Geophys. J. Int. (2003) **152,** 280–301

Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography

Hans-Peter Bunge,[1] C. R. Hagelberg[2] and B. J. Travis[2]

[1]*Department of Geosciences, Princeton University, Princeton NJ* 08544, *USA. E-mail: bunge@princeton.edu*
[2]*Los Alamos National Laboratory, Los Alamos NM* 87545, *USA. E-mails: hagelberg@lanl.gov; bjtravis@lanl.gov*

Accepted 2000 July 17. Received 2002 July 01; in original form 2002 August 10

SUMMARY

Mantle convection models require an initial condition some time in the past. Because this initial condition is unknown for Earth, we cannot infer the geological evolution of mantle flow from forward mantle convection calculations even for the most recent Mesozoic and Cenozoic geological history of our planet. Here we introduce a fluid dynamic inverse problem to constrain unknown mantle flow back in time from seismic tomographic observations of the mantle and reconstructions of past plate motions using variational data assimilation. We derive the generalized inverse of mantle convection and explore the initial condition problem in high-resolution, 3-D spherical mantle circulation models for a time period of 100 Myr, roughly comparable to half a mantle overturn. We present a synthetic modelling experiment to demonstrate that mid-Cretaceous mantle structure can be inferred accurately from fluid dynamic inverse modelling, assuming present-day mantle structure is well-known, even if an initial first guess assumption about the mid-Cretaceous mantle involved only a simple 1-D radial temperature profile. We also demonstrate that convecting present-day mantle structure back in time by reversing the time-stepping of the energy equation is insufficient to model the mantle structure of the past. The difficulty arises, because such backward convection calculations ignore thermal diffusion effects, and therefore cannot account for the generation of thermal buoyancy in boundary layers as we go back in time. Inverse mantle convection modelling should make it possible to infer a number of flow parameters from observational constraints of the mantle.

Key words: Earth's interior, geodynamics, inverse problems, mantle convection.

1 INTRODUCTION

Mantle convection is a time-dependent process as evidenced by the large-scale reorganizations of plate motions observed in geological reconstructions of the lithosphere (Gordon & Jurdy 1986; Lithgow-Bertelloni & Richards 1998). Indeed, looking at continents geologists recognized such plate reorganizations early on as continental drift (Wegener 1912). The breakup of Gondwanaland, the long-lived southern supercontinent which had been in existence for more than 400 million years (Myr) by the time it rifted apart early in the Cretaceous, offers one of the best examples to study the large-scale dispersal of continental plates (Norton & Sclater 1979; Scotese 1991; van der Voo 1993). For oceanic lithosphere, a casual inspection of past plate motion maps reveals similarly dramatic surface velocity variations occurred in the ocean basins. In fact, even during the late Mesozoic and Cenozoic history of our planet, that is during the past 100 Myr, Earth witnessed such significant events as the creation and destruction of large areas of the ocean floor. This is illustrated by tectonic reconstructions of the lithosphere over the past 120 Myr, a time period for which reliable plate reconstructions are available (see Fig. 1).

Oceanic lithosphere is the upper thermal boundary layer of mantle convection. Unfortunately our relatively detailed reconstructions of past plate-motion are not matched by equivalent knowledge of the flow history in the underlying mantle. It has long been suspected that changes in surface plate-motion are accompanied by changes in mantle flow. Anderson (1982) noted the correlation of the prominent African geoid high with the location of the Mesozoic supercontinent Pangea, while Chase & Sprowl (1983) showed that modern geoid lows correspond closely to areas of subduction some 125 Myr ago. In a remarkable paper, Davies (1984) explained this somewhat surprising lag between surface tectonics and mantle flow as the consequence of mantle convection, reflecting the fact that mantle up and downwellings effectively preserve a memory of earlier plate tectonic regimes. Despite these early efforts nearly twenty years ago to comprehend the flow history of

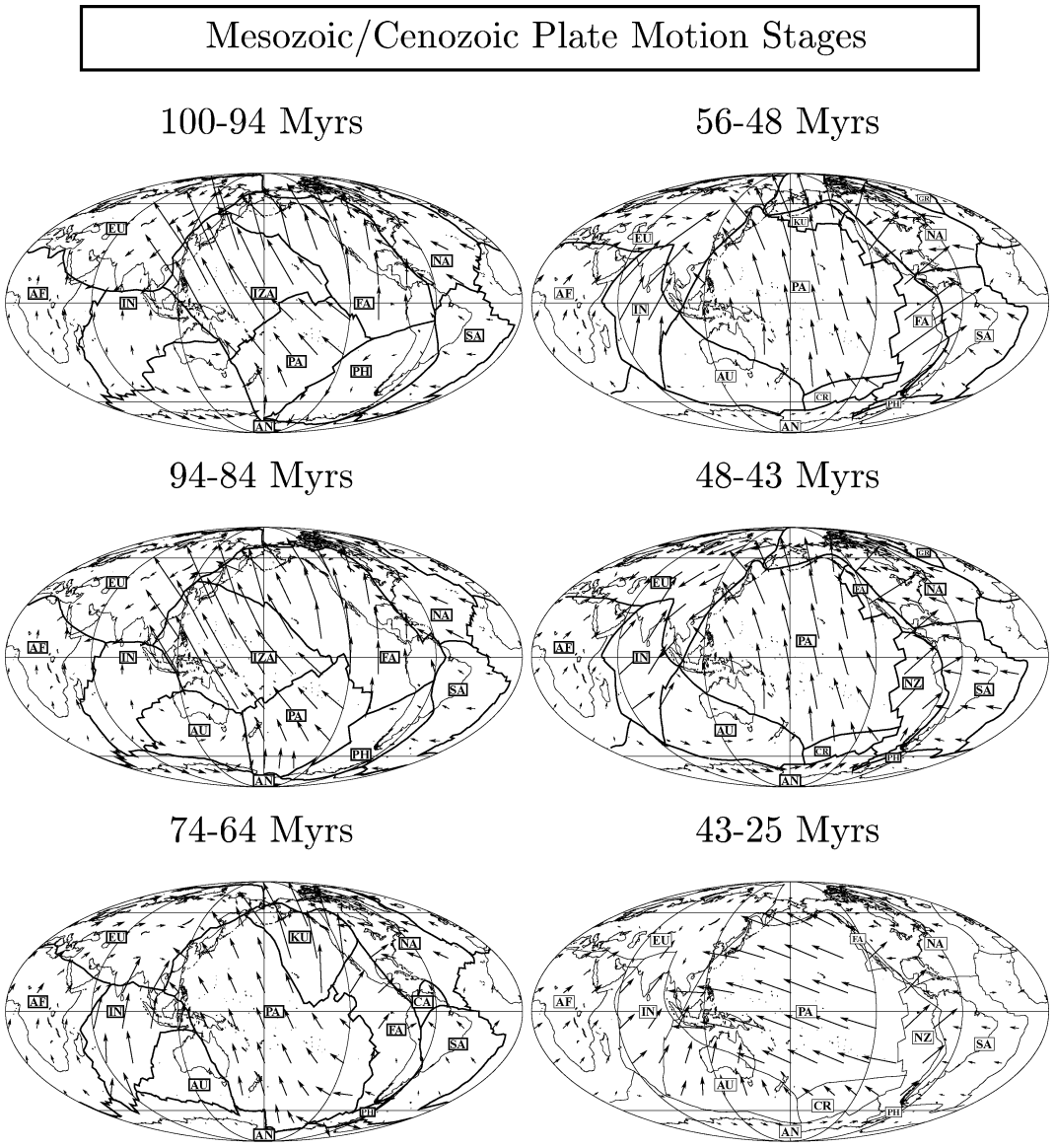

Figure 1. Plate boundary maps for Mesozoic and Cenozoic plate stages from Lithgow-Bertelloni & Richards (1998) with Cenozoic plate motion compiled from the global reconstructions of Gordon & Jurdy (1986). Arrows indicate the direction of plate motion. Their length is proportional to the plate speed. AF: Africa; AN: Antarctica; AR: Arabia; AU: Australia; CA: Caribbean; CO: Cocos; CR: Chatham Rise; EU: Eurasia; FA: Farallon; IN: India; KU: Kula; NA: North America; NZ: Nazca; PA: Pacific; PL: Philippine; PH: Phoenix; and SA: South America. In the Mesozoic the ancient Izanagi (IZA), Farallon (FA), Kula (KU) and Phoenix (PH) plates occupy most of the Pacific basin. Today these plates have largely disappeared. In the Cenozoic the Pacific (PA) plate is inferred to undergo a rapid change from predominantly northward to westward motion at about 43 Myr.

Earth's mantle, we still lack a rigorous understanding of past mantle flow even for the most recent Cenozoic and Mesozoic geological history of our planet.

There are numerous important reasons for trying to model the flow history of Earth's mantle. Continental platform stratigraphy and marine inundations are controlled in large part by time variations of Earth's dynamic topography in response to mantle convection (Gurnis 1990, 1993). The current topographic highstand and relative lack of continental shelf area in southern Africa (Nyblade & Robinson 1994) is indicative of tectonic uplift, and probably supported by lower mantle flow (Lithgow-Bertelloni & Silver 1998; Gurnis *et al.* 2000). A dynamic origin of southern Africa topography is entirely consistent with independent tomographic evidence that the most significant deep mantle seismic low velocity anomaly is located under southern Africa (Su *et al.* 1994; Li & Romanowicz 1996; Grand *et al.* 1997; Ritsema *et al.* 1998). Equally important evidence for deep mantle circulation comes from palaeomagnetic observations of so-called 'True Polar Wander' (TPW), defined with respect to some global reference frame, e.g. hotspots or no-net lithospheric rotation. TPW is often taken to represent motion of the rotation axis with respect to the deep mantle (Jurdy 1981) arising from changes in Earth's inertia tensor in response to large-scale reorganizations of mantle density anomalies (Ricard *et al.* 1992). Geophysical modelling indicates the amount and rate of TPW inferred by palaeomagnetists (Courtillot & Besse 1987) is consistent with density variations associated with subduction (Ricard *et al.* 1993).

TPW estimates for Earth are also in general agreement with TPW calculated from computer simulations of vigorous 3-D spherical mantle convection (Richards *et al.* 1999). In addition to the long-standing TPW problem, there is yet another reason to attempt to better understand the flow history of Earth's mantle. Recent magneto hydrodynamic simulations of the core suggest the geodynamo is sensitive to variations in Core–Mantle Boundary (CMB) heterogeneity (Glatzmaier *et al.* 1999; Bloxham 2000). Palaeomagnetists have long speculated about possible causes for the Cretaceous Normal Superchron (CNS), a remarkable time period lasting from approximately 120 Myr to 85 Myr ago, when the geodynamo occupied a single magnetic polarity. It appears entirely plausible that the great stability of the geodynamo in the mid-Cretaceous occurred in response to heterogeneity variations at the CMB associated with Mesozoic mantle convection. While these examples illustrate the influence of mantle convection on the evolution of our planet, they also demonstrate clearly that we cannot hope to investigate these events without a better knowledge of the temporal character of mantle flow.

We can understand why it is difficult to infer mantle flow at some previous time by recalling that convection is an initial value problem in addition to being a boundary value problem, implying that mantle convection is uniquely determined by an initial condition some time in the past. Stated in mathematical terms, the equations that govern the temporal evolution of the mantle (see, for example, Jarvis & McKenzie 1980) are derived from considerations of conservation of mass, momentum and energy. If we express the upper surface of the mantle by S, its lower surface at the CMB by C, then the mantle is contained in the volume V with boundaries $\partial V = S \cup C$. Denoting the time interval of interest $I = (t_0, t_1)$, where t_1 is the present-day and t_0 is some point in the past, say 100 Myr ago, the governing equations for $x \in V, t \in I$ are given by:

$$0 = \nabla \cdot \mathbf{u} \tag{1}$$

$$0 = -\nabla p + \nabla \cdot (\nu \nabla \mathbf{u}) + R(\bar{T} - T)\hat{\mathbf{k}} \tag{2}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \nabla^2 T + h \tag{3}$$

Here $\mathbf{u}$ is the velocity, p is the pressure, T is the temperature, $\bar{T}$ is the radial temperature profile, ν is the (non-dimensional) kinematic viscosity (assumed here for simplicity to be Newtonian), $\hat{\mathbf{k}}$ is the unit vector in gravitational direction and h is the internal heating rate. The parameter R is the Rayleigh number given by:

$$R = \frac{\beta d^3 \alpha g}{\nu_0 \kappa}, \tag{4}$$

where d is the depth of the mantle, α and κ are the coefficients of volume expansion and thermometric conductivity respectively, g is the acceleration due to gravity, β is a parameter to nondimensionalize temperature ($\theta = \beta T$), and ν_0 is a parameter to nondimensionalize viscosity ($\nu^* = \nu_0 \nu$). For purely bottom heated Rayleigh–Bénard convection, β is the temperature difference between the boundaries. The first equation is a condition to conserve mass which collapses into a volume conservation ($\nabla \cdot \mathbf{u} = 0$), if we make the simplifying assumption that the mantle is incompressible. The second equation is the momentum equation. The momentum equation balances forces that either induce or inhibit convective motion. In the mantle the main force balance is between viscous ($\nabla \cdot (\nu \nabla \mathbf{u})$) and buoyancy ($R(\bar{T} - T)\hat{\mathbf{k}}$) forces. Pressure gradient forces (∇p) are relatively minor, and one can safely assume the pressure distribution is nearly hydrostatic. We note the lack of inertial forces in the momentum equation. The absence of inertial effects in (2) is characteristic for the creeping flow regime of the mantle reflecting the fact that convective flow velocities and temporal variations of $\mathbf{u}$ are small. The momentum balance therefore is instantaneous. The time dependence of convection is introduced through the energy eq. (3). In this equation temperature variations ($\partial T/\partial t$) are governed by heat advection ($\mathbf{u} \cdot \nabla T$) and diffusion ($\nabla^2 T$) in addition to heat production from, say, radioactive decay (h). We supply boundary conditions by specifying the temperature at the surface, S and the CMB, C:

$$T(x, t)|_S = T_S(x, t) \quad x \in S, t \in I \tag{5}$$

$$T(x, t)|_C = T_C(x, t) \quad x \in C, t \in I \tag{6}$$

Here T_S and T_C are constants. Appropriate boundary conditions for the vector velocity field $\mathbf{u}$ are free-slip at the CMB and, for example, no-slip at the surface with specified velocities through time from, say, plate motion reconstructions, together with a no-penetration condition at either surface:

$$\mathbf{u}(x, t) = \mathbf{U}_S(x, t) \quad x \in S, t \in I \tag{7}$$

$$\nabla \mathbf{u} \cdot \hat{n}(x, t) = 0 \quad x \in C, t \in I \tag{8}$$

$$\mathbf{u}(x, t) \cdot \hat{n} = 0 \quad x \in \partial V, t \in I. \tag{9}$$

To study the temporal evolution of the system, we need an initial condition on the temperature at t_0:

$$T(x, t_0) = T_I(x) \quad x \in \bar{V}. \tag{10}$$

Eq. (10) presents us with an insurmountable obstacle to studying mantle flow back in time. This is because while we may constrain present-day mantle structure from seismic tomography, the structure of the mantle some time in the past is necessarily unknown. It is this lack of initial

condition information that ultimately dictates that a predictive mantle convection model capable of reproducing the temporal evolution of the mantle over the past 100 Myr is in a sense inescapably doomed to failure.

The work of Hager & O'Connell (1979) showed the initial condition problem can be at least partially overcome through data assimilation. By including, i.e. assimilating, present-day plate motion as a velocity boundary condition into analytic flow calculations Hager & O'Connell (1979) computed global mantle flow consistent both with their model and with plate motions observed today. Hager & O'Connell (1981) went one step further with the data assimilation problem by adding the buoyancy forces from subducted slabs to their model. Simply put, the purpose of data assimilation in a mantle convection model can be described as follows: using all available information determine as accurately as possible the state of mantle flow. The available information consists first of the observations proper, plate motions or buoyancy forces in the case of Hager & O'Connell (1979, 1981). The second source of information is the dynamic model, and more generally the physical laws governing the flow. These physical laws are fundamentally the principles of conservation of mass, momentum and energy, and a computational model is nothing else than a numerically usable statement of these principles.

From an algorithmic point of view data assimilation is usually implemented in one of two forms: sequential filtering and global smoothing (Wunsch 1996). In sequential filtering the model is integrated forward in time for the period for which observations have been made. Whenever an instant is reached where observations are available, the model is 'updated' or 'corrected'. The amplitude of the correction may be determined in an optimal sense, such as the Kalman filter (Wunsch 1996), or it may be derived suboptimally by guessing (often called 'nudging'). The model is then restarted from the updated state and the process repeated until all available information has been used. Bunge *et al.* (1998) used the nudging method to compute mantle circulation models. Starting from assumed initial conditions for the late Cretaceous mantle (necessarily an approximation for 'true' Cretaceous mantle heterogeneity, which is, of course, unknown) they assimilated the record of Mesozoic and Cenozoic plate motion into high resolution 3-D spherical mantle convection calculations. They performed the assimilation by updating surface velocities in their simulation with plate motions reconstructed for the past 120 Myr. Bunge *et al.* (1998) found the present-day mantle heterogeneity structure predicted from sequential assimilation of past plate motion agrees quite well with mantle heterogeneity mapped by seismic tomographers (Li & Romanowicz 1996; Grand *et al.* 1997).

An appealing aspect of sequential filtering is the constant updating it performs on the convection model: each new observation is used for correcting the latest model state. This makes sequential data assimilation well adapted to mantle convection studies. There is, however, a fundamental drawback. Precisely because of the sequential character of the assimilation each individual observation is used only once, and influences the model state only at later times. Information propagates from the past into the future. But no information is carried back into the past. This limitation is of great disadvantage in mantle convection studies, where our knowledge of the mantle today is infinitely more detailed than knowledge of the mantle at some earlier time. It is therefore advantageous to explore algorithms capable of carrying information explicitly back in time. Arguably the most important information on mantle convection comes from tomographic imaging studies of Earth's interior (e.g. Su *et al.* 1994; Masters *et al.* 1996). Steinberger & O'Connell (1997) explored an unusual approach to include (i.e. assimilate) this tomographic information into mantle convection studies. Taking the present-day mantle heterogeneity as an 'initial condition' and reversing the time-stepping of the energy equation in their mantle flow model, they 'advected' mantle heterogeneity back in time. Thermal diffusion is, of course, unconditionally unstable when one reverses the time-stepping in a mantle convection model. Thus they ignored it altogether. Remarkably Steinberger & O'Connell (1997) found reasonable agreement between the amount and rate of TPW predicted from advecting seismic mantle heterogeneity back in time and TPW estimates inferred for Earth (Courtillot & Besse 1987).

Clearly it is an approximation to ignore thermal diffusion processes in the mantle, which is why Steinberger & O'Connell (1997) limited their approach to flow of the past 64 Myr and estimated the associated error in Steinberger & O'Connell (1998) (see their appendix A2). Therefore we should not have any mathematical expectation that running a mantle convection code into the past provides us with reliable approximations of 'true' mantle heterogeneity at some earlier time. The results of Steinberger & O'Connell (1997), however, do suggest the adverse effects of thermal diffusion are probably minor outside of thermal boundary layers (where diffusion per definition is important), as long as we restrict our attention to time-periods of less than one mantle overturn, i.e. to time-periods of less than 100–200 Myr. More importantly, Steinberger & O'Connell (1997) motivates us to seek out algorithms capable of projecting information into the past that are mathematically correct. One such approach is global smoothing based on fluid dynamic inverse theory. Derived from a variational approach, the method relies on a formal *Adjoint-Model* in conjunction with the forward code (Le Dimet & Talagrand 1986; Talagrand 1997). Looking at the initial condition problem, the task is to compute an optimal initial state consistent with observations of the present state. By computing sensitivities of a performance measure with respect to changes in the initial state, the initial condition can be modified to achieve this optimal state. The adjoint model allows us to calculate the sensitivities of the performance measure in an efficient way, and to infer an optimal flow initialization in a least squares sense. Essentially one fits the mantle convection model to optimal initial conditions, not unlike the inverse problem of finding optimal seismic velocity structures faced by seismologists (Backus & Gilbert 1968; Tarantola 1987).

Exploring adjoint fluid dynamic inverse theory for mantle circulation models is the main theme of this paper, which we organize as follows: first we present the basic theory. In the process we derive the generalized inverse of mantle convection from a variational approach. We present a set of equations, known as the 'Euler–Lagrange' equations, that includes the 'adjoint equations' of mantle convection. We then turn our attention more specifically to the initial condition problem. We present a simplified formulation of the inverse problem with a simplified set of the adjoint equations aimed at constraining unknown mantle heterogeneity of the past from observations of the present state. We use these simplified adjoint equations in numerical mantle convection simulations to explore the method, and conclude for relatively simple mantle convection models that unknown initial conditions can be inferred back in time for at least 100 Myr.

2 THE GENERALIZED INVERSE OF MANTLE CONVECTION

We begin this section by introducing an inverse problem for mantle convection. The method is familiar in oceanography and meteorology (see Bennett (1992) and Wunsch (1996) for aspects related to oceanography, and Courtier *et al.* (1993) for an annotated bibliography of the meteorological literature). But it is novel in mantle convection studies. Consequently we present the basic theory first, before we proceed to the initial-condition problem. For the sake of completeness we derive the most general system of a set of equations known as the Euler–Lagrange equations. These equations consist of the adjoint equations coupled to the forward equations. Solutions of the Euler–Lagrange equations are called the generalized inverse of mantle convection and represent the optimal fit between a model solution and observational data in a least squares sense. Unfortunately the generalized inverse presents a formidable computational challenge and is not easily solved in practice. We will demonstrate that the generalized inverse can be simplified considerably for the initial condition problem.

The generalized inverse is derived from a variational approach to assimilation (Talagrand & Courtier 1987; Courtier & Talagrand 1987). We seek optimal parameters that minimize the difference between model predictions and some observable in a weighted least squares sense. We pose the problem as a control problem, where the control variables are model errors and data misfits. The misfit, or performance measure, is defined through a functional, J, often called the objective functional. (In finite dimensional applications it is simply an objective function.) J maps model and measurement residuals to a scalar through a sum of integrals. To minimize the objective functional we apply the calculus of variations. The necessary condition for a minimum of J, that $\nabla J = 0$, is satisfied by solutions to the Euler–Lagrange (EL) equations. In the process of deriving the EL system we define a set of adjoint variables. The adjoint variables may be thought of as responses to measurements similar to Green's functions (sensitivities), and obey a set of adjoint equations (adjoint with respect to the forward equations through an inner product defined by the performance measure, *cf.* Talagrand & Courtier 1987) nearly identical to the forward model except for forcing terms. The EL system consists of a coupled set of the adjoint and forward equations. Iterative techniques are generally required as a solution method for the EL system. In Section 3 we make some simplifying assumptions to introduce the initial condition inverse problem. The forward and adjoint equations in the EL system for the initial condition problem remain coupled, but in a less complicated way. An iterative scheme for the initial condition problem is shown to be equivalent to a gradient algorithm. A simple gradient algorithm for finding the minimum of J may then be applied. The close correspondence of forward and adjoint equations makes the adjoint approach so attractive in minimizing J because essentially the same computer code can be used to solve the adjoint system as is used for the forward system.

Our discussion of the objective functional starts by breaking J into three components representing respectively model errors, data misfit and the misfit of parameters. Model errors are relatively easy to understand. They arise because the mathematical model equations themselves tend to idealize a physical system. This is evident from the continuity eq. (1). It is easy to see that the assumption of perfect incompressibility in (1) is an approximation, albeit a fairly good one for the mantle. A somewhat more subtle example is apparent from the energy eq. (3). Here the slow decay of radioactive elements inside the Earth results in a time-dependent heat source function (h). Time variations of h have been incorporated into some mantle convection studies (Daly 1980; Christensen 1985). But for the sake of simplicity geodynamicists usually take h to be a constant. Model errors also arise from numerical inaccuracies of the convection model, and are part of any computational approximation to mathematical model equations. We acknowledge such model errors collectively as non-zero terms in the forward mantle convection equations and include divergence, momentum and temperature residuals in the equations below:

$$\nabla \cdot \mathbf{u} = D(x,t) \tag{11}$$

$$\nabla \cdot (\nu \nabla \mathbf{u}) + R(\bar{T} - T)\hat{\mathbf{k}} - \nabla p = \mathbf{M}(x,t) \tag{12}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T - \nabla^2 T - h = \Theta(x,t) \tag{13}$$

Here $D(x,t)$ is the divergence residual, $\mathbf{M}(x,t)$ is the momentum residual, and $\Theta(x,t)$ is the temperature residual. The model error part, J_{eq}, of the objective functional J is the weighted space-time integral over the squared residuals:

$$J_{\mathrm{eq}} = \int_I dt \int_V dx \int_I dt' \int_V dx' D(x,t)\mathbf{W}_D(x,t,x',t')D(x',t') + \int_I dt \int_V dx \int_I dt' \int_V dx' \mathbf{M}(x,t)^T \mathbf{W}_M(x,t,x',t')\mathbf{M}(x',t')$$
$$+ \int_I dt \int_V dx \int_I dt' \int_V dx' \Theta(x,t)\mathbf{W}_\Theta(x,t,x',t')\Theta(x',t'), \tag{14}$$

where we perform the integration over the model space (V) and the assimilation period (I) with appropriate weighting functions W. The weighting functions W could be chosen as the inverse of a covariance C in the sense of (Tarantola 1987):

$$\int_I dt' \int_V dx' \mathbf{W}(x,t,x',t')\mathbf{C}(x',t',y,\xi) = \delta(x-y, t-\tau) \tag{15}$$

Our attention next turns to the data misfit component of J. This part of the objective functional measures how much our mantle convection model departs from a particular set of observations. It is thus at the heart of the assimilation procedure. In order to derive the misfit functional, we must relate model variables to observations that are obtained often from a complicated measurement process. Here we do this by defining measurement functionals that act on the model variables. To make things plain, we focus on two kinds of observables available for data assimilation in mantle convection studies: present-day mantle heterogeneity provided through seismic tomography, and reconstructions of past plate motion histories, discrete in time and non-uniform in space. Looking first at the assimilation of seismic data, we require the mantle convection model to pass within some tolerance of present-day mantle heterogeneity, and express the misfit between model and observation by:

$$J_{\text{temp}} = \int_V dx \int_V dx' \epsilon_T(x) \mathbf{W}_T(x, x') \epsilon_T(x'), \tag{16}$$

$$\epsilon_T(x) = T_d(x) - \mathcal{T}(T(x)) \tag{17}$$

Here $T_d(x)$ is a measure of mantle temperature 'observed' from seismic tomography. $T_d(x)$ could be obtained by relating mantle seismic velocity variations to temperature heterogeneity using, for example, the available experimental data on the temperature sensitivity of seismic sound velocities for mantle silicates (Duffy & Ahrens 1992). The remaining terms are $T(x)$, the model computed mantle temperature, and $\mathcal{T}$, a tomographic operator. $\mathcal{T}$ is a linear measurement operator that accounts for the spatial filtering of mantle structure by seismic models, while $\epsilon_T(x)$ is the data misfit. We perform the integral over the present-day, the only period for which tomographic observations are available, and include a weighting function $\mathbf{W}_T(x, x')$, which could be the inverse of a tomographic covariance operator $\mathbf{C}_{\mathcal{T}}$ (e.g. Tarantola 1987). $\mathbf{W}_T(x, x')$ expresses our 'confidence' in tomographic observations for different regions of the mantle. Indeed, precisely because tomography performs a complex spatial filtering on mantle structure, we don't force the mantle convection model to interpolate the tomographic data. In other words, $\mathbf{W}_T(x, x')$ allows for differences between model solution and observation. Constraints from reconstructions of past plate motion are assimilated into mantle flow in much the same way. We require surface velocities of the convection model to pass within some tolerance of discrete plate motion data $U_d^i(x)$. As in tomography, we express our 'confidence' in plate motion reconstructions through a weighting function $\mathbf{W}_u(x, x')$. Generally speaking, we expect that $\mathbf{W}_u(x, x')$ decreases from the present to the past, reflecting the increasing uncertainties of plate reconstructions as we go back in time. Similarly to tomography we also include a measurement operator, which we may call a tectonic measurement operator here. The operator acts on the surface velocities $\mathbf{u}(x, t)$ computed from the convection model and accounts for the spatial and temporal filtering effects associated with reconstructions of Earth's plate motion history:

$$J_{\text{Svel}} = \int_S dx \int_S dx' [\boldsymbol{\epsilon}_u(x)]^{\text{T}} \mathbf{W}_u(x, x') \boldsymbol{\epsilon}_u(x'), \tag{18}$$

Each component of the velocity data misfit $\epsilon_u^i(x)$ is of the form:

$$\epsilon_u^i(x) = U_d^i(x) - \mathcal{U}^i(u(x, t)), \tag{19}$$

and is either a u-component or a v-component of the ith discrete plate velocity measurement, i.e. the plate motion stage. The weight $\mathbf{W}_u(x, x')$ is a 2×2 matrix with weighting functions for elements representing the inverses of velocity error covariances. Eq. (18) is just a shorthand notation for the weighted sum of squares of surface velocity misfits. The sum includes each velocity component at each measurement time. Temperature and velocity data portions of the objective functional together give us the performance of the total data misfit:

$$J_{\text{data}}(\mathbf{u}, T) = J_{\text{temp}} + J_{\text{Svel}} \tag{20}$$

Any other observable is introduced into the cost function through the same procedure.

We complete our discussion of J with two components related to uncertainties in initial conditions, J_I, and uncertainties in plate motion histories J_{bc}. We introduce these two components, so that we may treat initial conditions and past plate motion histories as model parameters that are given but poorly constrained. In other words, we may regard initial conditions and the history of surface velocities computed from a mantle convection model as model parameters that can be optimized based on observations, for example, of present-day mantle heterogeneity. We begin with J_I. To see how this component enters the objective functional, we take the 'first guess' temperature initial condition $T_I(x)$ as given but poorly known, and represent the difference between 'first guess' field and the optimal model field through the initial condition residual $i(x)$:

$$J_I(T(x, t_0)) = \int_V dx \int_V dx'\, i(x) \mathbf{W}_i(x, x') i(x') \tag{21}$$

$$i(x) = T(x, t_0) - T_I(x) \quad x \in \bar{V} \tag{22}$$

The integral is, of course, nothing other than a distance measure from the 'first guess' field in some weighted ($\mathbf{W}_i(x, x')$) least squares sense. The component J_{bc} is introduced to represent the uncertainties of plate motion histories as we go back in time. Indeed, precisely because past plate motions are difficult to reconstruct the further we go back in time (due in large part because reconstructions are dependent upon the magnetic isochron record of sea-floor spreading, which is limited to the characteristic maximum age for oceanic lithosphere), we could take a 'first guess' plate motion history $\mathbf{U}_S(x, t)$, and allow for corrections based on tomographic observations of subducted oceanic slabs. The distance between 'first guess' plate motion history and subsequent surface velocity updates is measured through a surface velocity residual, $v_s(x, t)$, in an integrated least squares sense:

$$\mathbf{v}_S(x, t) = \mathbf{u}(x, t) - \mathbf{U}_S(x, t) \quad x \in S, t \in \bar{I}. \tag{23}$$

$$J_{\text{bc}} = \int_I dt \int_S ds \int_I dt' \int_S ds'\, \mathbf{v}_S(x, t)^* \mathbf{W}_{v_S}(x, t, x', t') \mathbf{v}_S(x', t') \tag{24}$$

$\mathbf{U}_S(x, t)$ is the 'first guess' plate motion history, $\mathbf{u}(x, t)$ is the surface velocity computed from the circulation model, $\mathbf{v}_S(x, t)$ is the distance from the 'first guess' velocity field, and the spatial integration is carried over the model surface (S). Of course, other boundary condition errors can be added to the cost function in a similar way. Indeed, any unknown model parameter is introduced to the cost function through

this procedure, and we could, for example, estimate an optimal radial mantle viscosity profile from adjoint mantle circulation modelling. We add model, data and parameter residuals to obtain the total objective functional:

$$J(\mathbf{u}, T) = J_{\rm eq} + J_{\rm data} + J_I + J_{\rm bc} \tag{25}$$

To simplify the cost function, we introduce the adjoint variables $\chi(x,t)$, $\phi(x,t)$ and $\tau(x,t)$:

$$\chi(x,t) = \int_I dt' \int_V dx'\, \mathbf{W}_D(x,t,x',t') D(x',t') \tag{26}$$

$$\phi(x,t) = \int_I dt' \int_V dx'\, \mathbf{W}_M(x,t,x',t') \mathbf{M}(x',t') \tag{27}$$

$$\tau(x,t) = \int_I dt' \int_V dx'\, \mathbf{W}_\Theta(x,t,x',t') \Theta(x',t') \tag{28}$$

as weighted integrals over the model error and rewrite J:

$$\begin{aligned} J(\mathbf{u}, T) = & \int_I dt \int_V dx (\nabla \cdot \mathbf{u}) \chi(x,t) + \int_I dt \int_V dx [\nabla \cdot (\nu \nabla \mathbf{u}) + R(\bar{T} - T)\hat{k} - \nabla p]^T \phi(x,t) \\ & + \int_I dt \int_V dx \left(\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T - \nabla^2 T - h \right) \tau(x,t) + \int_V dx \int_V dx'\, \epsilon_T(x) \mathbf{W}_T(x,x') \epsilon_T(x') \\ & + \int_S dx \int_S dx' [\boldsymbol{\epsilon}_u(x)]^{\rm T} \mathbf{W}_u(x,x') \boldsymbol{\epsilon}_u(x') + \int_V dx (T(x,t_0) - T_I(x)) \int_V dx'\, \mathbf{W}_i(x,x') \left(T(x',t_0) - T_I(x')\right) \\ & + \int_I dt \int_S ds \int_I dt' \int_S ds'\, \mathbf{v}_S(x,t)^T \mathbf{W}_{v_S}(x,t,x',t') \mathbf{v}_S(x',t') \end{aligned} \tag{29}$$

The first variation of J, δJ, with respect to the unknown model parameters is given by the following expression, where we assume the weighting functions are symmetric:

$$\begin{aligned} 2\delta J = & \int_I dt \int_V dx (\nabla \cdot \delta \mathbf{u}) \chi(x,t) + \int_I dt \int_V dx [\nabla \cdot (\nu \nabla \delta \mathbf{u}) - R \delta T \hat{k} - \nabla \delta p]^T \phi(x,t) \\ & + \int_I dt \int_V dx \left(\frac{\partial \delta T}{\partial t} + \delta \mathbf{u} \cdot \nabla T + \mathbf{u} \cdot \nabla \delta T - \nabla^2 \delta T \right) \tau(x,t) + \int_V dx \int_V dx'\, \delta T(x,t_1) \mathbf{W}_T(x,x') \epsilon_T(x') \\ & + \int_S dx \int_S dx' [-\mathcal{M}(\delta \mathbf{u})(x,t)]^{\rm T} \mathbf{W}_u(x,x') \boldsymbol{\epsilon}_u(x') + \int_V dx (\delta T(x,t_0)) \int_V dx'\, \mathbf{W}_i(x,x') \left(T(x',t_0) - T_I(x')\right) \\ & + \int_I dt \int_S ds \int_I dt' \int_S ds' [\delta \mathbf{u}(x,t)]^T \mathbf{W}_{v_S}(x,t,x',t') \mathbf{v}_S(x',t') \end{aligned} \tag{30}$$

Here we denote variations of temperature and velocity with the prefix δ, while the expression $\mathcal{M}(\delta \mathbf{u})$ represents measurements of the variation of model computed surface velocities. We integrate δJ by parts in space and time to isolate terms involving $\delta \mathbf{u}$ and δT, and to rearrange the differential operators to act upon the adjoint variables. The formal criterion $\delta J = 0$ at a minimum of the cost function implies that forward and adjoint variables must satisfy a set of so-called *Euler–Lagrange* equations, consisting of adjoint equations coupled to the forward model through error estimates. (We note that deriving the Euler–Lagrange equations is somewhat tedious. As an example of the formalism, we include a detailed derivation of the adjoint energy equation in the Appendix.) Solutions to the Euler–Lagrange equations are candidates for minima of J, and constitute the generalized inverse of mantle convection. We write the adjoint equations for $x \in V$ and $t \in I$ below:

$$\nabla \cdot \phi = 0 \tag{31}$$

$$\nabla \cdot (\nu \nabla \phi) + \tau \nabla T = \nabla \chi \tag{32}$$

$$-\frac{\partial \tau}{\partial t} - \nabla \cdot (\tau \mathbf{u}) + R \mathbf{k} \cdot \phi = \nabla^2 \tau + \delta(x, t - t_1) \left(T_d(x) - \mathcal{T}(T(x))\right), \tag{33}$$

where $\delta(x)$ is the Dirac delta function, and $\mathbf{u}$ and T are forward velocity and temperature that couple the adjoint to the forward system. The adjoint variables also satisfy the following adjoint boundary conditions:

$$\phi = \int_I dt' \int_S ds' [\mathcal{U}(\delta)]^* \mathbf{W}_u(x,t,x',t') \boldsymbol{\epsilon}_u(x',t') \quad x \in S, t \in I \tag{34}$$

$$\frac{\partial \phi}{\partial n} = 0 \quad x \in \partial C,\ t \in I \tag{35}$$

$$\phi \cdot \hat{n} = 0 \quad x \in \partial V,\ t \in I \tag{36}$$

$$\chi = 0 \qquad x \in \partial V, t \in I \tag{37}$$

$$\tau(x,t) = 0 \quad x \in \partial V,\ t \in I \tag{38}$$

The forward system of the Euler–Lagrange equations for $x \in V$ and $t \in I$ are the usual mantle convection equations plus boundary conditions, coupled to the adjoint estimates on the divergence, momentum and energy residual (see eqs 11–13):

Table 1. Adjoint variables and their relationship to model residuals.

Adjoint variable	Definition	Model residual	Estimate
Pressure, χ	$\mathbf{W}_D \bullet D$	Divergence, D	$C_D \bullet \chi$
Velocity, ϕ	$\mathbf{W}_M \bullet \mathbf{M}$	Momentum, $\mathbf{M}$	$C_M \bullet \phi$
Temperature, τ	$\mathbf{W}_\theta \bullet \theta$	Energy, θ	$C_\theta \bullet \tau$

$$\nabla \cdot \mathbf{u} = C_D \bullet \chi \tag{39}$$

$$\nabla \cdot (\nu \nabla \mathbf{u}) + R(\bar{T} - T)\hat{k} - \nabla p = C_M \bullet \phi \tag{40}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T - \nabla^2 T - h = C_\theta \bullet \tau \tag{41}$$

Here we have written the error estimates $D(x, t)$, $\mathbf{M}(x, t)$, $\Theta(x, t)$ of the forward eqs (11–13) explicitly in terms of the adjoint variables, χ, ϕ, τ, and covariances, C_D, C_M, C_θ, exploiting the fact that weights W and covariances C are inverses of each other in the sense of (Tarantola 1987):

$$\int_I dt' \int_V dx' \, \mathbf{W}_D(x, t, x', t') \mathbf{C}_D(x', t', y, \xi) = \delta(x - y, t - \tau), \tag{42}$$

and we simplified the writing by denoting the volume–time integral over $V \times I$ by $\bullet$ in the sense of:

$$\int_I dt' \int_V dx' C_D(x, t, x', t') \chi(x', t') = C_D \bullet \chi(x, t). \tag{43}$$

Using this notation the relationship between model residuals and the adjoint variables is listed compactly in Table 1. The model surface velocities $\mathbf{u}$ satisfy the boundary condition $\mathbf{U}_S(x, t)$ together with an error estimate on the plate motion history:

$$\mathbf{u}(x, t) = \mathbf{U}_S(x, t) + \int_I dt' \int_S ds' C_{\mathbf{v}_S} \nu \frac{\partial \phi}{\partial n} \quad x \in S, t \in I. \tag{44}$$

The starting condition of the forward system is an initial 'first guess' temperature together with an error estimate:

$$T(x, t_0) = T_I(x) + \int_V dx' \, C_i(x, x') \tau(x', t_0), \tag{45}$$

while the starting condition for the adjoint system is a final time condition on the adjoint variable τ:

$$\tau(x, t_1) = 0 \quad x \in V \tag{46}$$

We can understand the physical significance of the Euler–Lagrange equations by closer examination of the adjoint and the forward system. Looking first at the adjoint energy eq. (33), we note the equation is integrated backward in time from the adjoint final time condition $\tau(x, t_1)$. The equation contains a forcing term $\delta(x, t - t_1)(T_d(x) - \mathcal{T}(T(x)))$ involving the residual between model temperature $\mathcal{T}(T(x, t_1))$ and observation $T_d(x)$ at the final state. The forcing term, in effect, implies that temperature residuals at the final time act as the initial condition for the adjoint variable τ. This somewhat unusual character of the adjoint energy equation allows us to relate final state residuals of the mantle convection model to errors in 'first guess' model parameters at some earlier time. The adjoint diffusion operator involves the opposite sign of the forward diffusion operator in the energy equation. This change, of course, makes the adjoint energy equation unconditionally stable to backward-in-time integration. But otherwise the adjoint energy equation is remarkably similar to the forward energy equation in mantle convection. The adjoint mass (31) and momentum (32) equations are even more similar to the mass (1) and momentum (2) equations of forward mantle convection. Here again error information on past plate motion is introduced into the adjoint momentum equation through a forcing term. The term, however, enters through the adjoint boundary condition on ϕ driven by residuals of surface velocities in the mantle convection model relative to the plate motion history. Turning to the forward system of the Euler–Lagrange equations, we see the system also remains largely unchanged from the equations of forward mantle convection, the main difference resulting from error estimates in terms of the adjoint variables that couple the forward to the adjoint system.

We summarize the relationship between model residuals, adjoint variables and estimates of the model residuals in terms of the adjoint variables compactly in Table 1. Error estimates in the forward system in terms of the adjoint variables couple the forward Euler–Lagrange equations to the adjoint equations. The adjoint Euler–Lagrange equations in turn are related to the forward system through data residuals and the forward variables. The coupling of forward and adjoint equations in the Euler–Lagrange equations makes the generalized inverse difficult to solve in practice, although some results have been obtained (e.g. Bennett *et al.* 1993, 1998, 2000; Hagelberg *et al.* 1996). It is instructive to turn our attention to a simplified system of the adjoint mantle convection equations.

3 ADJOINT EQUATIONS TO ESTIMATE TEMPERATURE INITIAL CONDITIONS IN MANTLE CONVECTION

A computationally more tractable approach for the initial condition problem is obtained if we make the following simplifying assumption: we ignore all model errors in the objective functional J except for initial condition error. That is, we assume our mantle convection model

captures all the relevant physical processes and the numerical solution contains at most insignificant error. The assumption amounts to taking the limit as the relevant model covariances tend to zero (weights in the objective functional become infinite) in the Euler–Lagrange equations of the generalized inverse (31–41). We also assume that final state errors result only from variations in the initial temperature distribution. In other words, we choose the temperature initial condition as the only model parameter we want to optimize. We arrive at the Euler–Lagrange equations for this simplified case by taking the appropriate limits in the generalized inverse, and obtain the following system of adjoint and forward equations (see also Talagrand & Courtier 1987) for a detailed derivation of the adjoint vorticity equation related to the assimilation of meteorological observations, and the Appendix for a detailed derivation of the adjoint energy equation of the generalized inverse of mantle convection). The adjoint equations for $x \in V$ and $t \in I$ are now given by:

$$\nabla \cdot \boldsymbol{\phi} = 0 \tag{47}$$

$$\nabla \cdot (\nu \nabla \boldsymbol{\phi}) + \tau \nabla T = \nabla \chi \tag{48}$$

$$-\frac{\partial \tau}{\partial t} - \nabla \cdot (\tau \mathbf{u}) + R\hat{\mathbf{k}} \cdot \boldsymbol{\phi} = \nabla^2 \tau + \delta(x, t - t_1)\,(T_d(x) - \mathcal{T}(T(x)), \tag{49}$$

together with their adjoint boundary conditions:

$$\boldsymbol{\phi} = 0 \quad x \in S, t \in I \tag{50}$$

$$\frac{\partial \boldsymbol{\phi}}{\partial n} = 0 \quad x \in \partial C, t \in I \tag{51}$$

$$\boldsymbol{\phi} \cdot \hat{n} = 0 \quad x \in \partial V, t \in I \tag{52}$$

$$\chi(x, t) = 0 \quad x \in \partial V, t \in I \tag{53}$$

$$\tau(x, t) = 0 \quad x \in \partial V, t \in I, \tag{54}$$

and an adjoint final time condition:

$$\tau(x, t_1) = 0, \tag{55}$$

where the adjoint system is again integrated backward in time from t_1 to t_0 from the adjoint final time condition on τ, just as we saw earlier in the case of the generalized inverse. The forward system is the usual forward model:

$$\nabla \cdot \mathbf{u} = 0 \tag{56}$$

$$\nabla \cdot (\nu \nabla \mathbf{u}) + R(\bar{T} - T)\hat{\mathbf{k}} - \nabla p = 0 \tag{57}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T - \nabla^2 T - h = 0, \tag{58}$$

with boundary conditions, where the error term in the initial temperature condition $T(x, t_0)$ remains as in the generalized inverse:

$$T(x, t_0) = T_I(x) - \int_V dx'\, C_i(x, x')\tau(x', t_0) \tag{59}$$

Here we note, however, that precisely because we ignore all model errors, the forward equations no longer couple to the adjoint equations in the complicated way seen in the forward eqs (39–41) of the generalized inverse. Instead, the coupling between forward and adjoint systems occurs now only through the initial condition, which greatly simplifies the solution of the Euler–Lagrange equations. One method of solving this system is to iterate between forward and adjoint solutions and updating the initial condition error term on each iterate. This procedure is essentially a gradient descent method, where the estimate of the gradient of the cost functional with respect to the initial condition is given by $-\tau(x, t_0)$. The gradient descent algorithm amounts to an iteration on the Euler–Lagrange system in such a way that:

$$T^{n+1}(x, t_0) = T_I(x) - \int_V dx'\, C_i(x, x')\tau^n(x', t_0) \tag{60}$$

If $C_i(x, x') = \delta(x - x')$ (with a weighting function of 1), the expression for the gradient of the objective functional with respect to the initial temperature field is given by:

$$\frac{\delta J}{\delta T_0} = -\tau(x, t_0) \tag{61}$$

A standard conjugate gradient algorithm (Fletcher & Reeves 1964) allows us to minimize the objective functional. The algorithm proceeds as follows:

(1) Compute $T^n(x, t)$ for $t_0 \le t \le t_1$ using the nth initial temperature estimate from eqs (56–58). The 0th initial temperature estimate is $T_I^0(x) = T_I(x)$, the 'first guess' initial condition.

(2) Compute the adjoint $\tau^n(x, t)$ using $\tau^n(x, t_1) = 0$, and forcing with $T_d(x) - T^n(x, t_1)$ from eqs (47–55). In effect, $\tau^n(x, t_1)$ is the final temperature difference between the model estimated temperature, $T^n(x, t_1)$, and the data $T_d(x)$.

(3) The gradient of the cost function with respect to the unknown initial condition is given by $d^n(x) = -\tau^n(x, t_0)$ from eq. (61).

(4) Update the initial temperature to begin the next iteration: $T^{n+1}(x, t_0) = T^n(x, t_0) - \rho d^n(x)$, where ρ is an arbitrary parameter for the step size. Optimal conditions for ρ can be computed, but we did not do so for our numerical experiments.

3.1 Numerical modelling results

Physical oceanographers have made great progress in applying sophisticated adjoint theory to circulation studies of the ocean (Bennett 1992; Wunsch 1996). Adjoint mantle circulation models by comparison are still in their infancy. We begin therefore with a modest adjoint mantle convection experiment aimed at understanding the initial condition problem in simple mantle flow calculations. Our goal is emphatically not to present a comprehensive adjoint mantle convection study, but to explore the efficiency of the method in constraining unknown mantle flow back in time. We use the numerical modelling code TERRA (Bunge & Baumgardner 1995). The code solves for momentum and energy balance of mantle convection at infinite Prandtl number (no inertial forces) in a spherical shell, with the inner radius being that of the outer core and the outer radius corresponding to Earth's surface. TERRA is benchmarked for numerical accuracy (Bunge 1996b; Richards *et al.* 2001) and has been applied to a range of global mantle flow problems. We cast the momentum and energy equation into their weak form and represent the primitive variables locally in grid space through piecewise linear finite elements. The elliptic operator associated with the momentum balance is inverted with an efficient multigrid approach. The method's key advantage is that computational expense scales optimally with the number of numerical grid points (Brandt 1977). This makes multigrid competitive in cost with fast transform schemes (FFT) available for spectral codes. We determine the velocity field $\mathbf{u}$ by solving Stokes equation subject to an incompressibility constraint. Algebraically this is a saddle point problem, where we find solutions to the coupled mass and momentum balance using Uzawa iterations (Cahouet & Chabard 1988). The scheme solves (2) repeatedly, so as to apply the incompressibility constraint from (1). For the energy balance we use an explicit second-order accurate version of the Multi Dimensional Positive Definite Advection Transport Algorithm (MPDATA) (Smolarkiewicz 1984). Our discretization of the mantle is based on the regular icosahedron (Williamson 1968). The icosahedral grid provides an almost uniform triangulation of the sphere allowing us to avoid the 'pole-problem' of conventional latitude-longitude grids. The associated regular data structure is well suited for modern parallel computers and can readily be mapped onto distributed processor arrays such as in Beowulf PC clusters via domain decomposition and explicit message passing (Bunge & Dalton 2001). We use 10 million finite elements in all calculations to follow, resulting in a gridpoint resolution of 50 km throughout the mantle. With this resolution, we are able to resolve a characteristic thermal boundary layer thickness of order 200 km.

We model mantle convection with a Rayleigh number (based on internal heating) of 10^8, using material values corresponding to the surface (sublithospheric) values of the relevant parameters in the Rayleigh number. The value is about an order of magnitude smaller than estimates for the Earth due to computational limitations. Arguably one of the more significant effects of lowering the Rayleigh number is to increase the importance of irreversible thermal diffusion processes relative to time-reversible advection. To keep things simple in our mantle convection model, we make the Boussinesq approximation (Boussinesq 1903) by assuming the mantle is incompressible—that is we keep density constant everywhere except for the buoyancy term of the momentum eq. (2). We also raise the lower mantle viscosity by a factor of 40 relative to the upper-mantle. A significant mantle viscosity increase with depth is suggested by studies of the geoid (Hager & Richards 1989; Ricard *et al.* 1993) and post glacial rebound (Mitrovica 1996; Lambeck *et al.* 1998). A high viscosity lower mantle also induces a long-wavelength convective planform in 3-D mantle convection studies similar to the subduction dominated planform observed for the Earth (Bunge *et al.* 1996a; Tackley 1996). We furthermore assume heat in the mantle is generated mostly internally by radioactive decay of Uranium, Thorium, and Potassium (e.g. Wasserburg *et al.* 1964), with only a minor component of 10 per cent bottom heating from an isothermal core. A relatively small component of core heating is consistent with observations of heat transport associated with hotspot volcanism, or mantle plumes (Morgan 1972), which account for at most about 10 per cent of the total mantle heat flux (Davies 1988; Sleep 1990). These modelling assumptions roughly constitute a 'standard model' for whole mantle convection (Davies & Richards 1992), and their effects have been explored in previous convection studies (Glatzmaier 1988; Bercovici *et al.* 1989; Tackley *et al.* 1994; Bunge *et al.* 1997; Zhong *et al.* 2000).

We start our numerical experiment by computing a reference temperature initial condition. The field is derived by running convection forward in time for ten billion years until the mean surface heatflux variations have dropped below 0.1 per cent. At the upper boundary of the convection model we impose (assimilate) plate motions corresponding to the mid-Cretaceous (119 Myr) as shown in Fig. 2(c). Plate rms velocities are scaled such that convective motion is neither increased nor reduced by the imposed surface velocity field, i.e. we reduce the rms plate velocity to match the rms surface velocity of convection with no assimilated plate motion, thus keeping the model Peclet number unchanged. The temperature initial condition represents mantle flow in quasi steady-state with mid-Cretaceous plate motion, and in lack of other information we could regard it as a crude approximation of the large-scale thermal heterogeneity structure in the mid-Cretaceous mantle. Fig. 2(a) shows a cut-away of the 3-D temperature field. In the western Pacific a mantle downwelling is located at the convergent boundary of the Izanagi and the Eurasian plate, while near surface temperatures are elevated in a narrow zone adjacent to the palaeo-East Pacific Rise due to passive mantle upwelling at the Izanagi, Farallon and Phoenix spreading centers. There is a lack of pronounced active mantle upwellings (plumes) in the deeper mantle as expected from the relatively small amount of bottom heating (10 per cent). Otherwise the overall heterogeneity character is similar to the mid-Cretaceous initial condition assumed by Bunge *et al.* (1998). We start from the initial condition, and compute a simple mantle circulation model by running convection forward in time for 100 Myr. At the surface we impose present-day plate motions (Fig. 2d) over the entire integration period, scaled to match the convective vigor of the calculation. We could, of course, assimilate a record

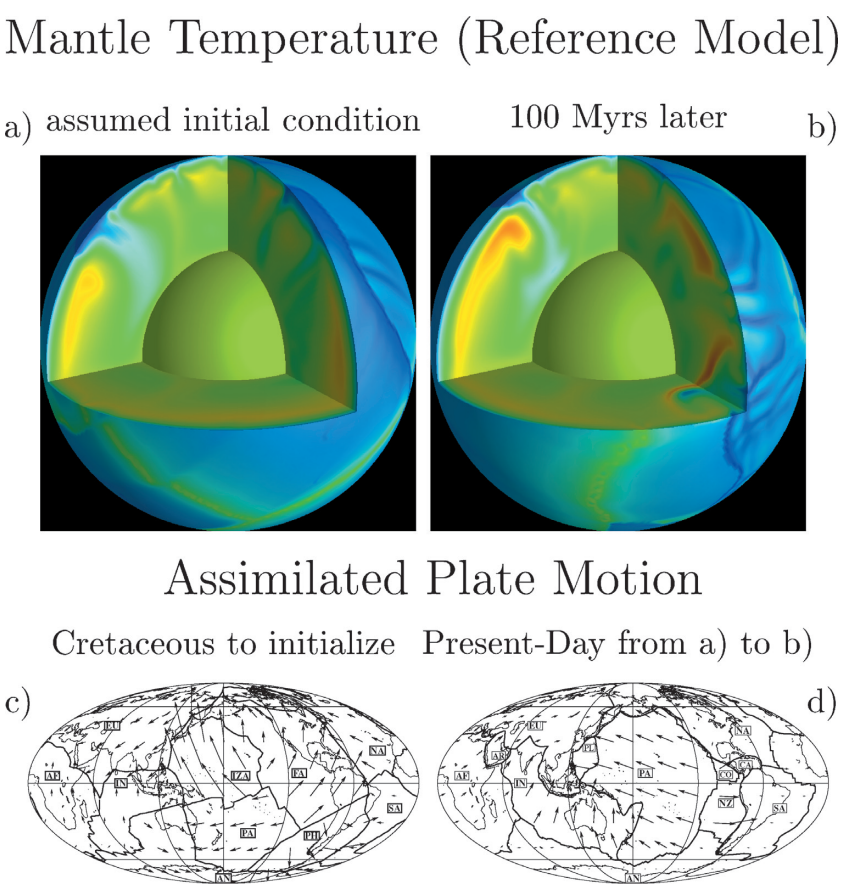

Figure 2. (a) Cut-away of the 3-D temperature initial condition field for the reference mantle circulation model (see text) seen from the Pacific hemisphere. The model is obtained by imposing (assimilating) mid-Mesozoic plate motions (c) until quasi steady-state is reached. Blue is cold, and red is hot and the linear color scale ranges from 0 to 2300 °C. The upper 100 km of the mantle are removed to show the convective planform. Narrow hot zones near the surface reflect passive mantle upwelling at the Izanagi (IZA), Farallon (FA), Pacific (PA) and Phoenix (PH) spreading centers. The cold downwelling in the cross-sectional view under the northwestern Pacific results from subduction of the Izanagi and Farallon plates. (b) Same as (a) but after 100 Myr of present-day plate motion (d) have been imposed. (c) Map of plate boundaries and velocities for the 119–100 Myr stage from Lithgow-Bertelloni & Richards (1998). The ancient Izanagi, Farallon and Phoenix plates occupy most of the Pacific basin. (d) Same as (c) but for the present-day from Gordon & Jurdy (1986). The Izanagi, Farallon and Phoenix plates have largely disappeared.

of Mesozoic and Cenozoic plate motions (Lithgow-Bertelloni & Richards 1998) into the circulation model. But we prefer a simple one-stage plate motion history, in order to avoid any unnecessary complication in our calculation. After 100 Myr the model reaches a final state shown in Fig. 2(b). A mantle downwelling is now located at the convergent boundary of the Pacific and the Eurasian plate, while the East Pacific Rise is evident from elevated temperatures near the Pacific and Nazca spreading plate boundary. These mantle heterogeneity variations are comparable to the overall differences in the plate tectonic regime of the mid-Cretaceous relative to present-day plate motion.

Having defined the reference mantle circulation model we proceed to explore the adjoint method. We compute a ‘perturbed’ mantle circulation model started from a ‘perturbed’ initial state rather than the reference initial condition in Fig. 2(a). The temperature perturbation constitutes an error in our first guess at the initial state, and is implemented by lowering the mid-mantle temperatures locally by 1400 °C under the Pacific plate in a cubic region of 1000 km side-length. We choose the amplitude of the anomalous temperature perturbation such that it is approximately equal in magnitude to the temperature drop across the cold upper thermal boundary layer. The anomalous buoyancy arising from the temperature perturbation is thus comparable to thermal buoyancy variations associated with subduction. Our modified initial state represents ‘imperfect’ knowledge of the ‘true’ initial state, and we may regard it as a ‘first guess’ of the reference initial condition. We run convection forward in time for 100 Myr and arrive at a ‘first guess’ final state, which of course differs from the ‘true’ final state of the reference calculation. Temperature differences between ‘first guess’ and reference model are shown in Fig. 3, separately for the initial and the final state. A look at the final state (Fig. 3b) reveals the cubic temperature anomaly (Fig. 3a) has been sheared and thermally diffused by mantle flow, with most of the advective transport being directed parallel to the motion of the overlaying Pacific plate. The final state temperature residual (or misfit) is the input for an adjoint backward-in-time integration, as we noted earlier in our derivation of the adjoint equations. We solve the adjoint equations and compute an updated (or improved) ‘second guess’ initial condition from the gradient of the cost function J. In other words, we use the adjoint calculation to ‘back-project’ final state temperature residuals to corresponding errors in

the initial state. Our 'second guess' initial state is the starting point for a new forward integration of mantle convection and by implication another backward-in-time integration of the adjoint system based on the 'second guess' final state residual.

We repeat the forward/adjoint calculation one hundred times. The remaining temperature residual after 100 forward/adjoint iterations is illustrated in Figs 3(c) and (d), where we show the temperature difference between 'best guess' and reference model separately for the initial and the final state. The final-state temperature residual has all but disappeared, while a small temperature residual remains in the initial state. We recall, however, that the adjoint is driven by temperature errors in the final state. Hence we are not surprised that the small remaining temperature residual of the final state induces almost no further improvement in the initial condition, implying that mantle flow has become nearly insensitive to the small remaining initial condition error. It is insightful to monitor the norm of the rms temperature residual (true-estimate), both for the final state (present-day) and the initial condition as a function of the adjoint iteration. We show this global error measure in Fig. 7(a). As expected, the residual temperature rms norm drops rapidly, both for initial and the final state in the first 50 adjoint iterations. Later on the convergence levels out. The final error reduction of the initial state exceeds 60 per cent (the residual temperature rms norm drops from 13.2 to 5.8 °C), while the error reduction in the final state (dropping from 13.4 to 0.7 (°C) levels off somewhat better at 95 per cent.

Our results are instructive. But there is a pressing question. Does the adjoint approach perform well, because we made a good first guess? We address this question in a second modelling experiment. Taking a rather extreme view, we pretend to have no knowledge at all of the initial condition and cast our lack of initial condition information into a 1-D temperature profile, wiping out all lateral temperature heterogeneity of the reference initial state (Fig. 2a). We note, however, that our 1-D temperature profile preserves the total heat content of the reference initial condition, allowing us to avoid any artificial secular temperature variation in the calculation. The new 'first guess' initial state is shown in Fig. 4(a). Integrating mantle convection forward in time for 100 Myr with present-day plate motion imposed as in our previous modelling experiment, we arrive at the 'first guess' final state shown in Fig. 4(b). The assimilated present-day plate motion produces mantle downwellings under subduction zones, as expected. There is, for example, a cold downwelling beneath the northwestern Pacific corresponding to subduction of the Pacific plate under Japan and the Kuriles islands. The downwelling, however, does not reach into the lower most mantle indicating the assimilated surface plate motion history (100 Myr) is insufficient to affect large-scale heterogeneity in the deep mantle, which preserves a memory of our 'first guess' initial state. Our poor choice for the initial state results in significant error associated with our 'first guess' calculation. This is evident from inspection of Figs 5(a) and (b), where we show temperature residuals of the 'first guess' calculation relative to our reference model, both for the initial and the final state. Figs 5(a) and (b) reveal our 'first guess' model is characterized by large temperature anomalies concentrated in subduction dominated regions of the circum Pacific, both at the initial and the final state.

We perform 100 forward/adjoint iterations and show mantle temperatures for our 'best guess' case in Figs 4(c) and (d), both for the initial and the final state. We also show the temperature residuals associated with our 'best guess' calculation (Figs 5c and d). It is clear from inspection of Figs 4 and 5 that the temperature in our 'best guess' case corresponds closely to the 'true' temperature distribution of the reference case (Figs 2a and b) even in the deep mantle. The agreement is borne out by the near absence of significant deep mantle temperature residuals in Figs 5(c) and (d), both for the initial and the final state. Closer to the upper thermal boundary layer, however, where thermal diffusion per definition exceeds the effects of heat advection, our adjoint calculation is less effective in reducing the initial temperature errors as indicated by small near surface temperature residuals concentrated at mid oceanic ridges. Again we monitor the norm of the rms temperature residual, both for the initial and the final state per adjoint iteration (Fig. 7b). The rms drops rapidly for the initial and the final state in the first 50 adjoint iterations. Later on the convergence levels out. The convergence rate is similar to the behaviour we noted earlier in our previous adjoint experiment. The final error reduction in the initial state approaches 50 per cent (the rms norm of the temperature residual drops from 195 to 102 °C), while the error reduction in the final state (where the rms norm of the temperature residual drops from 237 to 17 °C) levels off somewhat better at >90 per cent.

3.2 Backward-in-time convection

The success of the adjoint approach raises an interesting question. Could we achieve similar results simply by running our mantle convection model back in time? In other words, could we predict the 'true' initial state in Fig. 2(a) if we took the 'true' final state in Fig. 2(b) as the 'initial condition' for a backward-in-time integration? The question is important, because the adjoint approach requires the repeated evaluation of the forward model together with the adjoint. Our two modelling experiments above each required 100 forward/adjoint iterations. Thus the computational expense associated with adjoint sensitivity analysis exceeds the evaluation of a single forward model by a factor of 200. We examine the question in Fig. 6, where we run mantle convection back in time from the 'true' final state by reversing the time-stepping of the energy equation in our convection model. Thermal diffusion cannot be reversed in time, as we noted earlier. Thus in order to numerically stabilize the backward problem, we continue to run thermal diffusion from hot to cold. In other words, we arbitrarily reverse the sign of the diffusion operator. The modified energy equation of the backward problem is given by:

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = -\nabla^2 T + h \tag{62}$$

The predicted initial temperature after we ran mantle convection back in time for 100 Myr from the 'true' final state is shown in Fig. 6(a). The most noticeable result in our backward calculation is a large cold thermal anomaly located under the East Pacific Rise. There is also a lack of cold downwelling material in the deeper mantle, for example, beneath the northwestern Pacific in the subduction zone beneath Japan. Both results are easily understood. At the East Pacific Rise the cold thermal boundary layer is artificially forced into the upper-mantle, because the

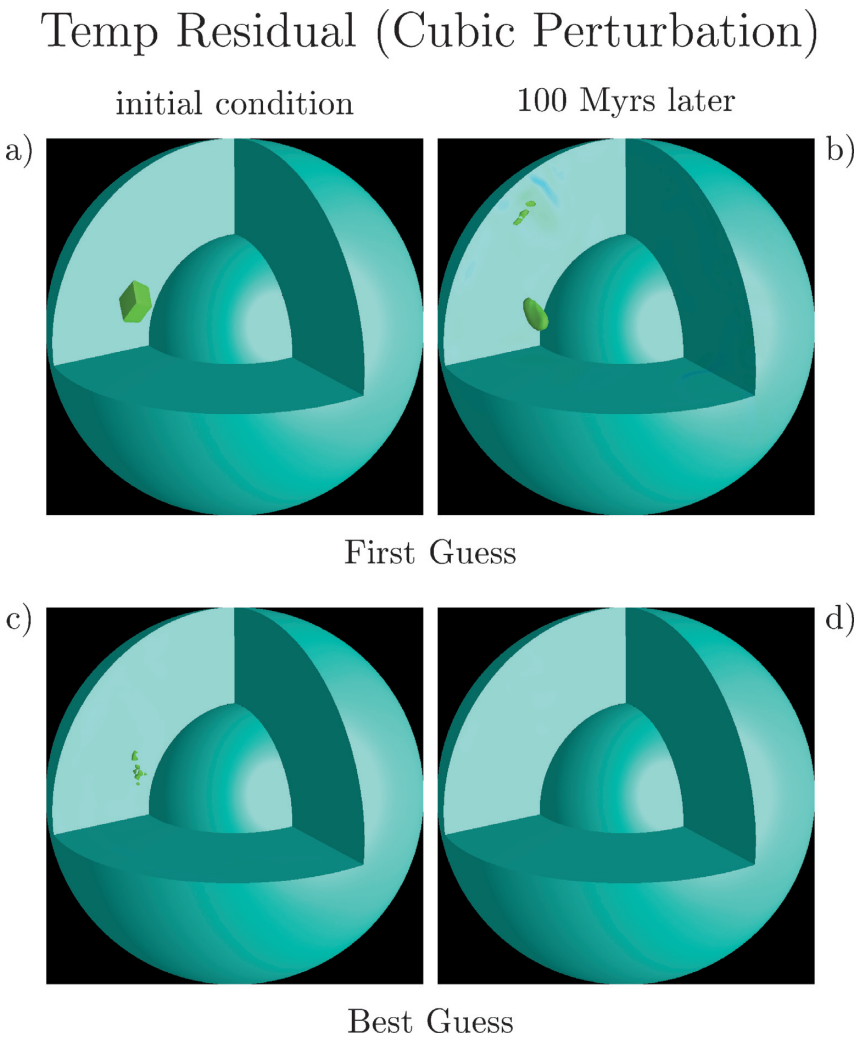

Figure 3. Temperature residual (true—estimate) for a mantle circulation model with cubic initial condition perturbation shown for First Guess (a/b) and Best Guess (c/d) case at the initial and final state (see text). Residuals are plotted on a linear colour scale ranging from −500 to 1400 °C with view angle identical to Fig. 2. The isosurface is pinned to the +400°C temperature residual. The First Guess case has the initially cubic perturbation in the mantle beneath the Pacific diffused and sheared into the direction of Pacific plate motion after 100 Myr of forward integration. For the Best Guess case only a small temperature residual beneath the Pacific remains at the initial condition, while at the final state temperature residuals are nearly zero everywhere.

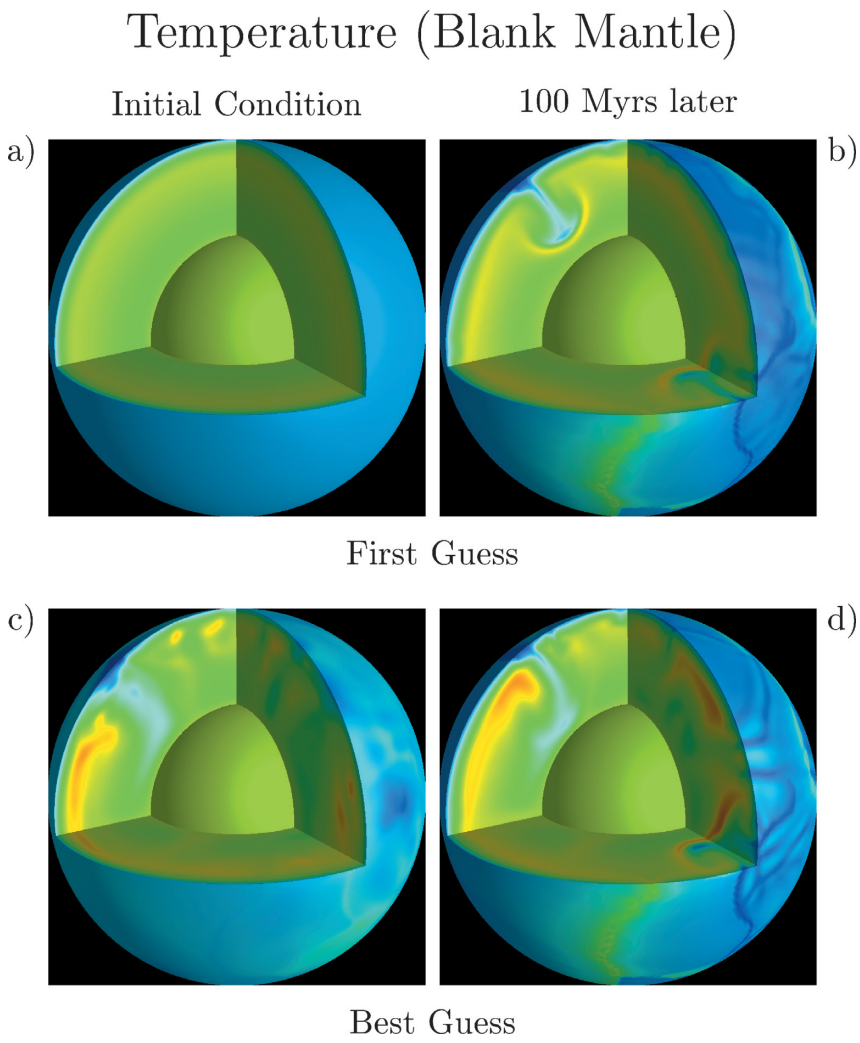

Figure 4. Temperatures for a perturbed mantle circulation model where the assumed initial condition is a 1-D radial temperature profile (see text) shown for First Guess (a/b) and Best Guess (c/d) case at the initial and final state. Blue is cold, red is hot and the color scale and view angle are identical to Figs 2(a) and (b), for comparison. The upper 100 km of the mantle are removed to show the convective planform. The initial condition of the First Guess case shows the assumed 1-D profile. At the final state the First Guess case shows slabs reaching into the mid mantle but not below, because the assimilated 100 Myr plate motion history is insufficient to structure deep mantle flow, which preserves a memory of the initial condition. In the Best Guess case temperatures at the initial and final state agree closely with the 'true' temperatures of the reference model in Figs 2(a) and (b).

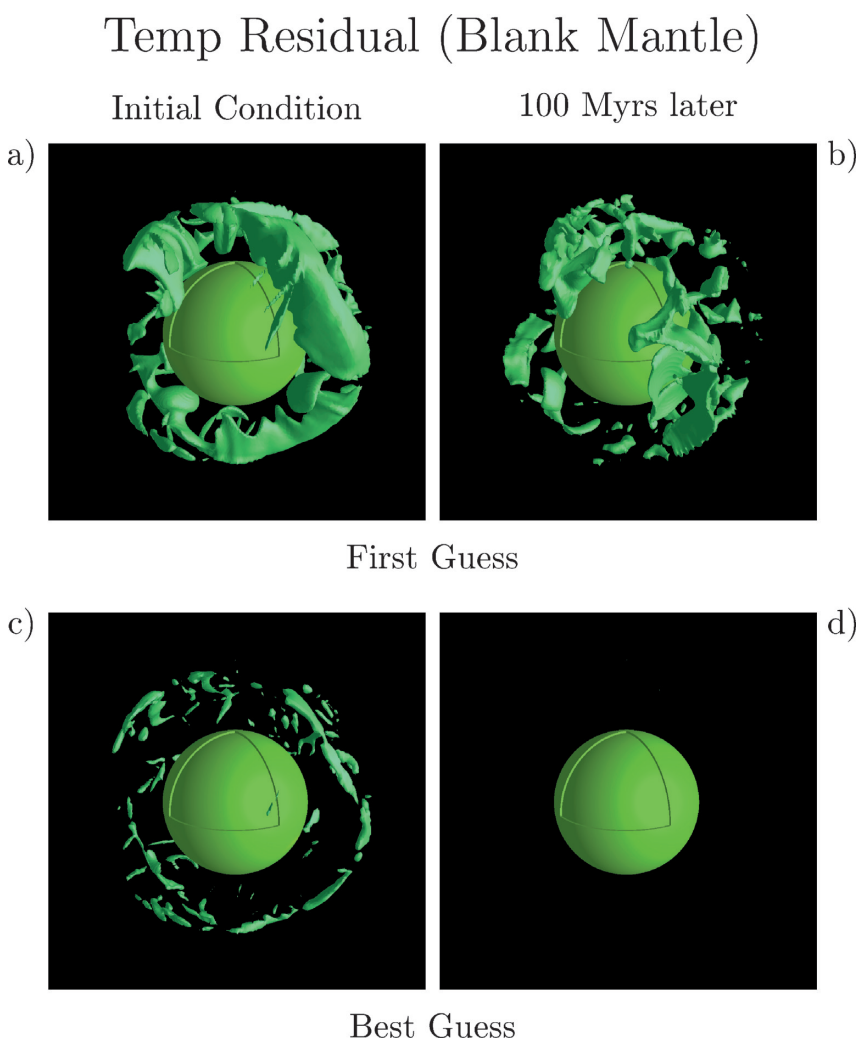

Figure 5. Temperature residuals (true—estimate) for the mantle circulation model of Fig. 4 shown for First Guess (a/b) and Best Guess (c/d) case at the initial and final state. Residuals are plotted on a linear colour scale ranging from ±1700 °C with view angle identical to Fig. 4. The isosurface is pinned to −400°C. In the First Guess case temperature residuals are large in the subduction dominated circum Pacific mantle both at the initial and final state, because the assumed 1-D temperature profile in Fig. 4(a) is a poor guess for the 'true' initial condition in Fig. 2(a). In the Best Guess case temperature residuals at the initial and final state have largely disappeared due to the adjoint calculation, except near the upper thermal boundary layer at the initial state.

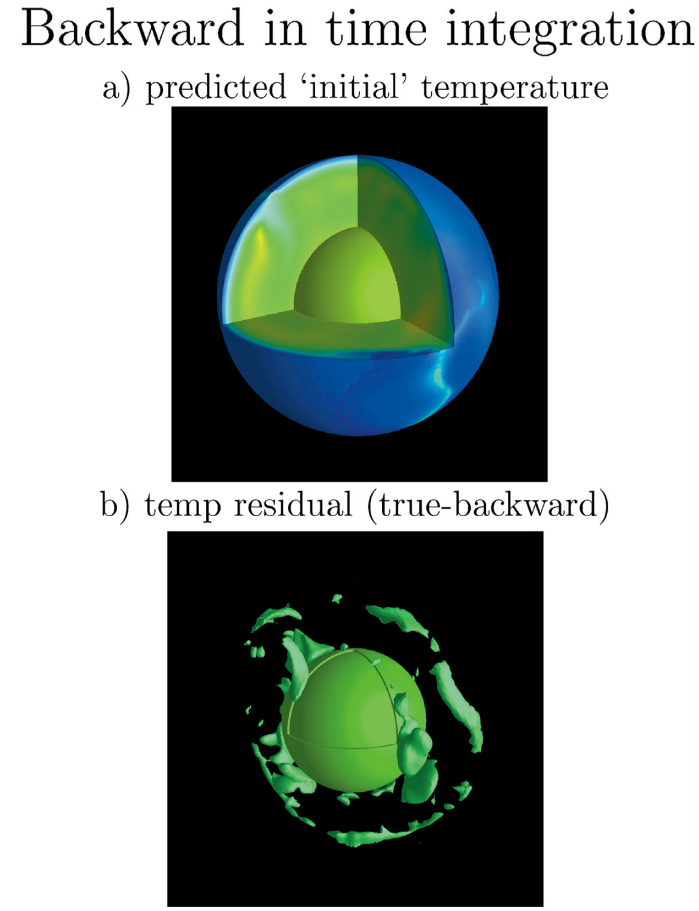

Figure 6. (a) Temperatures after the 'true' final state of Fig. 2(b) has been convected back in time 100 Myr to approximate the 'true' initial state of Fig. 2(a). Colour scale and view angle are identical to Figs 2(a) and (b), for comparison. The upper 100 km of the mantle are removed. Cold mantle is located beneath the East Pacific Rise due to the artificial reversal of the spreading process. There is also a lack of deep mantle slabs, because the backward integration results in slabs rising to the surface. At the surface the reversal of the subduction process results in elevated temperatures in subduction zones beneath South and Central America. (b) Temperature residuals (true—backward) for the backward case. Colour scale, view angle and isosurface value are identical to Fig. 5, for comparison. Large temperature residuals exist throughout the subduction dominated circum Pacific mantle.

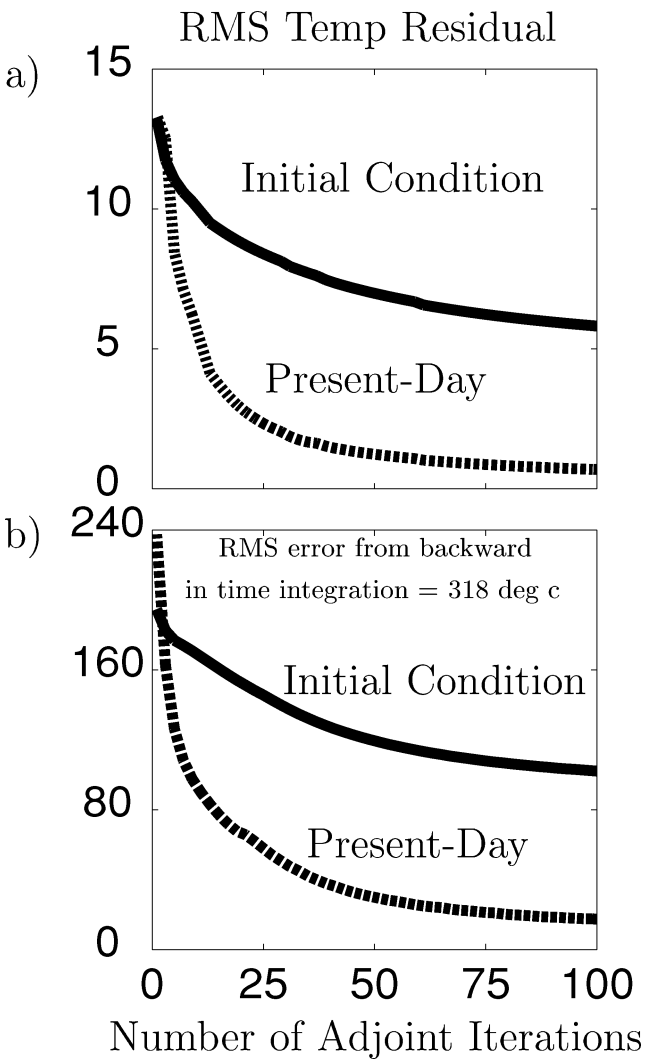

Figure 7. RMS temperature residual at the initial and final state for the locally (a) and globally (b) perturbed initial condition case as a function of adjoint iteration. Temperature residuals decrease rapidly during the first 50 forward/adjoint iterations. Later on the convergence levels off. RMS residual drops from 13.2 to 5.8 °C (a) and from 195 to 102 °C (b) for the initial condition, and from 13.4 to 0.7 °C (a) and from 237 to 17 °C (b) for the present-day. For comparison, the rms temperature initial condition error of the backward calculation (which does not depend on the adjoint iteration) has a much larger value of 318 °C (see text).

spreading process is reversed in our backward calculation. Similarly, reversing the subduction process results in cold slabs rising from the deeper mantle to the surface. The deeper mantle has no 'sources' of new slabs. Therefore, the reversal of the subduction process inevitably leads to a depletion of slabs in the lower mantle.

The great difficulty to predict past mantle flow from backward convection calculations is borne out in Fig. 6(b). Here we show the temperature residual of the initial state (true-backward) predicted from backward integration relative to the 'true' reference initial condition in Fig. 2(a). We plot the temperature residual of the backward model on the same isotherm (−400 °C) we used for the adjoint experiment in Fig. 5, in order to facilitate comparison with our 'best guess' model. From inspection of Fig. 6(b) it is clear that the backward model is characterized by very large temperature anomalies, especially in subduction dominated regions of the circum Pacific, as expected. Not surprisingly in Fig. 7 we find that the large temperature residuals also lead to a large rms temperature residual norm. In fact, the rms norm of the residual temperature in our backward model (318 °C) exceeds the rms norm of our 'best guess' model (102 °C) in Figs 4 and 5 by about a factor of three. This large residual temperature rms norm exceeds even the rms norm of our rather poor 'first guess' model (195 °C) in Fig. 4(a), although we note that the result would be improved, if we had excluded the thermal boundary layers from our backward calculation (Steinberger & O'Connell 1997).

4 DISCUSSION

We have derived the generalized inverse of mantle convection, and we have investigated the initial condition problem with simple, numerical, adjoint mantle circulation models. But we have by no means explored the full potential of the adjoint approach for global geodynamic mantle

studies. Our parameter choice, using an incompressible equation of state and a simplified mantle rheology with no lateral viscosity variations arising from lateral variations in temperature and strain rate (e.g. Tackley 1996; Trompert & Hansen 1998; Zhong *et al.* 2000; Richards *et al.* 2001) is motivated in part by our desire to keep the numerical modelling experiment as simple as possible, in order to better understand the efficiency of the adjoint methodology to constrain mantle flow back in time. Our choice is also motivated by computational considerations. Each adjoint experiment requires 100 evaluations of the forward model together with 100 evaluations of the adjoint code and demands several weeks of computing time on a large parallel computer (a LINUX based Beowulf PC-cluster, Bunge & Davies 2001). However, our results are sufficient to infer that the adjoint approach can be applied to constrain unknown mantle flow back in time for at least 100 Myr, assuming the present-day mantle structure is well known. It is also clear from our results that the adjoint approach is superior to backward-in-time calculations.

Among the approximations we made for the adjoint experiment, our assumption of 'perfect' knowledge of the final state is almost certainly too optimistic. It is clear that seismologists have made great progress in illuminating the internal heterogeneity structure of the mantle. As a result of this development new high-resolution seismic images of the mantle allow us to 'see' into our planet both globally (van der Hilst *et al.* 1997; Grand *et al.* 1997; Ritsema & van Heijst 2000) and locally (van der Lee & Nolet 1997; Ritsema *et al.* 1999) with greater clarity than ever before. However, any seismic inversion invariably imposes a complex spatial filter on mantle structure and must be considered by geodynamicists who look for tomographic information to constrain mantle flow in time. We can explore the effects of tomography, for example, by filtering geodynamic models seismically. A number of studies have attempted to invert geodynamic models with seismic tools (e.g. Johnson *et al.* 1993; Megnin *et al.* 1997; Bunge & Davies 2001). These studies show that geodynamic heterogeneity after inversion does not differ substantially from the input model, suggesting that, although seismic filtering effects are complex and not well understood, their influence on large-scale mantle structure is probably minor. In addition to seismic filtering, we must also consider the competing effects of thermal and chemical heterogeneity in tomographic mantle models. There is increasing evidence that substantial heterogeneity, especially in the deepest mantle, is the result of chemical variations (Ishii & Tromp 1999; van der Hilst & Kárason 1999). Independent geochemical considerations (Hofmann 1997) and dynamic models (Christensen & Hofmann 1994; Tackley 1998; Kellogg *et al.* 1999; Coltice & Ricard 1999) support this view. It is not immediately obvious how to best distinguish geochemical from thermal mantle heterogeneity, or for that matter how to best approach the uncertainty associated with seismic resolution apart from further improving the seismic models. Our results, therefore, suggest that assimilation of tomographic mantle heterogeneity into geodynamic models must be approached with caution.

We must justify another approximation in our adjoint experiment. In calculating mantle flow we assumed 'perfect' knowledge of the surface plate motion history. Just like our assumption of knowing the final state heterogeneity everywhere in our convection model, this premise is almost certainly incorrect. When assimilating past plate motion models into mantle convection calculations, it is important to realize that the tectonic reconstructions of past plate motion, particularly for those plates that have completely disappeared (Izanagi, Phoenix, Kula) or are in the process of disappearing (Farallon), are only approximate, especially the positions of subduction zones and to a lesser extent the poles of rotation. As an example for the uncertainty associated with past plate motion we take the ancient Izanagi subduction zone in the northwestern Pacific. The location of this Mesozoic plate boundary is largely unknown, leading to small but significant differences in mantle heterogeneity inferred from seismic tomography and geodynamic heterogeneity models predicted from sequential assimilation of past plate motion (Bunge *et al.* 1998). The fact that geodynamic heterogeneity structure is sensitive to small variations in the assimilated plate motion model suggests to use forward sensitivity analysis, to explore the consistency of tomographic and tectonic data-sets (Bunge & Grand 2000). Indeed, we could include the history of surface plate motions as another parameter into the cost functional J, especially for plate motion data that is poorly constrained by surface tectonic observations. Our derivation of the generalized inverse of mantle convection shows how to do this. And it is clear, at least from a theoretical point of view, that an adjoint approach would provide a far better way to explore the uncertainties of past plate motion models, than a relatively crude forward sensitivity analysis.

Our finding that the present-day mantle structure could be used to infer heterogeneity at some earlier time is of great interest, because there is much independent support for the view that large-scale mantle heterogeneity changed significantly over the past 100 Myr. One line of evidence comes from palaeomagnetic studies of the reversal frequency of Earth's core (Constable 2000; Gallet & Hulot 1997; McFadden & Merrill 1984). Palaeomagnetists have long speculated about possible causes for the Cretaceous Normal Superchron (CNS), when the geodynamo occupied a single magnetic polarity. Since the end of the CNS the reversal frequency of Earth's magnetic field has been doubling roughly every 20 Myr. It has been proposed that mantle controlled variations in the thermal structure of the CMB might be responsible for increasing the reversal activity of the dynamo (Hide 1967; Cox 1975). Support for this view comes from core convection calculations with imposed lateral heatflux variations (Zhang & Gubbins 1992, 1993), as well as numerical dynamo models (Sarson *et al.* 1997). Lateral variations in CMB heatflux also increase the reversal frequency of the 'Glatzmaier–Roberts' dynamo (Glatzmaier *et al.* 1999), and could cause a breakdown of the geocentric axial dipole (GAD) hypothesis as noted recently by Bloxham (2000). From these studies it is evident that we must seek to improve our understanding of the temporal evolution of the mantle, in order to explore the range of plausible variations in large-scale CMB heterogeneity. Our poor result in predicting lower mantle structure in the 'first guess' model of Fig. 4(b) after 100 Myr of assimilated plate motion shows quite conclusively that we cannot hope to infer temporal variations of CMB structure from sequential assimilation of plate motion histories. In fact, even if we were to assimilate the entire record of late Mesozoic and Cenozoic plate motions into a mantle circulation model, we could probably not model the evolution of deep mantle structure, simply because the lower most mantle preserves a memory of whatever we may assume as the initial condition for the mid-Cretaceous mantle. The adjoint approach, therefore, offers an attractive alternative to estimate the evolution of deep-mantle flow over the past 100 Myr.

There is yet another important reason to pursue studies of variational data assimilation through the adjoint method in mantle convection models. Geodynamicists are now faced with increasingly detailed observational constraints on large-scale structure and dynamics of the mantle. New seismic initiatives suggest we will soon be able to image mantle heterogeneity, for example, under North America on a scale of 100 km or less (Meltzer *et al.* 1999). This dramatic increase in seismic resolution approaches the resolving power of massively parallel circulation models of the mantle. Revised reconstructions of past plate motion (Mueller *et al.* 1993) add further constraints on the temporal evolution of the mantle. So do bounds of global dynamic topography from the Phanerozoic flooding record (Gurnis 1990). The latter, when studied in regions with well controlled tectonic uplift rates, provide important information on the radial mantle viscosity structure as shown recently by Gurnis *et al.* (2000). It is not obvious how to model these new data sets efficiently with forward geodynamic simulations aside from performing an enormous range of forward sensitivity studies (Tackley *et al.* 1994; Bunge *et al.* 1997; Zhong *et al.* 2000; Bunge & Grand 2000; Gurnis *et al.* 2000), effectively computing the gradient of the objective functional J at significant computational cost. It is, however, clear from theoretical considerations, as we explored in this paper, that the adjoint method compared to simple forward sensitivity analysis is a far superior way to obtain ∇J.

5 CONCLUSION

We have introduced an inverse problem for mantle convection and we have modelled mantle circulation with variational data assimilation, investigating the potential to infer past mantle flow and structure from plate motion histories and seismic tomography using an adjoint mantle circulation approach. Our results are relatively straightforward: Large-scale mantle structure can be constrained back in time for at least 100 Myr in simple mantle circulation models, assuming the present-day mantle structure is well known. By comparison, backward calculations of mantle convection that attempt to predict past mantle structure from reversing the time-stepping of the energy equation and running circulation back in time from the present-day are unable to model past mantle flow. The failure of backward-in-time convection studies—in addition to ignoring the effects of thermal diffusion—is primarily a consequence of the fact that such models are unable to infer the generation of thermal buoyancy in boundary layers as we go back in time. Thus the adjoint approach offers an attractive alternative to constrain mantle flow into the past. The potential to model the relatively recent Cenozoic and Mesozoic mantle flow history suggests to explore a broad range of important geophysical and geological problems associated with large-scale mantle flow, including true polar wander, time variations in Earth's dynamic topography as well as the temporal evolution of large-scale heterogeneity at the CMB.

ACKNOWLEDGMENTS

The authors are grateful to Bruce Buffett, Mark Richards, Bernhard Steinberger, Olivier Talagrand, and Carl Wunsch, who provided careful reviews that greatly improved the manuscript. This work was supported by NSF grants EAR-0106651 and EAR-9980457 to HPB. All computations for this work were carried out on a dedicated Beowulf cluster for geophysical modelling at Princeton's Geosciences Department funded by NSF grant EAR-9814635 to HPB. BJT acknowledges support from the DOE Office of Basic Energy Sciences. HPB, CRH and BJT acknowledge support from the Los Alamos branch of the Institute of Geophysics and Planetary Physics (IGPP).

REFERENCES

Abraham, R., Marsden, J. & Ratiu Manifolds, T., 1983. Tensor Analysis, and Applications, Addison–Wesley Pub, Reading, MA.

Anderson, D.L., 1982. Hotspots, polar wander, Mesozoic convection and the geoid, *Nature,* **297,** 391–393.

Backus, G.E. & Gilbert, J.F., 1968. The resolving power of gross earth data, *Geophys. J. R. astr. Soc.,* **16,** 169–205.

Bennett, A.F., 1992. Inverse Methods in Physical Oceanography, Cambridge University Press, Cambridge.

Bennett, A.F., Leslie, L.M., Hagelberg, C.R. & Powers, P.E., 1993. Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method, *Mon. Wea. Rev.,* **121,** 1714–1729.

Bennett, A.F., Chua, B.S., Harrison, D.E. & McPhaden, M.J., 1998. Generalized inversion of tropical atmosphere-ocean data and a coupled model of the tropical Pacific, *J. Climate,* **11,** 1768–1792.

Bennett, A.F., Chua, B.S., Harrison, D.E. & McPhaden, M.J., 2000. Generalized inversion of tropical atmosphere-ocean (TAO) data and a coupled model of the tropical Pacific. Part II: The 1995–1996 La Nina and 1997–1998 El Nino, *J. Climate,* **13,** 2770–2785.

Bercovici, D., Schubert, G. & Glatzmaier, G.A., 1989. Three-dimensional spherical models of convection in the earth's mantle, *Science,* **244,** 950–955.

Bloxham, J., 2000. Sensitivity of the geomagnetic axial dipole to thermal core–mantle interactions, *Nature,* **405,** 63–65.

Boussinesq, J., 1903. Théorie analytique de la chaleur, mise en harmonie avec la thermodynamique et avec la théorie mécanique de la lumière, Vol. 2 Paris: Gauthier-Villars.

Brandt, A., 1977. Multi-level adaptive solutions to boundary value problems, *Math. compu.,* **31,** 333–390.

Bunge, H.-P. & Baumgardner, J.R., 1995. Mantle convection modelling on parallel virtual machines, *Comp. Phys.,* **9,** 207–215.

Bunge, H.-P. & Dalton, M., 2001. Building a high-performance linux cluster for large-scale geophysical modeling, in *Linux Clusters: The HPC Revolution, NCSA Conference Proceedings,* CSM Publications, Urbana-Champaign. http://archive.ncsa.uiuc.edu/LinuxRevolution.

Bunge, H.-P. & Davies, J.H., 2001. Tomographic images of a mantle circulation model, *Geophys. Res. Lett.,* **28,** 77–80.

Bunge, H.-P. & Grand, S.P., 2000. Mesozoic plate-motion history below the northeast Pacific ocean from seismic images of the subducted farallon slab, *Nature,* **405,** 337–340.

Bunge, H.-P., Richards, M.A. & Baumgardner, J.R., 1996a. Effect of depth-dependent viscosity on the planform of mantle convection, *Nature,* **376,** 436–438.

Bunge, H.-P., 1996b. Global mantle convection models, *PhD Thesis,* University of California, Berkeley.

Bunge, H.-P., Richards, M.A. & Baumgardner, J.R., 1997. A sensitivity study of three-dimensional spherical mantle convection at 10(8) Rayleigh number: Effects of depth-dependent viscosity, heating mode, and an endothermic phase change, *J. geophys. Res.,* **102,** 11 991–12 007.

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

Cahouet, J. & Chabard, J.P., 1988. Some fast 3D finite element solvers for the generalized Stokes problem, *Int. J. Numer. Methods Fluids,* **8,** 869–895.

Chase, C.G. & Sprowl, D.R., 1983. The modern geoid and ancient plate boundaries, *Earth planet. Sci. Lett.,* **62,** 314–320.

Christensen, U.R., 1985. Heat transport by variable viscosity convection II: pressure influence, non-newtonian rheology and decaying heat sources, *Phys. Earth planet. Inter.,* **37,** 183–205.

Christensen, U.R. & Hofmann, A.W., 1994. Segregation of subducted oceanic crust in the convecting mantle, *J. geophys. Res.,* **99,** 19 867–19 884.

Coltice, N. & Ricard, Y., 1999. Geochemical observations and one layer mantle convection, *Earth planet. Sci. Lett.,* **174,** 125–137.

Constable, C.G., 2000. On rates of occurence of geomagnetic reversals, *Phys. Earth planet. Inter.,* **118,** 181–193.

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

Courtier, P., Derber, J., Errico, R., Louis, J.-F. & Vukićević, T., 1993. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology, *Tellus,* **45** A, 342–357.

Courtillot, V.E. & Besse, J., 1987. Magnetic field reversals, polar wander, and core-mantle coupling, *Science,* **237,** 1140–1147.

Cox, A., 1975. The frequency of geomagnetic reversals and the symmetry of the nondipole field, *Rev. Geophys.,* **13,** 35–51.

Daly, S.F., 1980. Convection with decaying heat sources: constant viscosity, *Geophys. J. R. astr. Soc.,* **61,** 519–547.

Davies, G.F., 1984. Lagging mantle convection, the geoid and mantle structure, *Earth planet. Sci. Lett.,* **69,** 187–194.

Davies, G.F., 1988. Ocean bathymetry and mantle convection, 1. Large-scale flow and hotspots, *J. geophys. Res.,* **93,** 10 467–10 480.

Davies, G.F. & Richards, M.A., 1992. Mantle convection, *J. Geol.,* **100,** 151–206.

Duffy, T.S. & Ahrens, T.J., 1992. Sound velocities at high pressure and temperature and their geophysical implications, *J. geophys. Res.,* **97,** 4503–4520.

Fletcher, R. & Reeves, C.M., 1964. Function minimization by conjugate gradients, *Comp. J.,* **7,** 149–154.

Gallet, Y. & Hulot, G., 1997. Stationary and nonstationary behaviour within the geomagnetic time scale, *Geophys. Res. Lett.,* **24,** 1875–1878.

Glatzmaier, G.A., 1988. Numerical simulations of mantle convection: Time-dependent, three-dimensional, compressible, spherical shell, *Geophys. Astrophys. Fluid Dyn.,* **43,** 223–264.

Glatzmaier, G.A., Coe, R., Hongre, L. & Roberts, P., 1999. The role of the earth's mantle in controlling the frequency of geomagnetic reversals, *Nature,* **401,** 885–890.

Gordon, R.G. & Jurdy, D.M., 1986. Cenozoic global plate motions, *J. geophys. Res.,* **91,** 12 389–12 406.

Grand, S.P., van der Hilst, R.D. & Widiyantoro, S., 1997. Global seismic tomography: a snapshot of convection in the earth, *GSA Today,* **7,** 1–6.

Gurnis, M., 1990. Bounds on global dynamic topography from Phanerozoic flooding of continental platforms, *Nature,* **344,** 754–756.

Gurnis, M., 1993. Phanerozoic marine inundation of continents driven by dynamic topography above subducting slabs, *Nature,* **364,** 589–593.

Gurnis, M., Mitrovica, J.X., Ritsema, J. & van Heijst, H.-J., 2000. Constraining mantle density structure using geological evidence of surface uplift rates: The case of the African superplume, G^3*, 1,* Paper number 1999GC000035.

Hagelberg, C.R., Bennett, A.F. & Jones, D.A., 1996. Local existence results for the generalized inverse of the vorticity equation in the plane, *Inver. Prob.,* **12,** 437–454.

Hager, B.H. & O'Connell, R.J., 1979. Kinematic models of large-scale flow in the earth's mantle, *J. geophys. Res.,* **84,** 1031–1048.

Hager, B.H. & O'Connell, R.J., 1981. A simple global model of plate dynamics and mantle convection, *J. geophys. Res.,* **86,** 4843–4867.

Hager, B.H. & Richards, M.A., 1989. Long-wavelength variations in earth's geoid—physical models and dynamical implications, *Phil. Trans. R. Soc. Lond.,* A., **328,** 309–327.

Hide, R., 1967. Motions of the earth's core and mantle, and variations in the main geomagnetic field, *Science,* **157,** 55–56.

Hofmann, A.W., 1997. Mantle geochemistry: The message from oceanic volcanism, *Nature,* **385,** 219–229.

Ishii, M. & Tromp, J., 1999. Normal-model and free-air gravity constraints on lateral variations in velocity and density of earth's mantle, *Science,* **285,** 1231–1236.

Jarvis, G.T. & McKenzie, D.P., 1980. Convection in a compressible fluid with infinite Prandtl number, *J. Fluid Mech.,* **96,** 515–583.

Johnson, S., Masters, T.G., Tackley, P.J. & Glatzmaier, G.A., 1993. How well can we resolve a convecting earth with seismic data?, (abstract), *EOS, Trans. Am. geophys. Un.,* **74,** Fall Meeting suppl., 80.

Jurdy, D.M., 1981. True polar wander, *Tectonophysics,* **74,** 1–16.

Kellogg, L.H., Hager, B.H. & van der Hilst, R.D., 1999. Compositional stratification in the deep mantle, *Science,* **283,** 1881–1884.

Lambeck, K., Smither, C. & Johnston, P., 1998. Sea-level change, glacial rebound and mantle viscosity for northern Europe, *Geophys. J. Int.,* **134,** 102–144.

Le Dimet, F.X. & Talagrand, O., 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus,* **38** A, 97–110.

Li, X.D. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.,* **101,** 22 245–22 272.

Lithgow-Bertelloni, C. & Richards, M.A., 1998. The dynamics of Cenozoic and Mesozoic plate motions, *Rev. Geophys.,* **36,** 27–78.

Lithgow-Bertelloni, C. & Silver, P.G., 1998. Dynamic topography, plate driving forces and the African superswell, *Nature,* **395,** 269–272.

Masters, G., Johnson, S., Laske, G. & Bolton, H., 1996. A shear-velocity model of the mantle, *Phil. Trans. R. Soc. Lond.,* A., **354,** 1385–1410.

McFadden, P. & Merrill, R., 1984. Lower mantle convection and geomagnetism, *J. geophys. Res.,* **89,** 3354–3362.

Megnin, C., Bunge, H.-P., Romanowicz, B. & Richards, M.A., 1997. Imaging 3-D spherical convection models: What can seismic tomography tell us about mantle dynamics?, *Geophys. Res. Lett.,* **24,** 1299–1302.

Meltzer, A. *et al.*, 1999. USArray initiative, *GSA Today,* **9,** 8–10.

Mitrovica, J.X., 1996. Haskell [1935] revisited, *J. geophys. Res.,* **101,** 555–569.

Morgan, W.J., 1972. Deep mantle convection plumes and plate motions, *Am. Assoc. Petr. Geol. Bull.,* **56,** 203–213.

Mueller, R.D., Royer, J.-Y. & Lawver, L.A., 1993. Revised plate motions relative to the hotspots from combined Atlantic and Indian ocean hotspot tracks, *Geology,* **21,** 275–278.

Norton, I.O. & Sclater, J.G., 1979. A model for the evolution of the Indian ocean and the breakup of Gondwanaland, *J. geophys. Res.,* **84,** 6803–6830.

Nyblade, A.A. & Robinson, S.W., 1994. The African superswell, *Geophys. Res. Lett.,* **21,** 765–768.

Ricard, Y., Sabadini, R. & Spada, G., 1992. Isostatic deformations and polar wander induced by redistribution of mass within the earth, *J. geophys. Res.,* **97,** 14 223–14 236.

Ricard, Y., Richards, M.A., Lithgow-Bertelloni, C. & Le Stunff, Y., 1993. A geodynamic model of mantle density heterogeneity, *J. geophys. Res.,* **98,** 21 895–21 909.

Richards, M.A., Bunge, H.-P., Ricard, Y. & Baumgardner, J.R., 1999. Polar wandering in mantle convection models, *Geophys. Res. Lett.,* **26,** 1777–1780.

Richards, M.A., Yang, W.-S., Baumgardner, J.R. & Bunge, H.-P., 2001. The role of a low viscosity zone in stabilizing plate tectonics: Implications for comparative terrestrial planetology, G^3, **2,** 1–16.

Ritsema, J. & van Heijst, H.J., 2000. Seismic imaging of structural heterogeneity in earth's mantle: Evidence for large-scale mantle flow, *Sci. Progr.,* **83,** 243–259.

Ritsema, J., Ni, S., Helmberger, D.V. & Crotwell, H.P., 1998. Evidence for strong shear velocity reductions and velocity gradients in the lower mantle beneath Africa, *Geophys. Res. Lett.,* **25,** 4245–4248.

Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 1999. Complex shear wave velocity structure imaged beneath Africa and Island, *Science,* **286,** 1925–1928.

Sarson, G.R., Jones, C.A. & Longbottom, A.W., 1997. The influence of boundary region heterogeneities on the geodynamo, *Phys. Earth planet. Inter.,* **101,** 13–32.

Scotese, C.R., 1991. Jurassic and Cretaceous plate tectonic reconstructions, *Palaeogeog. Palaeoclimat. Palaeoecol.,* **87,** 493–501.

Sleep, N.H., 1990. Hotspots and mantle plumes: Some phenomenology, *J. geophys. Res.,* **95,** 6715–6736.

Smolarkiewicz, P.K., 1984. A fully multidimensional positive definite advection transport algorithm with small implicit diffusion, *J. Comput. Phys.,* **54,** 325–362.

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

Steinberger, B. & O'Connell, R.J., 1998. Advection of plumes in mantle flow; implications for hot spot motion, mantle viscosity and plume distribution, *Geophys. J. Int.,* **132,** 412–434.

Su, W.J., Woodward, R.L. & Dziewonski, A.M., 1994. Degree 12 model of shear velocity heterogeneity in the mantle, *J. geophys. Res.,* **99,** 6945–6980.

Tackley, P.J., 1996. Effects of strongly variable viscosity on three-dimensional compressible convection in planetary mantles, *J. geophys. Res.,* **101,** 3311–3332.

Tackley, P.J., 1998. Three-dimensional simulations of mantle convection with a thermal-chemical boundary layer in D″?, in The Core-Mantle Boundary Region, *Geodynamics,* **28,** 231–253, AGU, Washington D.C.

Tackley, P.J., Stevenson, D.J., Glatzmaier, G.A. & Schubert, G., 1994. Effects of multiple phase transitions in a three-dimensional spherical model of convection in earth's mantle, *J. geophys. Res.,* **99,** 15 877–15 901.

Talagrand, O., 1997. Assimilation of observations, an introduction, *J. Meteorol. Soc. Jap.,* **75,** 191–209.

Talagrand, O. & Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.,* **113,** 1311–1328.

Tarantola, A., 1987. *Inverse Problem Theory,* Elsevier, Amsterdam.

Trompert, R. & Hansen, U., 1998. Mantle convection simulations with rheologies that generate plate-like behaviour, *Nature,* **395,** 686–689.

van der Hilst, R.D. & Kárason, H., 1999. Compositional heterogeneity in the bottom 1000 kilometers of earth's mantle: Toward a hybrid convection model, *Science,* **283,** 1885–1888.

van der Hilst, R.D., Widiyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature,* **386,** 578–584.

van der Lee, S. & Nolet, G., 1997. Seismic image of the subducted trailing fragments of the Farallon plate, *Nature,* **386,** 266–269.

van der Voo, R., 1993. *Palaeomagnetism of the Atlantic, Tethys and Iapetus ocean,* Cambridge University Press, Cambridge.

Wasserburg, G.J., MacDonald, G.J.F., Hoyle, F. & Fowler, W.A., 1964. Relative contributions of Uranium, Thorium, and Potassium to heat production in the earth, *Science,* **143,** 465–467.

Williamson, D.L., 1968. Integration of the barotropic vorticity equation on a spherical geodesic grid, *Tellus,* **20,** 642–653.

Wegener, A., 1912. Die Entstehung der Kontinente, *Geologische Rundschau,* **3,** 276–292.

Wunsch, C., 1996. *The Ocean Circulation Inverse Problem,* Cambridge University Press, Cambridge.

Zhang, K. & Gubbins, D., 1992. On convection in the earth's core driven by lateral temperature variations in the lower mantle, *Geophys. J. Int.,* **108,** 247–255.

Zhang, K. & Gubbins, D., 1993. Convection in a rotating spherical fluid shell with an inhomogeneous temperature boundary condition at infinite Prandtl number, *J. Fluid Mech.,* **250,** 209–232.

Zhong, S., Zuber, M.T., Moresi, L. & Gurnis, M., 2000. Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection, *J. geophys. Res.,* **105,** 11 063–11 082.

APPENDIX A: DERIVING THE GENERALIZED INVERSE FOR THE ENERGY EQUATION

Although the derivation of the generalized inverse is straightforward, it is somewhat tedious. To explore the approach and to provide an example of the methodology, we present here a detailed derivation of the generalized inverse of the energy equation. The objective functional J, consisting of initial condition error ($J_{\rm I}$), error associated with the model equations ($J_{\rm eq}$), and data error ($J_{\rm data}$) is obtained by letting:

$$J_{\rm I}(T(x,t_0)) = \int_V dx \int_V dx'\, i(x)\mathbf{W}_i(x,x')i(x') \quad \text{with} \quad i(x) = T(x,t_0) - T_I(x) \quad x \in \bar{V} \tag{A1}$$

be the initial condition error component, and:

$$J_{\rm eq} = \int_I dt \int_V dx \int_I dt' \int_V dx' \Theta(x,t)\mathbf{W}_\Theta(x,t,x',t')\Theta(x',t') \tag{A2}$$

the model error component, where:

$$\frac{\partial T}{\partial t} + \mathbf{u}\cdot\nabla T - \nabla^2 T - h = \Theta(x,t) \tag{A3}$$

is the model (temperature) residual. Note that because we focus our derivation on the generalized inverse of the energy equation, we include only error components $\Theta(x,t)$ associated with the forward energy equation. In other words, $J_{\rm eq}$ ignores error components associated with the forward momentum and mass conservation equations that we would include otherwise in a full derivation of the generalized inverse of mantle

convection. We complete the objective functional with a component for the data misfit. We let the finite set of temperature data at the final time, t_1, be given by a linear measurement functional acting on the temperatures:

$$T_d^k = \mathcal{M}_x^k(T(x,t_1)) + \epsilon_d^k = \int_V dx'\, M^k(x')T(x',t_1) + \epsilon_d^k \tag{A4}$$

Here T_d^k is the kth temperature datum, ϵ_d^k is the kth measurement error, and $M^k(x)$ is the kth measurement kernel. We assume the kernel is a smoothing kernel. But it could also have the properties of a distribution, such as the Dirac delta. The contribution to the objective functional J from the temperature data residual is:

$$J_{\text{data}}(\mathbf{u},T) = \sum_{i,j}\left(T_d^i - \mathcal{M}_x^i(T(x,t_1))\right) W_T^{ij}\left(T_d^j - \mathcal{M}_x^j(T(x,t_1))\right), \tag{A5}$$

which we write in vector form as:

$$J_{\text{data}}(\mathbf{u},T) = \boldsymbol{\epsilon}_{\mathbf{d}}{}^T \mathbf{W}_T \boldsymbol{\epsilon}_{\mathbf{d}} = (\mathbf{T}_{\mathbf{d}} - \mathcal{M}_x(T(x,t_1))^T \mathbf{W}_T(\mathbf{T}_{\mathbf{d}} - \mathcal{M}_x(T(x,t_1)). \tag{A6}$$

We sum these contributions into the objective functional $J = J_{\text{I}} + J_{\text{eq}} + J_{\text{data}}$, where $\mathbf{W}_i$ and $\mathbf{W}_\Theta$ are symmetric weighting functions and $\mathbf{W}_T$ is a symmetric weight matrix. A sure approach for finding the first variation of J is through a classical perturbation technique. We consider the functions $T(x,t) + \epsilon\delta T$ and $\mathbf{u}(x,t) + \epsilon\delta\mathbf{u}$ for admissible functions δT and $\delta\mathbf{u}$ as the argument for J, where ϵ is a scalar parameter not to be confused with the data residuals $\boldsymbol{\epsilon}_{\mathbf{d}}$ from (A4), and note that δT and $\delta\mathbf{u}$ satisfy all required initial and boundary conditions. Taking the derivative of J with respect to ϵ and looking at the case where $\epsilon = 0$, we arrive at the Gâteaux (or directional) derivative of J (Abraham *et al.* 1983):

$$\begin{aligned}
J(\mathbf{u}(x,t) + \epsilon\,\delta\mathbf{u}(\mathbf{x},\mathbf{t}), T(x,t) + \epsilon\delta T(x,t)) &= \int_V dx \int_V dx'\,(T(x,t_0) + \epsilon\delta T(x,t_0) - T_I(x))\mathbf{W}_i(x,x')\,(T(x',t_0) + \epsilon\delta T(x',t_0) \\
&- T_I(x')) + \int_I dt \int_V dx \int_I dt' \int_V dx' \Bigg[\frac{\partial}{\partial t}(T+\epsilon\delta T) + (\mathbf{u}+\epsilon\delta\mathbf{u})\cdot(\nabla T + \epsilon\delta T) \\
&- \nabla^2(T+\epsilon\delta T) - h\Big]_{(x,t)} \mathbf{W}_\Theta(x,t,x',t')\left[\frac{\partial}{\partial t}(T+\epsilon\delta T) + (\mathbf{u}+\epsilon\delta\mathbf{u})\cdot(\nabla T + \epsilon\delta T) - \nabla^2(T+\epsilon\delta T) - h\right]_{(x',t')} \\
&+ \sum_{i,j}\left(T_d^i - \mathcal{M}_x^i(T(x,t_1) + \epsilon\delta T(x,t_0))\right) W_T^{ij} \times \left(T_d^j - \mathcal{M}_x^j(T(x,t_1) + \epsilon\delta T(x,t_0))\right)
\end{aligned} \tag{A7}$$

Assuming $\mathbf{W}_\Theta$ and $\mathbf{W}_i$ are symmetric in their arguments (e.g. $\mathbf{W}_\Theta(x,t,x',t') = \mathbf{W}_\Theta(x',t',x,t)$ and $W_T^{ij} = W_T^{ji}$), the first derivative of J with respect to ϵ is given by:

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}J &= 2\int_V dx \int_V dx'\,\delta T(x,t_0)\mathbf{W}_i(x,x')(T(x',t_0) + \epsilon\delta T(x',t_0) - T_I(x')) \\
&+ 2\int_I dt \int_V dx \int_I dt' \int_V dx' \left[\frac{\partial}{\partial t}(\delta T) + \delta\mathbf{u}\cdot(\nabla T + \epsilon\delta T) + (\mathbf{u}+\epsilon\delta\mathbf{u})\cdot\nabla\delta T - \nabla^2\delta T\right]_{(x,t)} \mathbf{W}_\Theta(x,t,x',t') \\
&\times \left[\frac{\partial}{\partial t}(T+\epsilon\delta T) + (\mathbf{u}+\epsilon\delta\mathbf{u})\cdot(\nabla T + \epsilon\delta T) - \nabla^2(T+\epsilon\delta T) - h\right]_{(x',t')} \\
&+ 2\sum_{i,j}\left(-\mathcal{M}_x^i(\delta T(x,t_1))\right) W_T^{ij} \times \left(T_d^j - \mathcal{M}_x^j(T(x,t_1) + \epsilon\delta T(x,t_0))\right)
\end{aligned} \tag{A8}$$

We set $\epsilon = 0$, to arrive at the first variation of J:

$$\begin{aligned}
\delta J &= 2\int_V dx \int_V dx'\,\delta T(x,t_0)\mathbf{W}_i(x,x')\,(T(x',t_0) - T_I(x')) + 2\int_I dt \int_V dx \int_I dt' \int_V dx' \Bigg[\frac{\partial}{\partial t}\delta T \\
&+ \delta\mathbf{u}\cdot\nabla T + \mathbf{u}\cdot\nabla\delta T - \nabla^2\delta T\Big]_{(x,t)} \mathbf{W}_\Theta(x,t,x',t') \\
&\times \left[\frac{\partial}{\partial t}T + \mathbf{u}\cdot\nabla T - \nabla^2 T - h\right]_{(x',t')} + 2\sum_{i,j}\left(-\mathcal{M}_x^i(\delta T(x,t_1))\right) W_T^{ij} \times \left(T_d^j - \mathcal{M}_x^j(T(x,t_1))\right),
\end{aligned} \tag{A9}$$

and define an adjoint variable, $\tau(x,t)$, as the weighted residual of the model temperature equation. That is:

$$\tau(x,t) = \int_I dt' \int_V dx'\,\mathbf{W}_\Theta(x,t,x',t')\,\Theta(x',t') = \int_I dt' \int_V dx'\,\mathbf{W}_\Theta(x,t,x',t')\left[\frac{\partial}{\partial t}T + \mathbf{u}\cdot\nabla T - \nabla^2 T - h\right]_{(x',t')} \tag{A10}$$

Note that if we define the covariance operator $C_\Theta(x,t,x',t')$ to be the functional inverse of $\mathbf{W}_\Theta(x,t,x',t')$, that is:

$$\int_I dt' \int_V dx'\,\mathbf{W}_\Theta(x,t,x',t)C_\Theta(x',t',y,\xi) = \delta(x-y,t-\xi), \tag{A11}$$

then from eq. (A10) we establish an estimate of the model residual, Θ, in terms of the adjoint variable, τ:

$$\Theta(x,t) = \int_I dt' \int_V dx'\tau(x',t')\,C_\Theta(x',t',x,t). \tag{A12}$$

We rewrite the first variation of J in terms of the adjoint variable τ:

$$\frac{1}{2}\delta J = \int_V dx \int_V dx'\, \delta T(x,t_0)\mathbf{W}_i(x,x')\,(T(x',t_0)-T_I(x')) + \int_I dt \int_V dx \left[\frac{\partial}{\partial t}\delta T + \delta\mathbf{u}\cdot\nabla T + \mathbf{u}\cdot\nabla\delta T - \nabla^2\delta T\right]_{(x,t)} \tau(x,t)$$
$$+ \sum_{i,j}\left(-\mathcal{M}_x^i(\delta T(x,t_1))\,W_T^{ij}\left(T_d^j - \mathcal{M}_x^j(T(x,t_1)\right)\right). \tag{A13}$$

A necessary condition for a minimum of the performance functional J is that the first variation δJ (A13) be zero at the extremum. The coupled Euler–Lagrange system and the natural boundary conditions are the result of finding conditions for which $\delta J = 0$. Here we accomplish this in a term-by-term fashion, in order to isolate products of variations, δT, in the integrals comprising the penalty functional. Lets consider for a moment just the term involving the time derivative of δT in δJ (A13). We exchange the order of integration and integrate by parts in time, in order to place the time derivative on the adjoint variable τ, leaving a term multiplied by δT plus some time-boundary terms. That is:

$$\int_I dt \int_V dx \left(\frac{\partial}{\partial t}\delta T(x,t)\right)(\tau(x,t)) = \int_V dx \int_I dt \left(\frac{\partial}{\partial t}\delta T(x,t)\right)(\tau(x,t)) = \int_V dx \Big\{[\delta T(x,t_1)\,\tau(x,t_1)$$
$$- \delta T(x,t_0)\,\tau(x,t_0)] - \int_I dt\,\delta T\frac{\partial}{\partial t}\tau(x,t)\Big\} = \int_V dx[\delta T(x,t_1)\,\tau(x,t_1) - \delta T(x,t_0)\,\tau(x,t_0)] + \int_I dt \int_V dx\,\delta T\left(-\frac{\partial\tau}{\partial t}\right). \tag{A14}$$

We will combine the boundary terms from the integration by parts in time with the initial condition term in a moment. Returning to the first variation δJ (A13) of the performance functional, we now use integration by parts in space (the divergence theorem) for the term containing $\nabla\delta T$:

$$\int_I dt \int_V dx(\mathbf{u}\cdot\nabla\delta T)\,(\tau(x,t)) = \int_I dt \left\{\int_V dx[(\mathbf{u}\cdot\nabla\delta T)\,(\tau(x,t))]\right\} = \int_I dt\Big\{\int_V dx[\nabla\cdot(\tau\,\delta T\mathbf{u})]$$
$$- \int_V dx[(\delta T\mathbf{u})\cdot\nabla\tau + \delta T\,\tau\nabla\cdot\mathbf{u}]\Big\} = \int_I dt\left\{\int_S dx[(\tau\,\delta T\mathbf{u})\cdot\mathbf{n}] - \int_V dx\,[\delta T\mathbf{u}\cdot\nabla\tau]\right\} = \int_I dt \int_V dx[\delta T[-\mathbf{u}\cdot\nabla\tau]], \tag{A15}$$

having used the fact that $\mathbf{u}\cdot\mathbf{n} = 0$ on the boundary S and $\nabla\cdot\mathbf{u} = 0$ everywhere. Similarly, we integrate the term in δJ (A13) involving $\nabla^2\delta T$ by parts to obtain:

$$\int_I dt \int_V dx\nabla^2\delta T(x,t)\tau(x,t) = \int_I dt\left\{\int_V dx\nabla^2\delta T\,\tau\right\} = \int_I dt\Big\{\int_V dx\,\nabla^2\tau\,\delta T + \int_S dx[\tau\nabla\delta T\cdot\mathbf{n} - \delta T\,\nabla\tau\cdot\mathbf{n}]$$
$$+ \int_C dx[\tau\nabla\delta T\cdot\mathbf{n} - \delta T\,\nabla\tau\cdot\mathbf{n}]\Big\} = \int_I dt \int_V dx\,\nabla^2\tau\,\delta T + \int_I dt \int_S ds[\tau\nabla\delta T\cdot\mathbf{n} - \delta T\,\nabla\tau\cdot\mathbf{n}] + \int_I dt \int_C ds[\tau\nabla\delta T\cdot\mathbf{n} - \delta T\,\nabla\tau\cdot\mathbf{n}]. \tag{A16}$$

We deduce the following natural boundary conditions for the adjoint variable τ on the core-mantle boundary (C) and the outer surface of the mantle (S) from (A16). On C, S, where temperatures are specified, i.e. where we assume that $\delta T = 0$, $\tau(x,t) = 0$ for $x \in C, S$ since $\nabla\delta T\cdot\mathbf{n}$ is arbitrary. We furthermore drop the term containing $\delta\mathbf{u}$ in δJ at this point, since it contributes to the adjoint momentum equation. (We would, of course, include this term in a full derivation of the generalized inverse of mantle convection.) Collecting terms we have:

$$\frac{1}{2}\delta J = \int_V dx \int_V dx'\,\delta T(x,t_0)\mathbf{W}_i(x,x')(T(x',t_0)-T_I(x')) + \int_I dt \int_V dx\left[-\frac{\partial\tau}{\partial t} - \mathbf{u}\cdot\nabla\tau + \nabla^2\tau\right]\delta T$$
$$+ \int_V dx[\delta T(x,t_1)\,\tau(x,t_1) - \delta T(x,t_0)\,\tau(x,t_0)] + \sum_{i,j}\left(-\mathcal{M}_x^i(\delta T(x,t_1))\,W_T^{ij}\times\left(T_d^j - \mathcal{M}_x^j(T(x,t_1)\right)\right), \tag{A17}$$

assuming the natural boundary conditions on τ hold, and neglecting the term containing $\delta\mathbf{u}$ (see above). The variation $\delta T(x,t_1)$ is arbitrary, which implies that $\tau(x,t_1) = 0$ providing the natural "final" time condition for the adjoint variable τ. Under this condition the first variation of J reduces to:

$$\frac{1}{2}\delta J = \int_V dx\left[\int_V dx'\,\delta T(x,t_0)\mathbf{W}_i(x,x')\times(T(x',t_0)-T_I(x')) - \delta T(x,t_0)\tau(x,t_0)\right]$$
$$+ \int_I dt \int_V dx\left[-\frac{\partial\tau}{\partial t} - \mathbf{u}\cdot\nabla\tau + \nabla^2\tau\right]\delta T + \sum_{i,j}\left(-\mathcal{M}_x^i(\delta T(x,t_0))\,W_T^{ij}\times\left(T_d^j - \mathcal{M}_x^j(T(x,t_1)\right)\right). \tag{A18}$$

We next collect the terms containing the initial variation $\delta T(x,t_0)$, that is:

$$\int_V dx\left[\int_V dx'\,\delta T(x,t_0)\mathbf{W}_i(x,x')\,(T(x',t_0)-T_I(x')) - \delta T(x,t_0)\,\tau(x,t_0)\right]$$
$$= \int_V dx\delta T(x,t_0)\left[\int_V dx'\,\mathbf{W}_i(x,x')(T(x',t_0)-T_I(x')) - \tau(x,t_0)\right]. \tag{A19}$$

For arbitrary $\delta T(x,t_0)$ the above term, as part of δJ (A13), is zero if the adjoint variable:

$$\tau(x,t_0) = \int_V dx'\,\mathbf{W}_i(x,x')\,(T(x',t_0)-T_I(x')). \tag{A20}$$

The above equations imply that the gradient of J with respect to the initial temperature may be thought of as:

$$\tau(x, t_0) - \int_V dx' \, \mathbf{W}_i(x, x') \, (T(x', t_0) - T_I(x')), \tag{A21}$$

where we note that an iterative method that reduces the norm of the gradient is a gradient descent method. Assuming that the weighting function $\mathbf{W}_i$ is defined such that it has a functional inverse C_i in the sense of (Tarantola 1987):

$$\int_V dx' \, \mathbf{W}_i(x, x') C_i(x', y) = \delta(x - y), \tag{A22}$$

we can invert the weighting operator and solve for the model initial condition $T(x, t_0)$. The condition:

$$\tau(x', t_0) = \int_V dx \, \mathbf{W}_i(x', x) \, (T(x, t_0) - T_I(x)) \quad \text{implies} \tag{A23}$$

$$\int_V dx' \, C_i(x', y) \tau(x', t_0) = \int_V dx' \, C_i(x', y) \int_V dx \, \mathbf{W}_i(x', x) \, (T(x, t_0) - T_I(x)), \tag{A24}$$

where the right hand side reduces to:

$$\int_V dx \int_V dx' \, C_i(x', y) \mathbf{W}_i(x', x) \, (T(x, t_0) - T_I(x)) = \int_V dx \, \delta(x - y) \, (T(x, t_0) - T_I(x)) = T(y, t_0) - T_I(y). \tag{A25}$$

Exchanging x for y we obtain:

$$T(x, t_0) = T_I(x) - \int_V dx' \, C_i(x', x) \tau(x', t_0). \tag{A26}$$

Thus, the natural boundary conditions in time emerge as a final time condition on the adjoint variable, $\tau(x, t_1) = 0$, and as a correction (coupling) to the forward model initial condition, $T(x, t_0)$, through the solution to the adjoint equation. The performance functional under the preceding assumptions reduces to:

$$\frac{1}{2} \delta J = \int_I dt \int_V dx \left[-\frac{\partial \tau}{\partial t} - \mathbf{u} \cdot \nabla \tau + \nabla^2 \tau \right] \delta T + \sum_{i,j} \left(-\mathcal{M}_x^i (\delta T(x, t_1)) \right) W_T^{ij} \left(T_d^j - \mathcal{M}_x^j (T(x, t_1)) \right). \tag{A27}$$

We note that the measurement functional acting on δT can be written in the following way:

$$\begin{aligned} \mathcal{M}_x^j (\delta T(x, t_1)) &= \int_V dx' \, M^k(x') \delta T(x', t_1) = \int_V dx' \, M^k(x') \int_V dx \, \delta(x - x') \, \delta T(x, t_1) \\ &= \int_V dx \, \delta T(x, t_1) \int_V dx' \, M^k(x') \, \delta(x - x') = \int_I dt \int_V dx \, \delta T(x, t) \, \mathcal{M}_y^j (\delta(x - y, t - t_1)). \end{aligned} \tag{A28}$$

The performance functional then becomes:

$$\begin{aligned} \frac{1}{2} \delta J &= \int_I dt \int_V dx \left[-\frac{\partial \tau}{\partial t} - \mathbf{u} \cdot \nabla \tau + \nabla^2 \tau + \sum_{i,j} \left(-\mathcal{M}_y^i (\delta(x - y, t - t_1)) \right) W_T^{ij} \left(T_d^j - \mathcal{M}_x^j (T(x, t)) \right) \right] \delta T(x, t) \\ &= \int_I dt \int_V dx \left[-\frac{\partial \tau}{\partial t} - \mathbf{u} \cdot \nabla \tau + \nabla^2 \tau - (\mathcal{M}_{\mathbf{y}} (\delta(x - y, t - t_1)))^T \mathbf{W}_T \epsilon_{\mathbf{d}} \right] \delta T(x, t). \end{aligned} \tag{A29}$$

For arbitrary variations δT, the integral is zero provided the following adjoint equation holds:

$$-\frac{\partial \tau}{\partial t} - \mathbf{u} \cdot \nabla \tau + \nabla^2 \tau - \mathcal{M}_{\mathbf{y}}{}^T (\delta(x - y, t - t_1)) \mathbf{W}_T \epsilon_{\mathbf{d}} = 0. \tag{A30}$$

Gathering the equations that we have derived, along with the natural boundary conditions on the adjoint variable τ, we obtain the coupled Euler–Lagrange system:

$$-\frac{\partial \tau}{\partial t} - \mathbf{u} \cdot \nabla \tau + \nabla^2 \tau = \mathcal{M}_{\mathbf{y}}{}^T (\delta(x - y, t - t_1)) \mathbf{W}_T \epsilon_{\mathbf{d}} \tag{A31}$$

$$\tau(x, t) = 0 \quad \text{on} \quad S, C \tag{A32}$$

$$\tau(x, t_1) = 0 \tag{A33}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T - \nabla^2 T = h + \int_I dt' \int_V dx' \tau(x', t') C_\Theta(x', t', x, t) \tag{A34}$$

$$T(x, t) = T_S(x, t) \quad \text{on} \quad S \tag{A35}$$

$$T(x, t) = T_C(x, t) \quad \text{on} \quad C \tag{A36}$$

$$T(x, t_0) = T_I(x) - \int_V dx' \, C_i(x', x) \tau(x', t_0). \tag{A37}$$

Alternatively, in many applications in the literature that use the strong constraint formulation for the initial condition, the initial condition residual is not explicitly included in the functional J. In that case an expression for the gradient of J with respect to the initial condition is derived and found to be $-\tau(x, t_0)$. The mathematical manipulations to arrive at that conclusion are essentially the same as those given above.