yang, W.-S., Baumgardner, J. R., and Bunge, H.-P.: Role of a low-viscosity zone in stabilizing plate tectonics: Implications for comparative terrestrial planetology, Geochemistry, Geophysics, Geosystems, 2, 2001.

5 Rolf, T., Coltice, N., and Tackley, P.: Statistical cyclicity of the supercontinent cycle, Geophysical Research Letters, 41, 2351–2358, 2014.

Seton, M., Müller, R., Zahirovic, S., Gaina, C., Torsvik, T., Shephard, G., Talsma, A., Gurnis, M., Turner, M., Maus, S., et al.: Global continental and ocean basin reconstructions since 200Ma, Earth-Science Reviews, 113, 212–270, 2012.

Seton, M., Whittaker, J. M., Wessel, P., Müller, R. D., DeMets, C., Merkouriev, S., Cande, S., Gaina, C., Eagles, G., Granot, R., et al.: Community infrastructure and repository for marine magnetic identifications, Geochemistry, Geophysics, Geosystems, 15, 1629–1641,
10 2014.

Stadler, G., Gurnis, M., Burstedde, C., Wilcox, L. C., Alisic, L., and Ghattas, O.: The dynamics of plate tectonics and mantle flow: From local to global scales, Science, 329, 1033–1038, 2010.

Steinberger, B. and O'Connell, R. J.: Changes of the Earth's rotation axis owing to advection of mantle density heterogeneities, Nature, 387, 169–173, 1997.

15 Tackley, P. J.: Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations, Geochemistry, Geophysics, Geosystems, 1, 2000.

Tackley, P. J.: Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid, Physics of the Earth and Planetary Interiors, 171, 7–18, 2008.

Tackley, P. J., Stevenson, D. J., Glatzmaier, G. A., and Schubert, G.: Effects of an endothermic phase transition at 670 km depth in a spherical
20 model of convection in the Earth's mantle, Nature, 361, 699–704, 1993.

Talagrand, O.: Assimilation of observations, an introduction, Journal-Meteorological Society of Japan Series 2, 75, 81–99, 1997.

Talagrand, O.: Errors. A posteriori diagnostics, in: Advanced Data Assimilation for Geosciences: Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012, p. 229, Oxford University Press, 2014.

Torsvik, T. H., Steinberger, B., Gurnis, M., and Gaina, C.: Plate tectonics and net lithosphere rotation over the past 150My, Earth and
25 Planetary Science Letters, 291, 106–112, 2010.

Van Heck, H. and Tackley, P.: Planforms of self-consistently generated plates in 3D spherical geometry, Geophysical Research Letters, 35, 2008.

van Leeuwen, P. J.: Comment on "Data assimilation using an ensemble Kalman filter technique", Monthly Weather Review, 127, 1374–1377, 1999.

30 Van Leeuwen, P. J.: An ensemble smoother with error estimates, Monthly Weather Review, 129, 709–728, 2001.

Walzer, U. and Hendel, R.: Mantle convection and evolution with growing continents, Journal of Geophysical Research: Solid Earth, 113, 2008.

Worthen, J., Stadler, G., Petra, N., Gurnis, M., and Ghattas, O.: Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow, Physics of the Earth and Planetary Interiors, 234, 23–34, 2014.

35 Yoshida, M.: Mantle convection with longest-wavelength thermal heterogeneity in a 3-D spherical model: Degree one or two?, Geophysical Research Letters, 35, 2008.

Zhang, N., Zhong, S., Leng, W., and Li, Z.-X.: A model for the evolution of the Earth's mantle structure since the Early Paleozoic, Journal of Geophysical Research: Solid Earth, 115, 2010.

Zhong, S., Yuen, D. A., and Moresi, L. N.: Numerical methods for mantle convection, Treatise on Geophysics, 7, 197–222, 2015.

Table 1. Values of the parameters of the forward model

Symbol	Meaning	value
Ra	Rayleigh number	10^6
R_h	Non-dimensional internal heating rate	20.5
L	number of grid points in longitude	384
M	number of grid points in radius	48
r_a	Radius of the top of the domain	2.2
r_b	Radius of the bottom of the domain	1.2
T_a	Temperature at the top of the domain	0
T_b	Temperature at the bottom of the domain	0.9
E_A	Activation Energy	23.03
T_1	Temperature at which $\mu_T = 1$	1
β	Factor of viscosity reduction for partial melting	10
T_{s_0}	Solidus Temperature at $r = r_a$	0.6
$\nabla_r T_s$	Radial gradient of the solidus temperature	2
σ_Y	Yield Stress	1.10^4
$\nabla_r \sigma_Y$	Radial gradient of the yield stress	2.10^5

Table 2. Notations and dimensions of data assimilation variables

Symbol	Meaning	Size (Literal)	Size (Value)
$\boldsymbol{x}$	state	$LM + L$	18 816
$\boldsymbol{y}$	data	$L + L$	768
$\mathbf{H}$	observation matrix operator	$(L + L) \times (LM + L)$	$768 \times 18\,816$
$\mathbf{R}$	observation error covariance matrix	$(L + L) \times (L + L)$	768×768
$\mathbf{P}$	state error covariance matrix	$(LM + L) \times (LM + L)$	$18\,816 \times 18\,816$
$\mathbf{X}$	ensemble state	$(LM + L) \times N$	$18\,816 \times N$,
			$(N = 96{,}288 \text{ or } 768)$

Table 3. Notations and range of values tested for data assimilation parameters

Symbol	Meaning	value
N	number of ensemble members	96 to 768
K	number of observation times	16
γ^{+}	maximum inflation factor	1.25
ℓ_v	vertical correlation length	0.3 to 1
ℓ_h	horizontal correlation angle	$\pi/10$ to $\pi/2$