Supplemental Material for:

How to assess similarities and differences between mantle circulation models and Earth using disparate independent observations

J. H. Davies[1], J. Panton[1], I. Altoe[2], M. Andersen[1], P. Béguelin[1], A. Biggin[3], C. Davies[4], T. Elliott[2], Y. A. Engbers[3,10], V. M. Fernandes[5,11], A. M. G. Ferreira[6], S. Fowler[2], S. Ghelichkhan[7], B. J. Heinen[2], P. Koelemeijer[8], F. Latallerie[8], W. Li[9,12], G. Morgan[1], S. J. Mason[4], R. Myhill[2], A. Nowacki[4], N. Récalde[1], C. O'Malley[5,13], A. Plimmer[1], D. Porcelli[8], G. G. Roberts[5], J. B. Rodney[2], J. Shea[9], O. Shorttle[9], W. Sturgeon[6], A. M. Walker[8], J. Ward[4,14], and J. Wookey[2]

[1]School of Earth and Environmental Sci., Cardiff Univ.
[2]School of Earth Sciences, Univ. of Bristol
[3]Dept. of Earth, Ocean and Ecological Sci., Univ. of Liverpool
[4]School of Earth and Environment, Univ. of Leeds
[5]Dept. of Earth Science and Engineering, Imperial College London
[6]Dept. of Earth Sciences, Univ. College London
[7]Research School of Earth Sci., Australian National Univ.
[8]Dept. of Earth Sciences, Univ. of Oxford
[9]Dept. of Earth Sciences, Univ. of Cambridge
[10]Now at Electromagnetic Signatures and Propagation, TNO, The Netherlands
[11]Now at GFZ, Potsdam
[12]Now at Dept. of Earth Sciences, Univ. of Hong Kong
[13]Now at Cathie Group, Newcastle upon Tyne, U.K.
[14]Now at HM Government Statistical Service, U.K.

1 Mantle Circulation Model (MCM)

Here we extend the description provided in Section 2 of the main manuscript to provide more details of the mantle circulation modelling, where an MCM uses plate motion history as the surface velocity boundary conditions. An MCM therefore has plate tectonic-like surface behaviour in locations consistent with geological history on Earth [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] .

The mantle dynamics is simulated by solving numerically the conservation of mass, momentum, energy and composition equations in 3D spherical geometry. The simulations presented here assume a compressible mantle under the anelastic fluid approximation, which approximates mass conservation through the equation:

$$\nabla \cdot (\rho \mathbf{u}) = 0, \tag{1}$$

where ρ is density, and $\mathbf{u}$ is the fluid velocity vector. The equation of motion is:

Table 1: Mantle properties

Parameter	**Symbol**	**Value**	**Units**
Thermal Conductivity	k	3	W K^{-1} m^{-1}
Specific Heat	C$_v$	1100	J kg^{-1} K^{-1}
Reference Viscosity	μ	4×10^{21}	Pa s
Surface Temperature	T$_0$	300	K
Acceleration due to Gravity	g	10	m s^{-2}
Initial Concentration K^{40}	K$_0^{40}$	1.62×10^{-9}	mol g^{-1}
Initial Concentration U^{235}	U$_0^{235}$	1.99×10^{-12}	mol g^{-1}
Initial Concentration U^{238}	U$_0^{238}$	1.01×10^{-10}	mol g^{-1}
Initial Concentration Th^{232}	Th$_0^{232}$	3.48×10^{-10}	mol g^{-1}

Concentrations of heat producing elements given at time of circulation

$$\frac{\partial}{\partial x_j}\left(\eta\left[\dot{\epsilon}_{ij} - \frac{2}{3}\delta_{ij}\frac{\partial u_k}{\partial x_k}\right]\right) - \frac{\partial p}{\partial x_i} = -\Delta\rho' g_r, \tag{2}$$

where η is viscosity, x_j is a spatial co-ordinate, $\dot{\epsilon}_{ij}$ is the strain-rate tensor, p is dynamic pressure, g_r is the radially directed acceleration due to gravity, and where $\Delta\rho'$, the lateral density is:

$$\Delta\rho' = -\alpha\rho_0(T - T_{\mathrm{ref}}) + \Delta\rho_C(C - C_{\mathrm{ref}}) + \chi_T\rho_0 p \tag{3}$$

with α the coefficient of thermal expansion and ρ_0 a reference density. T is temperature and T_{ref} is a radially varying reference temperature, with a constant temperature in the mid-mantle and thermal boundary layers associated with the top and bottom boundaries. C is a scalar variable that represents variations in bulk composition. The bulk composition with C = 0 represents a depleted harzburgitic composition, while C = 1 represents an enriched basaltic composition, and C = 0.2 represents a lherzolite composition. C_{ref} a reference bulk composition, and $\Delta\rho_c$ is a scaling factor that determines how the bulk composition parameter C affects the lateral density anomaly. χ_T is the isothermal compressibility.

Conservation of energy is approximated through the expression:

$$\rho C_p \left(\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) = \nabla \cdot (k \nabla T) + H + \Phi + \alpha T u_i \frac{\partial p}{\partial xi}, \tag{4}$$

where t is time, H is internal heat generation (by radioactive decay), C_p is specific heat capacity, k is thermal conductivity, and Φ is viscous dissipation.

The bulk composition C obeys the conservation equation:

$$\frac{\partial C}{\partial t} = -\nabla \cdot (C\mathbf{u}) + S. \tag{5}$$

where S is the source term representing melting, described next. Near the surface, melting is simulated following the methodology of Van Heck *et al.* [16] using a depth and composition dependent solidus and liquidus, with a linear increase in the degree of melting between these limits. Melting enriches particles (C increases) near the surface and depletes particles (C decreases) in the melting source region. The concentrations of different isotopes are also tracked using particles. These isotopes are fractionated during melting according to the degree of melting and the relevant partition coefficient. Further details on the implementation of the melting process and the parameter values used in these simulations can be found in Van Heck *et al.* [16]. Figure SM1shows the average distribution of the C value of particles beneath ocean basins at present day in the MCM. Assuming either 50% of particles have $C = 1$ or C > 0.5 we estimate mean oceanic crust thickness to be between 4.5 and 9.5 km.

The velocity boundary condition at the core-mantle boundary is free-slip, while the velocities at the surface are prescribed by a model of plate motion history, which has zero radial velocities. Many models of plate motion histories have been used in MCMs [17, 18, 19, 20, 21], here we use the Müller *et al.*[17] model describing 1Ga of plate motion history. This plate model is generated using a joint inversion for multiple constraints on absolute plate motion [22] in order to reconstruct paleo-latitudes and paleo-longitudes in relation to mantle structures. The surface temperature boundary condition is isothermal (Table 1), while the temperature boundary condition at the core-mantle boundary is also isothermal but varies in time, derived from a coupled model for the evolution of the core [23].

The initial temperature field is derived by running a mantle convection simulation for roughly 2 Gyr until the surface heat flux is near steady-state. We then apply the surface velocity field from the model of plate motion history from the 1 Ga time step for 200 Myr, to condition the mantle with structure related to that instant of plate motion history. As the simulation advances, the model has decreasing memory of its initial condition, therefore an initial condition over two mantle over-turn times ago (1 Ga) will have little influence on the present day structure. Figure SM2 shows the average mantle temperature and CMB

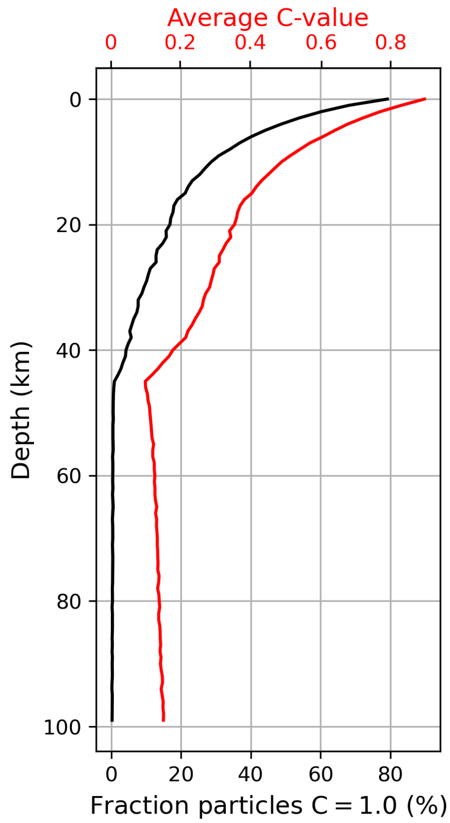

Figure 1: Profiles produced from particles in ocean basins at present day in the MCM, binned at 1 km depth intervals. Black profile shows the fraction of particles with a bulk composition of $C = 1$ and red profile shows the mean bulk composition.

temperature over model time, and also the thermal energy fluxes in and out of the mantle domain.

The initial bulk composition field is designed to mimic a partly processed mantle. Ten particles are distributed evenly in the volume nearest each node, with four particles with $C = 0$, one with $C = 1$, and five with $C = 0.2$. In simulations with a primordial layer, all particles within 150 km of the CMB are set to $C = 2$ at the start of the velocity field conditioning phase. We also track (and evolve by decay) the concentration of heat producing elements and their daughter isotopes. The initial concentrations of these isotopes are calculated following [24], see Table 1.

The properties assumed for the mantle are listed in Table 1. Viscosity varies

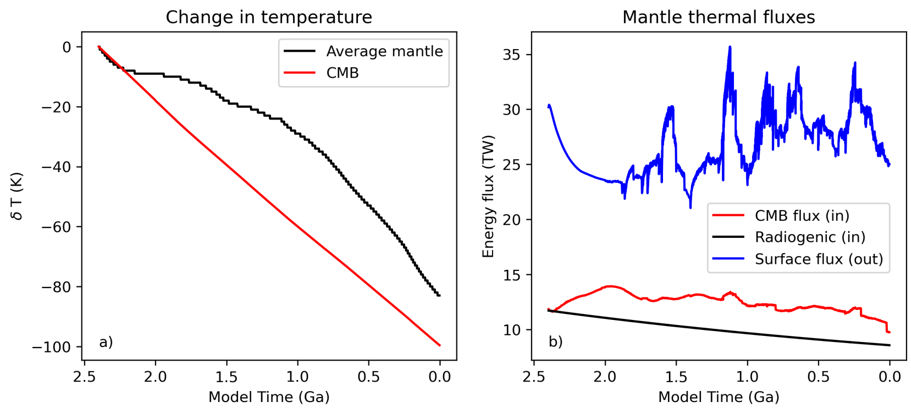

Figure 2: a) Change in average mantle temperature and CMB temperature over model time, b) thermal energy fluxes in and out of the mantle domain.

with both depth and temperature according to:

$$\eta = \eta_z \exp((z'V_a) - (E_a T')) \tag{6}$$

where η is the viscosity at a given node, η_z is the reference viscosity (η_0) multiplied by the radial viscosity factor at depth z, z' is the non-dimensional depth, V_a=1.0 and E_a=2.0 are non-dimensional constants that control the sensitivity of viscosity to depth and temperature, and T' is the non-dimensional temperature. Temperature is non-dimensionalised via $T' = (T - T_0)/(T_c - T0)$, where T is the nodal mantle temperature, T_0 is the temperature of the surface boundary, and T_c is the temperature of the lower boundary at the CMB. We non-dimensionalise depth by $z' = z/h$, where h is the total thickness of the mantle. The reference viscosity profile and viscosity range with depth are plotted in Fig. 3.

We also include the dynamic effects of the phase changes at 410 km and 660 km depth, using the sheet mass anomaly method [25, 26], using the parameters in table 2. We note that this simple sheet mass anomaly method is not strictly consistent in a mantle with laterally varying composition, for example a mantle with an increased fraction of basalt (with its very low olivine content) will reduce the proportion of mantle olivine.

Table 2: Olivine phase change parameters (the density change assumes 67% olivine).

Reference depth (km)	**$\Delta\rho$ ($\mathrm{kg\,m^{-3}}$)**	**Clapeyron slope ($\mathrm{MPa\,K^{-1}}$)**
410	230	2.25
660	380	-1.5

During the simulation, we impose the plate motion history from 1 Ga to the present day in steps of 1 Myr. We adjust the magnitude of the surface

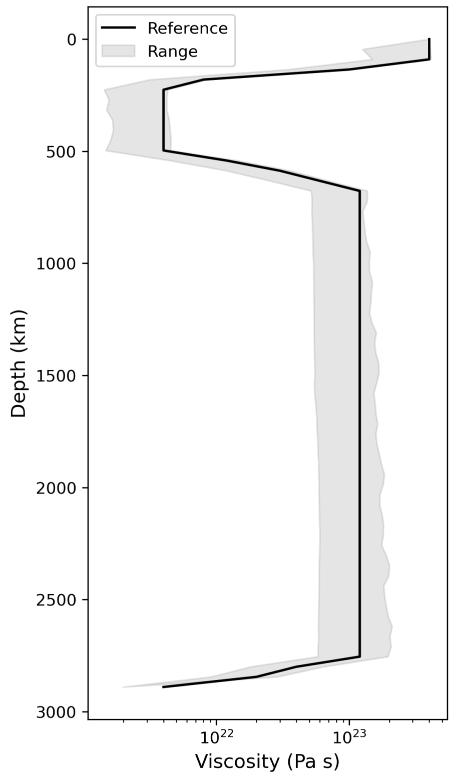

Figure 3: Reference viscosity profile (η_0 multiplied by radial factor, black line) and the viscosity range at each radial layer of simulation (grey shaded area).

velocity to avoid introducing energy into the system when applying the surface velocity boundary condition. We achieve this by first estimating the natural root mean square (r.m.s.) surface velocity in a free-slip surface simulation (without a high viscosity lithosphere) with the same model parameters. We then scale the applied surface velocity from the plate motion history to achieve this mean r.m.s. surface velocity using a constant scaling factor. The length of time over which each plate motion stage is applied, is increased by the same constant scale factor that the velocity is decreased by, this maintains Earth-like subduction and ridge mass fluxes. In the simulation described here, we scale velocities (and radioactive decay constants) by 50 %, thus doubling the time per plate stage.

The equations are solved using the benchmarked [27], parallel [28] code TERRA [29, 26, 30, 31, 3, 32, 16], where the unknown variables, velocity, dynamic pressure and temperature are solved on a structured grid. Each radial layer of the structured grid is based on a regular icosahedron, where 10 diamonds, made from pairing its 20 triangles are iteratively subdivided [33]. The sample simulation presented here has an average lateral resolution at mid-mantle depth of around 45 km, with similar radial spacing. We note that simulations with higher resolution are possible but at significantly higher computational cost, $\propto$ (grid spacing)$^{-4}$. They will allow a broader range of models to be investigated, e.g. including a thinner lower viscosity asthenosphere. The advection of bulk composition and isotope amounts is undertaken by ascribing the properties to particles and tracking their movement using a fourth-order Runge-Kutta scheme [31]. Any properties of the particles required at the node locations are obtained by a distance and mass weighted interpolation of those properties to the node, from the particles nearest to that node. We note that only one value of bulk composition, given by the parameter C, is tracked on each particle, and also on all grid nodes.

2 Producing Isotropic Seismic Structure from Mantle Circulation Models

The compositions assumed for the three independent lithologies used in this study (Baker and Beckett [34] (harzburgite), Walter [35] (lherzolite) and White and Klein [36] (basalt)) are presented in Table 3. For each lithology a separate look-up table of elastic properties (as a function of depth and temperature throughout the mantle) is produced. How these three tables are produced and used to predict the isotropic seismic structure from the thermocompositional simulation is described in the main text.

Table 3: Mole percent oxide compositions for the three independent lithologies used in this study.

	Harzburgite C = 0	**Lherzolite C = 0.2**	**Basalt C = 1**
$\mathbf{SiO_2}$	36.184	38.819	52.298
MgO	56.559	49.894	15.812
FeO	5.954	6.145	7.121
CaO	0.889	2.874	13.027
$\mathbf{Al_2O_3}$	0.492	1.963	9.489
$\mathbf{Na_2O}$	0.001	0.367	2.244

The C value for lherzolite is determined by finding the C-value value that minimizes the least-squares misfit in 6-component oxide space using the compositions reported in table 3. A C value of 0.2 for lherzolite gives a reasonable fit.

As mentioned in the main text, the final effective densities and seismic velocities throughout the domain are calculated by harmonic averaging of the lherzolite, harzburgite and basalt material, weighted by the mass fractions f_i^{M} of each bulk composition. This is physically correct averaging for densities, given as:

$$\frac{1}{\rho} = \frac{\sum_i V_i}{\sum_j M_j} = \sum_i \frac{M_i}{\sum_j M_j} \frac{V_i}{M_i} = \sum_i \frac{f_i^{\mathrm{M}}}{\rho_i} \tag{7}$$

where i and j are indexes for the end-member components, and are summed assuming a mechanical mixture of the appropriate combination (as described in the main text) of the independent compositions (explicitly listed above). We note there is no "correct" choice for multi-lithology (or even multi-phase) averaging of seismic velocities, as in reality this would depend on the textures and polarisation of seismic waves passing through the medium. However, the exact choice of averaging (arithmetic, harmonic, ...) has a relatively small effect on the velocities compared to small-scale variations in the fractions of lithologies in the mechanical mixture.

The thermodynamic dataset used to generate the look-up tables of elastic properties lacks a full covariance matrix, and therefore we cannot propagate uncertainties or calculate confidence bounds on the final effective properties. To obtain a first-order estimate of the uncertainties in the calculated seismic properties, we made several simplifying assumptions. First, we assume that the uncertainty in V_s is entirely dependent on the uncertainty in shear modulus, G. Secondly, we assume that the relative uncertainty in $G(P, T)$ is the same as for G_0, and that the published variances for individual phases are independent. We generated tables of modal phase proportions at regularly spaced P, T conditions between 10 and 130 GPa along both a cool and hot geotherm [37, 38] in the same manner as for the look-up tables. For phases modelled as solid solutions we first propagated the error in a Voigt-Reuss-Hill average of the molar proportions of each endmember phase. We then propagated the error in a Voigt-Reuss-Hill average of the volume proportions of each phase to obtain the uncertainties in G for each assemblage, before converting to a percentage error in V_s. The minimum, maximum, and average relative uncertainties for the three lithologies are presented in table 4.

The uncertainties in the mineral dataset are ultimately dependent on the uncertainties in the original experimental and computational data used in its construction [39]. This leads to increased uncertainties when particular phases are present. We note that for harzburgite, δV_s is highest in the upper mantle and at the base of the lower mantle, likely due to more complex mineralogy and greater uncertainties in post-perovskite respectively. The uncertainty in harzburgite is greater than for lherzholite due to the increased proportion of calcium perovskite. For basalt, the δV_s is more than double in the lower mantle compared to the upper mantle, likely due to a lack of experimental data at these conditions, and the larger error for basalt overall can be attributed to greater uncertainties in the elastic properties of Fe- and Al-endmember phases, Ca-perovskite, and free silica.

Table 4: Estimated uncertainty in V_s for the three lithologies.

	Harzburgite **C = 0**	**Lherzolite** **C = 0.2**	**Basalt** **C = 1**
Min	0.31%	0.32%	0.29%
Max	0.50%	0.48%	1.67%
Mean	0.39%	0.45%	1.04%

3 Testing models with seismic observations

3.1 Whole Mantle

3.1.1 1D Isotropic

The mantle circulation model (MCM) presented in this study has a 45 km vertical resolution. In order to simulate sharp seismic discontinuities, we interpolate temperature bilinearly from the neighbouring grid nodes onto a finer 2.5 km radial grid and interpolate the bulk composition from the particles onto this grid by a nearest neighbour scheme . We then build 1-D seismic profiles from the high-resolution MCM model, by radially averaging ρ, V_S, and V_P, since the icosahedral grid used by MCM distributes grid-nodes almost uniformly across the Earth's surface [33]. The global average is then interpolated at the same depths as PREM.

The comparison between 1D radially-averaged velocities and models such as PREM and AK135 is only intended to be a zeroth-order check on our models. In the main text, we briefly discussed the expected differences between our synthetic models and the models based on the real Earth. These differences have several origins:

Earth model approximations PREM was not designed to represent a 1D radial-average structure of the Earth, but rather the 1D structure that best-fit normal mode, travel time and attenuation data and the Earth's mass and moment of inertia [40]. Uneven ray path coverage and the requirement to fit the width and velocity jump across mantle discontinuities will both result in differences from a true 1D average. Furthermore, the deeper part of the PREM model (>670 km depth) uses the Adam-Williamson equation, which assumes an adiabatic temperature profile between hand-picked layer boundaries, homogeneous bulk composition, no sharp mineralogical discontinuities between layer boundaries and that any continuous reactions are able to reach equilibrium during the passage of a seismic wave.

Thermodynamic modelset approximations The thermodynamic modelsets used in this study are inevitably imperfect, both in terms of formulation and significant uncertainties in some parameter values (Section 2).

MCM model choices Some of the MCM model choices are the things that we want to understand, but other choices may also affect the fit to seismic observations. In the main text we mention the simplified lithosphere in our MCMs, but other pragmatic choices such as the simplified density and chemical parameterisations will certainly also affect the 1D seismic profiles.

While outside the scope of this study, an improved comparison might take an MCM, create synthetic seismic data similar to those used to construct PREM and AK135, and repeat the original inversion procedure on that synthetic data.

3.1.2 3D Long wavelength tomography

Once the MCM is re-parameterised and filtered using S40RTS [41], we can calculate the correlation between the predicted and observed seismic velocities. At each radial spline, we compute the spherical harmonic coefficients of degree l and order m for the seismic velocities predicted by the geodynamic model $\{a_{lm}, b_{lm}\}$ as well as the seismic tomography model $\{c_{lm}, d_{lm}\}$. The correlation per spherical harmonic degree at each radial spline (r^l) is given by:

$$r^l = \frac{\sum_{m=0}^{l} (a_{lm}c_{lm} + b_{lm}d_{lm})}{\sqrt{\sum_{m=0}^{l} (a_{lm}^2 + b_{lm}^2)}\sqrt{\sum_{m=0}^{l} (c_{lm}^2 + d_{lm}^2)}} \tag{8}$$

The total correlation at each radial layer up to degree l_{max} ($r^{tot}_{l_{\max}}$) is then given by:

$$r^{tot}_{l_{\max}} = \frac{\sum_{l=1}^{l_{\max}} \sum_{m=0}^{l} (a_{lm}c_{lm} + b_{lm}d_{lm})}{\sqrt{\sum_{l=1}^{l_{\max}} \sum_{m=0}^{l} (a_{lm}^2 + b_{lm}^2)}\sqrt{\sum_{l=1}^{l_{\max}} \sum_{m=0}^{l} (c_{lm}^2 + d_{lm}^2)}}. \tag{9}$$

Finally, we compute the weighted mean correlation ($\langle r_{l_{\max}} \rangle$) combining all depth splines (z_j) [42]

$$\langle r_{l_{\max}} \rangle = \frac{\sum_{j=1}^{M} w_j r^{t}_{l^{tot}_{\max}} (z_j)}{\sum_{i=1}^{M} w_j} \tag{10}$$

where the weighting $w_j = h_j(R - z_j)^2$ for layer j accounts for the change in area with depth, thus depending on the layer depth (z_j), layer thickness (h_j) and the radius of the Earth R.

3.1.3 Normal mode splitting

Normal modes are long-period oscillations of the whole Earth and thus only sensitive to its long-wavelength structure of the mantle. There are two types of modes; spheroidal modes ${}_nS_l$ and toroidal modes ${}_nT_l$, that are characterised by their radial order n and angular order l. Each mode multiplet consists of $2l + 1$ singlets with azimuthal order m, which all have the same resonance frequency for a spherically symmetric, isotropic, non-rotating Earth model. These singlets have different frequencies, i.e. the degeneracy is removed, in a more realistic Earth, an effect known as splitting. The splitting due to Earth's rotation and ellipticity can be calculated, while the additional splitting that is observed is related to 3D Earth structure.

Splitting function coefficients [43] are conveniently used to describe the splitting of a particular normal mode. Using perturbation theory, these are linearly related to perturbations of a reference Earth model:

$$c_{\mathrm{st}} = \int_0^a \delta m_{\mathrm{st}}(r) K_s(r) dr + \sum_d \delta h^d_{\mathrm{st}} H^d_s \tag{11}$$

where s and t are the angular order s and azimuthal order t describing lateral heterogeneity in the Earth. $K_s(r)$ and H^d_s are the sensitivity kernels associated with the perturbations, computed in the anisotropic PREM model [40]. δm_{st} are the coefficients for perturbations in shear-wave velocity (V_S), compressional-wave velocity (V_P) and density (ρ), while δh_{st} refer to perturbations in topography at internal boundaries. These splitting function coefficients are combined with spherical harmonics to visualise the normal mode splitting, i.e. the variations in resonance frequency of the normal mode. The resulting splitting function maps represent the radially averaged Earth structure, as sensed by a particular normal mode.

To compute synthetic splitting function maps, we first reparameterise the velocity and density structure of the MCM in spherical harmonics for each depth. Using Equation 3.1.3, we then compute the splitting function coefficients of each mode. Here, we restrict ourselves to two groups of modes with a particular sensitivity: high-frequency fundamental spheroidal modes that are primarily sensitive to the upper mantle and core-mantle boundary Stoneley modes that are sensitive to the CMB [44]. For each group, we analyse 10 different modes: fundamental modes ${}_0S_{21} -_0 S_{30}$ as upper mantle sensitive modes and CMB Stoneley modes ${}_1S_{10} -_1 S_{14}$, ${}_2S_{15} -_2 S_{17}$, ${}_2S_{25}$ and ${}_3S_{26}$ as lower mantle sensitive modes. We use the observations of [45, 44], which are publicly available. Some additional examples are given in Figure 4.

To quantitatively compare the predicted and observed splitting functions, we compute both the spectral correlation and the spectral amplitude ratio, both up to the maximum spherical harmonic degree of the observation. By evaluating both of these, we can directly assess whether the predicted heterogeneity is in the right geographic location irrespective of whether the amplitude matches (via the correlation). We can separately assess whether the strength of mantle heterogeneity is correct, even if the structure is not exactly in the right location (via the amplitude ratio).

3.2 Upper Mantle

3.2.1 1D Radial Anisotropy

In order to evaluate the elastic anisotropy at a particular location in the mantle, we trace pathlines back in time through the stored history of mantle flow and record the local velocity gradient tensor at each time step (back to 100 Ma, or when the particle leaves the upper mantle). This extends the approach of [46] by incorporating the time-varying history from the mantle circulation model. The resulting tensor is then scaled by the fraction of deformation calculated to be associated with dislocation creep, following the approach of [47]. We then

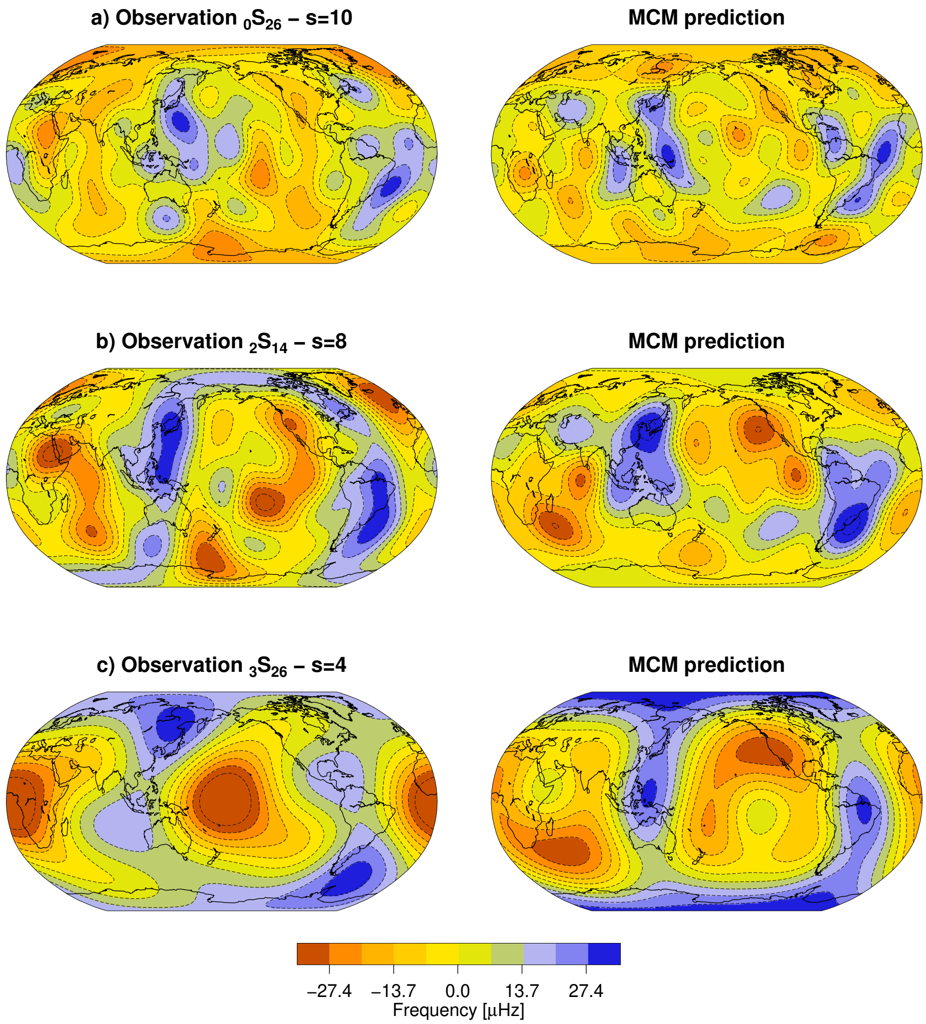

Figure 4: Further examples of predicted splitting functions based on the MCM, compared to observations, for a) mode $_0S_{26}$, b) mode $_2S_{14}$ and c) mode $_3S_{26}$.

provide these tensors as boundary conditions to a model of polycrystalline deformation (DRex [48], implemented in PyDRex [49]) which describes how the imposed macroscopic strain rate (the symmetric part of the velocity gradient tensor) is accommodated by microscopic strain in a collection of (initially random) crystals. We use a 2000-crystal assemblage comprising 70% olivine, 30% enstatite and default deformation parameters [48]. This process is computationally very expensive, limiting the number of total pathlines we can evaluate. We calculate depth profiles between 400 km depth and the surface, with a 25 km resolution. These are generated across 162 equal-area sampling points across the globe.

The result of this process is a set of crystal lattice orientations which can be used together with the single crystal elasticity to compute the macroscopic elasticity of the mantle at each path endpoint. For non-trivial 3D flow, this macroscopic elasticity generally exhibits triclinic symmetry with 21 independent elastic constants. For comparison with observation, we reduce this to a radial anisotropy by azimuthal averaging of the elasticities, and derive the S-wave anisotropic parameter ξ $(= (V_{\mathrm{SH}}/V_{\mathrm{SV}})^2)$ from the resulting wavespeeds.

3.2.2 Phase velocity

The description of utilising phase velocity maps to constrain mantle circulation models (MCMs) is given in the main text. We expand on a few aspects here.

We estimate data errors by clustering ray paths that start and end in the same 5x5° degree bins. The standard deviation of the path measurements within each cluster is used if there are more than 20 paths, otherwise the global standard deviation is used. Ray path density is also weighted in the inversion, to account for the uneven ray density.

Due to the ill-posed nature of the inverse problem, regularisation is applied in the inversions. To determine appropriate amounts of regularisation, we build L-curves that show the trade-off between the data misfit and the amount of norm-damping applied. Whilst it is common to use the 'elbow' of the L-curve, this provided maps that were too smooth and lost crucial detail, consequently we take the misfit of the model with no regularisation applied, multiply it by 110%, and use the corresponding amount of regularisation (Figure SM5). We also compute associated error maps by propagating the estimated model errors and ray density weighting through the inversion. Model errors provide important constraints for assessing the MCM.

In order to predict phase velocity maps for the MCM, we build 1D profiles at every 2 degrees in longitude and latitude. We first carry out a triangular interpolation laterally (on the spherical layers of the MCM) onto the lateral (longitude-latitude) locations of the tomographic grid nodes (voxels' center-points), followed by a linear interpolation with depth. At each location, we define 1D profiles of isotropic Vs and Vp, density, shear attenuation, bulk attenuation and eta. In the mantle Vs, Vp and density are taken from the MCM, whilst shear attenuation, bulk attenuation and eta are set to PREM [40]. The core is fixed to PREM and the crust is set to CRUST1.0 [50].

For completeness and for aiding the interpretation of the results presented in the main text, Fig. SM 6 shows the depth sensitivity kernels with respect to shear-wave speed for the fundamental mode phase velocity data used in this study.

Then to calculate the quantitative misfit between the real and predicted phase velocity maps at each period, we use the following equation:

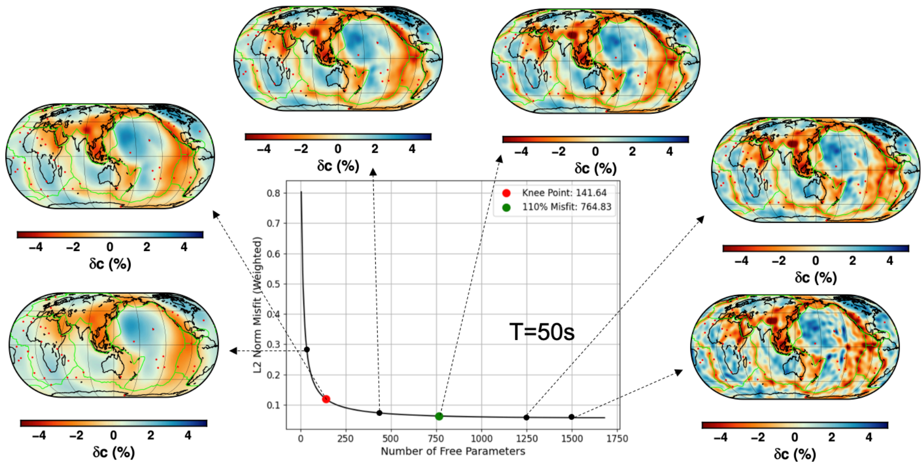

Figure 5: The effects is applying varying level of regularisation in the inversion. An L-curve is shown for an example phase velocity map at an illustrative period of 50 seconds. The number of effective parameters is given by the trace of the resolution matrix, which depends on the regularisation. The red dot shows the location along the L-curve and associated map using the traditional 'elbow' of the L-curve. The green dot shows the location along the L-curve and preferred map with the 110% misfit criteria.

$$\chi_m = \sqrt{\frac{1}{N} \sum_{i \in lat,lon}^{N} \frac{|(m_i^{\text{MCM}}(\omega) - m_i^{\text{Seismic}}(\omega))|^2}{\sigma_i^2(\omega)}} \tag{12}$$

where w is the frequency/period, N is the total number of locations i, which are a function of latitude and longitude, m^{MCM} is the predicted MCM, m^{Seismic} is the data-based seismic phase velocity, and σ^2 is the uncertainty in the seismic phase velocity maps.

3.3 Surface Wave Tomography

While the description of the SOLA surface wave tomography method and how to use it to constrain mantle circulation models is given in the main text, we provide some details on the computation of the misfit here.

Similar to the procedure explained in the previous section, we start by interpolating the MCM velocity predictions onto the tomographic grid. This is achieved first by using a triangular interpolation laterally (on a spherical shell of the MCM grid) to map predicted velocities at the latitudes and longitudes of the center points of the tomographic grid voxcells. Then we do a linear interpolation with depth to map velocity predictions at the depths of the center points of the tomographic voxcells.

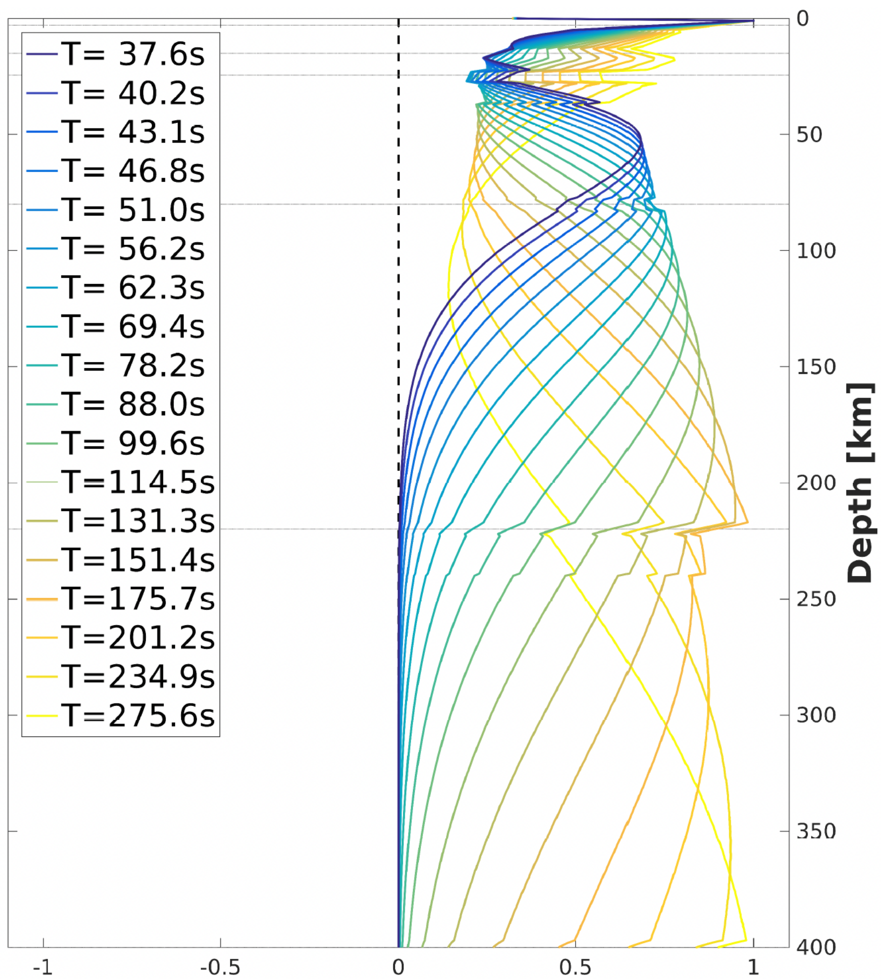

Figure 6: Normalised depth sensitivity kernels of fundamental mode Rayleigh wave phase velocity with wave periods between ~ 38 s and ~ 276 s with respect to shear wave speed for Earth model PREM.

To compute the misfit we first apply the SOLA resolution matrix $\boldsymbol{R}$ to the MCM prediction $\boldsymbol{m}^{\mathrm{MCM}}$, then we compute the misfit χ_m with the data-based tomography model $\boldsymbol{m}^{\mathrm{SOLA}}$, using the equation:

$$\chi_m = \sqrt{\frac{1}{\sum V_k} \sum V_k \frac{[(\boldsymbol{m}^{\mathrm{SOLA}})_k - (\boldsymbol{R}\boldsymbol{m}^{\mathrm{MCM}})_k]^2}{(\boldsymbol{\sigma}_{\boldsymbol{m}}^{\mathrm{SOLA}})_k^2}}, \tag{13}$$

where $\boldsymbol{\sigma}_{\boldsymbol{m}}^{\mathrm{SOLA}}$ are the tomographic uncertainties, k is the model parameter index, and V_k is the volume of grid cell k.

3.4 Other possible seismic constraints

We emphasise that we are only presenting a sub-set of possible disparate constraints on this sample test MCM. We here mention some other possible seismic constraints and methods of testing and constraining MCM models.

MCMs can be tested by predicting body wave traveltime residuals for MCMs [51]. This can be done for example by ray-tracing (e.g. [52, 53]) or with more sophisticated numerical tools such as SPECFEM3D_GLOBE to solve the 3-D wave equation [54, 55, 56]. An intermediate form of calculating traveltimes in MCMs (in terms of physical accuracy as well as computational effort) would be to employ finite-frequency sensitivity kernels (i.e. banana-doughnut kernels) [57]; an example of using this method to investigate MCMs is being prepared for this special issue, by Freissler et al..

Schuberth et al., (2009a,b) [51, 58] compared their predicted seismic structure with seismic tomography models, where they look at the radial profile of root-mean-square amplitudes, histograms of heterogeneity and spectral power and investigate the influence of applying the seismic filter. Schuberth et al. (2012) [54] took this further and used a spectral element method to simulate 3-D global wave propagation and compared the statistics of observed traveltimes with predictions from MCM, highlighting the potential significance of finite-frequency effects. Schuberth et al. (2015) [55] studied the dispersion of traveltime residuals in MCM derived models caused by diffraction as a function of period, observing pronounced dispersion. They discovered that wave-form healing is equally important for fast and slow seismic velocity structures in MCMs. Schuberth et al., (2021) [56] start directly from the variance of the temperature variations in the MCM and using wave propagation modelling compare the predicted traveltime residuals directly with characteristics of observed ones and through this quantify the uncertainties related to anelasticity.

4 Testing models with dynamic topography and geoid observations

Section 6 of the main manuscript summarises the testing of surface deflections and the geoid predicted by the MCM simulation. This section of Supplemental Material, first, extends the description of the calculation of surface deflections and the geoid using spherical harmonics and the propagator matrix technique. Secondly, the description of the methodologies used to assess the fidelity of the predictions are expanded from what is given in the main manuscript.

Following [59], surface deflection for each spherical harmonic coefficient, h_{lm} is calculated such that

$$h_{lm} = \frac{1}{\rho_m - \rho_w} \int_{R_{\mathrm{CMB}}}^{R} A_l \delta\rho_{lm}(r) \mathrm{d}r. \tag{14}$$

The products of the sensitivity kernel, A_l, and density anomalies, $\delta\rho_{lm}$, of spherical harmonic degree, l, and order, m, are integrated with respect to radius, r, between the core-mantle boundary (CMB), and Earth's surface. ρ_m and ρ_w are the mean densities of the surfical layer and overlying fluid, respectively, see e.g. [60, 61, 62] and body text of the main manuscript for more details.

Similarly, the geoid was calculated, such that

$$g_{lm} = \int_{R_{\text{CMB}}}^{R} K_l(r)\delta\rho_{lm}(r)\mathrm{d}r, \tag{15}$$

where K_l is the geoid sensitivity kernel. See [59] for extended methodology.

We expand on the overview of the four approaches used to compare estimates of surface deflections discussed in the main manuscript. As discussed in the main manuscript perhaps the harshest test is to, first, calculate root-mean-squared misfit between predicted surface deflections, h_n, and independent estimates, h_n^o, such that

$$\chi = \sqrt{\frac{1}{N}\sum_{n=1}^{N} w_\phi \left(h_n - h_n^o\right)^2}, \tag{16}$$

where N = number of estimates of surface deflection being compared. In the examples examined in this paper, surface deflection is calculated on a $1 \times 1^\circ$ grid such that $N = 65,341$. The prefactor w_ϕ is included to correct bias in cell size with latitude, ϕ, and is proportional to $\cos\phi$.

Secondly, to aid comparisons of surface deflections as a function of scale they are converted into the frequency domain using spherical harmonics. The degree-correlation spectrum, r_l, is calculated using pyshtools v4.10 [63], such that

$$r_l = \frac{Sf_1f_2}{\sqrt{Sf_1f_1 \cdot Sf_2f_2}} \tag{17}$$

where f_1 and f_2 are the spherical harmonic coefficients of the two estimates of surface deflection being compared. They vary as a function of order, m, and degree, l; $f = f_l^m$. Sf_1f_2 is the cross spectrum of the two functions. We note that $-1 \leq r_l \leq 1$, and we calculate the mean value, $\overline{r_l} = 1/L\sum_{l=1}^{L} r_l$, where L is total number of degrees. Thirdly, the correlation of the entirety of both functions can be estimated following [**?**], such that

$$r = \frac{\sum f_1^* f_2}{\sqrt{\sum f_1^* f_1}\sqrt{\sum f_2^* f_2}}, \quad \text{where} \quad \sum = \sum_{m=-l}^{+l}, \tag{18}$$

where $*$ indicates complex conjugation. This metric is not sensitive to the amplitudes of surface deflections.

Finally, differences in power spectra between between predicted and independent surface deflections are calculated such that

$$\chi_p = \sqrt{\frac{1}{L}\sum_{l=1}^{L}(\log_{10}P_l - \log_{10}P_l^o)^2 + \ldots}, \tag{19}$$

where L is the number of spherical harmonic degrees being considered. $P_l = \sum f_{lm}^2$ is the total power per degree of predicted surface deflections, where $\sum = \sum_{m=-l}^{l}$. P_l^o is total power per degree estimated independently, e.g. from residual oceanic age-depth measurements or Kaula's law [64, 65]. It is straightforward to incorporate multiple spectra into this calculation by simple addition (see Equation 19). Once power spectra are calculated it is straightforward to compare their spectral slopes, which can be used to assess whether broad patterns of surface deflections are similar even if their amplitudes are not.

5 Testing models with Geochemistry and Petrology

5.1 Identifying particles associated with ridges and plumes

Ridges are defined in the plate motion reconstruction [17]. As these are surface features, we project each ridge axis vertically down into the mantle and search for particles which are within 75 km from the projected plane, in a depth range of 135-300 km. This depth range is chosen so that we interrogate particles which are not within the melting zone of the model so that their composition represents the time integrated chemistry rather than modern melt, these particles represent MORB source material in our MCM.

For the detection of plumes we apply a K-means clustering algorithm, from the `sklearn` package [66] for `Python`, to the product of the non-dimensionalised temperature and radial velocity fields. The 'high' value cluster is deemed to be areas of the mantle which are plume-like. We search for plumes at radial model layers from 300 – 2500 km depth - the uppermost and lowermost mantle are omitted due to the difficulty of distinguishing plumes from ridges and lower mantle thermal structures [67]. Individual plumes are identified using a density based clustering approach (`sklearn.cluster.HDBSCAN`), which allows for plume tilt, splitting and merging. Plumes which are detected at a depth of 300 km are projected vertically upwards so that we can extract particles in a depth range of 135-300 km, as was done with the ridges. Particle which fall within this volume represent OIB source material in our MCM. Code for replicating this process is provided at 10.5281/zenodo.13960492.

5.2 Testing models against a geochemical inversion of MORB and OIB radiogenic isotope data

We perform a geochemical inversion of the global MORB and OIB dataset. 1031 MORB samples were compiled from the PetDB database in August 2023 and 1615 OIB samples were compiled from the GEOROC database in February 2024. Modeling is performed using the `NumPy` package [68] for `Python`. The data for six radiogenic isotope ratios ($^{87}Sr/^{86}Sr$, $^{143}Nd/^{144}Nd$, $^{176}Hf/^{177}Hf$, and $^{206,207,208}Pb/^{204}Pb$) for all the compiled samples are mapped into a 6-dimensional boolean array. Elements of this array are set to `True` if they correspond to the 6-isotope ratios composition of one or more samples, with a resolution of `Dataset range / 30` for each ratio, on par with the analytical precision of these measurements.

We run a `Monte Carlo` routine that calculates model isotope compositions for modern basalts through a mantle source evolution model. This model explores the parameter space for 16 variables relevant to the timing and magnitude of mantle source modification, from a primitive mantle (PM) composition [69] (listed in Table 5) at 4.57 Gyr to a basaltic melt at present-day. If model melts have an isotope composition corresponding to natural samples (as recorded by the boolean map), the values for the 16 model variables are logged. We calculate the median parameters leading to each natural sample composition (= `True` element of the boolean array). We then calculate the global MORB and OIB parameters means weighted by sample density (= number of natural samples corresponding to a given boolean array element) and by published plume buoyancy for OIB [70].

The mantle source evolution model used is fully local, with one given set of 16 parameters values corresponding to the full evolution history of a PM source at 4.57 Gyr to the mantle source of a single modern basalt. Note that the Pb concentration of the PM is reduced by 22% at 4.0 Gyr to allow a fit with MORB and OIB $^{206,207,208}Pb/^{204}Pb$. Each mantle source calculated contains two distinct solid components: peridotite and recycled crust, that then melt and mix to yield a basalt at present-day. The trace element and isotope composition of peridotite is modeled through 2 successive events of PM modification, one at time $t_{DM,Per}$ between 4.0 and 2.5 Gyr and one at time t_{dPer} between 2.5 Gyr and 0.5 Gyr. In the first event, a mass proportion $X_{DM,Per}$ between 0.0 and 1.0 of the PM is depleted through modal fractional melting with a melting degree $F_{DM,Per}$ between 0.0 and 0.1, before being re-homogenised with the rest of this local PM source. The second melt-depletion event affects the whole of this re-homogenised source with a modal fractional melting degree F_{dPer} between 0.0 and 0.1. The combined magnitude of these two peridotite depletion events is given by the overall degree of peridotite depletion F_d:

$$F_d = X_{DM,Per}F_{DM,Per} + F_{dPer}(1 - X_{DM,Per}F_{DM,Per}) \qquad (20)$$

At present-day, this model peridotite melts with a modal fractional melting degree F_{Per} between 0.02 and 0.15. This final peridotite melting event is not included in F_d and only serves to correct the mass balance of recycled crust in the final melt mixture, as the recycled crust has a higher degree of modal fractional melting F_{RC}, fixed at 0.65. Peridotite and recycled crust melt with solid/liquid partition coefficients Kd_{Per} and Kd_{RC} [71] (Table 5).

Recycled crust (RC) is modelled as a solid mixture between recycled mafic crust (MC) and recycled continental sediments. As for the peridotite, the source of MC is derived from the PM. For this MC source, a first event of PM modification occurs at time $t_{DM,MC}$ between 4.0 and 2.5 Gyr. A mass proportion $X_{DM,MC}$ between 0.0 and 1.0 of the PM is depleted through modal fractional melting with a melting degree $F_{DM,MC}$ between 0.0 and 0.1, before being re-homogenised with the rest of this local PM source. This re-homogenised source then melts at time t_{RC} between 2.5 Gyr and 0.5 Gyr with a modal fractional melting degree $F_{DM,MC}$ between 0.01 and 0.1. The resulting melt is the MC, which then gets altered (also at time t_{RC}) with an extent of alteration f_{Alt} between 0.0 and 1.0. This process models the addition of Rb and U to the MC by seawater before recycling into the mantle. The elemental budgets B_{Alt} [72] corresponding to f_{Alt} are in Table 5. Alteration changes magmatic elemental concentrations C_{MC0} to C_{MCAlt} through the following equation:

$$C_{MCAlt} = C_{MC0} + B_{Alt} f_{Alt} \tag{21}$$

Continental sediments are then added to MC to form RC with a mass proportion f_{Sed} between 0.0 and 0.1. Sediments are derived from a PM source at 4.57 Gyr that takes the average continental crust (CC) composition of [73] (see Table 5) at time t_{CC} between 4.0 and 2.5 Gyr. Note than the Th concentration of the CC is increased by 20% from the published value to allow a fit with MORB and OIB $^{208}Pb/^{204}Pb$. The CC source then takes the the global subducting sediment ($GLOSS$) composition of [74] (see Table 5) at time t_{RC}. The recycled crust RC (= altered MC + sediments) then gets dehydrated at t_{RC} with an extent of dehydration f_{Dhy} between 0.0 and 1.0 with the elemental mass loss ratios R_{Dhy} [75] listed in Table 5. This process models how fluid loss during subduction changes RC trace element abundances C_{RC0} to C_{RCDhy}:

$$C_{RCDhy} = C_{RC0}(1 - R_{Dhy} f_{Dhy}) \tag{22}$$

RC is mixed with peridotite with a mass proportion f_{RC} between 0.00 and 0.15. f_{RC} results are discussed in the main text, and the detailed results of this geochemical model will be discussed in an upcoming publication (Béguelin et al., in prep.).

Table 5: Reservoir compositions and budgets for the geochemical inversion

Element	**PM**	**CC**	**GLOSS**	B_{Alt}	R_{Dhy}	Kd_{Per}	Fd_{RC}
	$\mu g/g$	$\mu g/g$	$\mu g/g$	$+\mu g/g$	m_{ratio}	C_{sol}/C_{liq}	C_{sol}/C_{liq}
Rb	0.635	49	57.2	11.65	0.65	0.000321	0.003
Sr	21.1	320	327	0	0.408	0.031	0.0513
Sm	0.444	3.9	5.78	0	0.136	0.055	0.26122
Nd	1.354	20	27	0	0.309	0.03	0.14813
Lu	0.074	0.3	0.413	0	0.0	0.438	2.2911
Hf	0.309	3.7	4.06	0	0.136	0.061	0.2688
U	0.021	1.3	1.68	0.296	0.291	0.005	0.008404
Th	0.085	6.72	6.91	0	0.377	0.003	0.004646
Pb	0.185	11	19.9	0	0.846	0.005	0.04236

In the MCM, the f_{RC} value of a population of particles is calculated from the C values of these particles (one f_{RC} value per population using the following equation):

$$f_{RC} = \frac{\left(\frac{1}{m}\sum C_i^{>0.2} - 0.2\right)}{0.8}\left(\frac{m}{n}\right) \tag{23}$$

Where n is the total number of particles in a population, m is the number of these particles with $C > 0.2$, and $C_i^{>0.2}$ is the C value of an individual particle with $C > 0.2$.

To assess whether the early compositional heterogeneity of the mantle affects the distribution of heterogeneity later on, we run MCMs (not presented here) with three very different starting mixtures at 1 Ga, (a) 0% crustal material, (b) 10% crustal material (reference case), and (c) 20% crustal material. We then calculate the resulting vertical distribution of heterogeneity at present day recorded by the plume-ridge difference in recycled crustal material, as is done in Section 7a of the main text. The corresponding results of excess crustal material in plumes compared to ridges at present day are (a) 1.1% $\pm$0.9%, (b) 1.1% $\pm$1.2%, and (c) 1.9% $\pm$1.4%. The corresponding value obtained from a geochemical inversion of the MORB and OIB radiogenic isotope data that takes into account 4.57 Ga of Earth history is an excess of 1.3% (Section 7a). These results demonstrate (i) the MCM reaches a steady state in terms of distribution of mantle heterogeneity in less than 1 Ga and (ii) is equivalent to an independent geochemistry-based estimate. This resulting state at present-day is independent of the starting mixture at 1 Ga.

We note that the inter-plumes standard deviation (quoted in the main text) is the variability (or range) of the data, and not the uncertainty of the mean. To assess how robust the comparison between the MCM and Monte Carlo petrology model is, we compare the $\Delta f_{RC}^{\mathrm{MCM}}$ for 8 MCM runs (not presented here) that use the same thermal and compositional density parameters. We find a mean of $\Delta f_{RC}^{\mathrm{MCM}} = 1.7\% \pm 0.9\%$, meaning *this* result (i.e. the difference in recycled oceanic crust beneath OIB and MORB sources) is reproducible across MCM runs with similar inputs. Expectedly, MCM runs with different thermal and compositional density parameters yield different $\Delta f_{RC}^{\mathrm{MCM}}$ values across a range an order of magnitude larger than our reported uncertainty of ±0.9%. This means comparing $\Delta f_{RC}^{\mathrm{MCM}}$ to the robust $\Delta f_{RC}^{\mathrm{Geochem}}$ value is a useful tool to constrain thermal and compositional density input parameters of MCMs. We note that other aspects of geochemistry can be expected to differ between a model (like the Monte Carlo petrology model above) that relates to Earth's whole temporal evolution, and an MCM which simulates only part of the history (e.g. 1 Ga here).

References

[1] Bunge HP, Richards MA, Lithgow-Bertelloni C, Baumgardner JR, Grand SP, Romanowicz BA. 1998 Time scales and heterogeneous structure in geodynamic Earth models. *Science* **280**, 91–95.

[2] Davies JH, Bunge HP. 2001 Seismically "fast" geodynamic mantle models. *Geophys. Res. Lett.* **28**, 73–76.

[3] Bunge HP, Richards MA, Baumgardner JR. 2002 Mantle–circulation models with sequential data assimilation: inferring present–day mantle structure from plate–motion histories. *Phil. Trans. Roy. Soc. Lond. Series A* **360**, 2545–2567.

[4] McNamara AK, Zhong S. 2005 Thermochemical structures beneath Africa and the Pacific Ocean. *Nature* **437**, 1136–1139.

[5] Davies DR, Goes S, Davies J, Schuberth B, Bunge HP, Ritsema J. 2012 Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity. *Earth Planet. Sci. Lett.* **353-354**, 253–269. (10.1016/j.epsl.2012.08.016)

[6] Bower DJ, Gurnis M, Seton M. 2013 Lower mantle structure from paleogeographically constrained dynamic Earth models. *Geochem. Geophys. Geosys.* **14**, 44–63.

[7] Flament N, Williams S, Müller RD, Gurnis M, Bower D. 2017 Origin and evolution of the deep thermochemical structure beneath Eurasia. *Nature Communications* **8**, 14164.

[8] Cao X, Flament N, Müller RD. 2021 Coupled evolution of plate tectonics and basal mantle structure. *Geochem. Geophys. Geosys.* **22**, e2020GC009244.

[9] Zhang N, Zhong S. 2011 Heat fluxes at the Earth's surface and core-mantle boundary since Pangea formation and their implications for the geomagnetic superchrons. *Earth Planet. Sci. Lett.* **306**, 205–216.

[10] Li M, Black B, Zhong S, Manga M, Rudolph M, Olson P. 2016 Quantifying melt production and degassing rate at mid-ocean ridges from global mantle convection models with plate motion history. *Geochem. Geophys. Geosys.* **17**, 2884–2904.

[11] Li M, Zhong S. 2017 The source location of mantle plumes from 3D spherical models of mantle convection. *Earth Planet Sci. Lett.* **478**, 47–57.

[12] Li M, Zhong S, Olson P. 2018 Linking lowermost mantle structure, core-mantle boundary heat flux and mantle plume formation. *Phys. Earth Planet Int.* **277**, 10–29.

[13] Li M, Zhong S. 2019 Lateral motion of mantle plumes in 3-D geodynamic models. *Geophys. Res. Lett.* **46**, 4685–4693.

[14] Olson P, Deguen R, Rudolph M, Zhong S. 2015 Core evolution driven by mantle global circulation. *Phys. Earth Planet. Int.* **243**, 44–55.

[15] Zhang N, Zhong S, Leng W, Li Z. 2010 A model for the evolution of the Earth's mantle structure since the Early Paleozoic. *J. Geophys. Res.* **115**, B06401.

[16] Van Heck HJ, Davies JH, Elliott T, Porcelli D. 2016 Global-scale modelling of melting and isotopic evolution of Earth's mantle: Melting modules for TERRA. *Geoscientific Model Development* **9**, 1399–1411. (10.5194/gmd-9-1399-2016)

[17] Müller RD, Flament N, Cannon J, Tetley MG, Williams SE, Cao X, Bodur OF, Zahirovic S, Merdith A. 2022 A tectonic-rules-based mantle reference frame since 1 billion years ago – implications for supercontinent cycles and plate–mantle system evolution. *Solid Earth* **13**, 1127–1159. (10.5194/se-13-1127-2022)

[18] Merdith AS, Williams SE, Collins AS, Tetley MG, Mulder JA, Blades ML, Young A, Armistead SE, Cannon J, Zahirovic S, Müller RD. 2021 Extending full-plate tectonic models into deep time: Linking the Neoproterozoic and the Phanerozoic. *Earth-Science Reviews* **214**, 103477.

[19] Matthews KJ, Maloney KT, Zahirovic S, Williams SE, Seton M, Müller RD. 2016 Global plate boundary evolution and kinematics since the late Paleozoic. *Global and Planetary Change* **146**, 226–250. (https://doi.org/10.1016/j.gloplacha.2016.10.002)

[20] Seton M, Müller R, Zahirovic S, Gaina C, Torsvik T, Shephard G, Talsma A, Gurnis M, Turner M, Maus S, Chandler M. 2012 Global continental and ocean basin reconstructions since 200Ma. *Earth-Science Reviews* **113**, 212–270. (https://doi.org/10.1016/j.earscirev.2012.03.002)

[21] Lithgow-Bertelloni C, Richards MA. 1998 The dynamics of Cenozoic and Mesozoic plate motions. *Reviews of Geophysics* **36**, 27–78. (https://doi.org/10.1029/97RG02282)

[22] Tetley MG, Williams SE, Gurnis M, Flament N, Müller RD. 2019 Constraining Absolute Plate Motions Since the Triassic. *J. Geophys. Res.: Solid Earth* **124**, 7231–7258. (10.1029/2019JB017442)

[23] Davies CJ. 2015 Cooling history of Earth's core with high thermal conductivity. *Phys. Earth Planet. Int.* **247**, 65–79.

[24] Panton J, Davies JH, Elliott T, Andersen M, Porcelli D, Price MG. 2022 Investigating influences on the Pb pseudo-isochron using three-dimensional dantle convection models with a continental reservoir. *Geochem. Geophys. Geosys.* **23**, e2021GC010309.

[25] Tackley P, Stevenson D, Glatzmaier G, Schubert G. 1993 Effects of an endothermic phase transition at 670 km depth in a spherical model of convection in the Earth's mantle. *Nature* **361**, 699–704.

[26] Bunge HP, Richards MA, Baumgardner JR. 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.: Solid Earth* **102**, 11991–12007.

[27] Davies DR, Davies JH, Bollada PC, Hassan O, Morgan K, Nithiarasu P. 2013 A hierarchical mesh refinement technique for global 3-D spherical mantle convection modelling. *Geoscientific Model Development* **6**, 1095–1107. (10.5194/gmd-6-1095-2013)

[28] Bunge HP, Baumgardner JR. 1995 Mantle convection modeling on parallel virtual machines. *Computers in Physics* **9**, 207–215. (10.1063/1.168525)

[29] Baumgardner JR. 1985 Three-dimensional treatment of convective flow in the earth's mantle. *Journal of Statistical Physics* **39**, 501–511. (10.1007/BF01008348)

[30] Yang WS, Baumgardner JR. 2000 A matrix-dependent transfer multigrid method for strongly variable viscosity infinite Prandtl number thermal convection. *Geophys. & Astrophys. Fluid Dyn.* **92**, 151–195. (10.1080/03091920008203715)

[31] Stegman DR, Richards MA, Baumgardner JR. 2002 Effects of depth-dependent viscosity and plate motions on maintaining a relatively uniform mid-ocean ridge basalt reservoir in whole mantle flow. *J. Geophys. Res.: Solid Earth* **107**. (10.1029/2001JB000192)

[32] Walzer U, Hendel R, Köstler C, Müller M, Kley J, Viereck-Götte L. 2013 A forward model of mantle convection with evolving continents and a model

of the Andean subduction orogen. In Nagel WE, et al., editors, *High Performance Computing in Science and Engineering '12* , pp. 473–501. Berlin: Springer.

[33] Baumgardner JR, Frederickson PO. 1985 Icosahedral Discretization of the Two-Sphere. *SIAM Journal on Numerical Analysis* **22**, 1107–1115. (10.1137/0722066)

[34] Baker MB, Beckett JR. 1999 The origin of abyssal peridotites: a reinterpretation of constraints based on primary bulk compositions. *Earth Planet. Sci. Lett.* **171**, 49–61.

[35] Walter MJ. 2003 2.08 - Melt Extraction and Compositional Variability in Mantle Lithosphere. In Holland HD, Turekian KK, editors, *Treatise on Geochemistry* , pp. 363–394. Oxford: Pergamon.

[36] White WM, Klein EM. 2014 4.13 - Composition of the Oceanic Crust. In Holland HD, Turekian KK, editors, *Treatise on Geochemistry (Second Edition)* , pp. 457–496. Oxford: Elsevier.

[37] Brown J, Shankland T. 1981 Thermodynamic parameters in the Earth as determined from seismic profiles. *Geophys. J. Int.* **66**, 579–596.

[38] Anderson O. 1982 The Earth's core and the phase diagram of iron. *Philos. T. Roy. Soc. A* **306**, 21–25.

[39] Stixrude L, Lithgow-Bertelloni C. 2021 Thermal expansivity, heat capacity and bulk modulus of the mantle. *Geophys. J. Int.* **228**, 1119–1149. (10.1093/gji/ggab394)

[40] Dziewonski AM, Anderson DL. 1981 Preliminary reference Earth model. *Phys. Earth Planet. Int.* **25**, 297–356.

[41] Ritsema J, Deuss A, van Heijst HJ, Woodhouse JH. 2011 S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophys. J. Int.* **184**, 1223–1236.

[42] Peng D, Liu L. 2022 Quantifying slab sinking rates using global geodynamic models with data-assimilation. *Earth-Science Reviews* **230**, 104039. (10.1016/j.earscirev.2022.104039)

[43] Woodhouse J. 1985 Inversion for the splitting function of isolated low order normal mode multiplets. *Eos Trans. AGU* **66**, 300.

[44] Koelemeijer P, Deuss A, Ritsema J. 2013 Observations of core-mantle boundary Stoneley modes. *Geophys. Res. Lett.* **40**, 2557–2561.

[45] Deuss A, Ritsema J, van Heijst H. 2013 A new catalogue of normal-mode splitting function measurements up to 10 mHz. *Geophys. J. Int.* **193**, 920–937.

[46] Walker A, Forte A, Wookey J, Nowacki A, Kendall JM. 2011 Elastic anisotropy of D$''$ predicted from global models of mantle flow. *Geochem. Geophys. Geosys.* **12**, Q10006.

[47] Kendall E, Faccenda M, Ferreira AMG, Chang S. 2022 On the relationship between oceanic plate speed, tectonic stress, and seismic anisotropy. *Geophys. Res. Lett.* **49**, e2022GL097795.

[48] Kaminski E, Ribe NM, Browaeys JT. 2004 D-Rex, a program for calculation of seismic anisotropy due to crystal lattice preferred orientation in the convective upper mantle. *Geophys. J. Int.* **158**, 744–752. (10.1111/j.1365-246x.2004.02308.x)

[49] Bilton L, Duvernay T, Davies DR, Eakin CM. 2025 PyDRex: predicting crystallographic preferred orientation in peridotites under steady-state and time-dependent strain. *Geophysical Journal International* **241**, 35–57. (10.1093/gji/ggaf026)

[50] Laske G, Masters G, Ma Z, Pasyanos M. 2013 Update on CRUST1. 0—A 1-degree global model of Earth's crust. In *Geophysical research abstracts* vol. 15 p. 2658.

[51] Schuberth BSA, Bunge HP, Ritsema J. 2009 Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone?. *Geochem. Geophys. Geosys.* **10**. (https://doi.org/10.1029/2009GC002401)

[52] Bunge HP, Davies J. 2001 Tomographic images of a mantle circulation model. *Geophys. Res. Lett.* **28**, 77–80.

[53] Freissler R, Zaroli C, Lambotte S, Schuberth B. 2020 Tomographic filtering via the generalized inverse: A way to account for seismic data uncertainty. *Geophys. J. Int.* **223**, 254–269.

[54] Schuberth B, Zaroli C, Nolet G. 2012 Synthetic seismograms for a synthetic Earth: long-period P- and S-wave traveltime variations can be explained by temperature alone. *Geophys. J. Int.* **188(3)**, 1393–1412.

[55] Schuberth B, Zaroli C, Nolet G. 2015 Traveltime dispersion in an isotropic elastic mantle: Strong lower mantle signal in differential-frequency residuals. *Geophys. J. Int.* **203**, 2099–2118.

[56] Schuberth BSA, Bigalke T. 2021 From Mantle Convection to Seismic Observations. In *Mantle Convection and Surface Expressions* , pp. 97–119. American Geophysical Union.

[57] Dahlen F, Hung SH, Nolet G. 2000 Fréchet kernels for finite-frequency traveltimes—I. Theory. *Geophys. J. Int.* **141**, 157–174.

[58] Schuberth BSA, Bunge HP, Steinle-Neumann G, Moder C, Oeser J. 2009 Thermal versus elastic heterogeneity in high-resolution mantle circulation models with pyrolite composition: High plume excess temperatures in the lowermost mantle. *Geochem. Geophys. Geosys.* **10**.

[59] Ghelichkhan S, Bunge HP, Oeser J. 2021 Global mantle flow retrodictions for the early Cenozoic using an adjoint method: Evolving dynamic topographies, deep mantle structures, flow trajectories and sublithospheric stresses. *Geophys. J. Int.* **226**, 1432–1460.

[60] Parsons B, Daly S. 1983 The relationship between surface topography, gravity anomalies and temperature structure of convection. *J. Geophys. Res.* **88**, 1129–1144.

[61] McKenzie D. 1977 Surface deformation, gravity anomalies and convection. *Geophysical Journal of the Royal Astronomical Society* **48**, 211–238. (10.1111/j.1365-246X.1977.tb01297.x)

[62] Ricard Y. 2015 Physics of Mantle Convection. In Schubert G, editor, *Treatise on Geophysics* , pp. 23–71. Elsevier B.V. (10.1016/B978-044452748-6.00115-2)

[63] Wieczorek MA, Meschede M. 2018 SHTools: Tools for Working with Spherical Harmonics. *Geochem. Geophys. Geosys.* **19**, 1–19. (10.1029/2018GC007529)

[64] Hoggard MJ, White N, Al-Attar D. 2016 Suppl. Info. for "Global dynamic topography observations reveal limited influence of large-scale mantle flow". *Nature Geosci.* **9**, 1–34.

[65] Holdt MC, White NJ, Stephenson SN, Conway-Jones BW. 2022 Densely sampled global dynamic topographic observations and their significance. *J. Geophys. Res.: Solid Earth* **127**, 1–32.

[66] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. 2011 Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research* **12**, 2825–2830.

[67] Hassan R, Flament N, Gurnis M, Bower DJ, Müller D. 2015 Provenance of plumes in global convection models. *Geochem. Geophys. Geosys.* **16**, 1465–1489. Publisher: John Wiley & Sons, Ltd.

[68] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. 2020 Array programming with NumPy. *Nature* **585**, 357–362. (10.1038/s41586-020-2649-2)

[69] Sun SS, McDonough WF. 1989 Chemical and isotopic systematics of oceanic basalts: implications for mantle composition and processes. *Geol. Soc. Lond. Spec. Pub.* **42**, 313–345.

[70] Hoggard MJ, Parnell-Turner R, White N. 2020 Hotspots and mantle plumes revisited: Towards reconciling the mantle heat transfer discrepancy. *Earth Planet. Sci. Lett.* **542**, 116317.

[71] Stracke A, Bourdon B. 2009 The importance of melt extraction for tracing mantle heterogeneity. *Geochimica et Cosmochimica Acta* **73**, 218–238.

[72] Kelley KA, Plank T, Ludden J, Staudigel H. 2003 Composition of altered oceanic crust at ODP Sites 801 and 1149. *Geochem. Geophys. Geosys.* **4**.

[73] Rudnick R, Gao S. 2003 Major elements of Earth crust. *Treatise on Geochemistry* **3**.

[74] Plank T, Langmuir CH. 1998 The chemical composition of subducting sediment and its consequences for the crust and mantle. *Chem. Geol.* **145**, 325–394.

[75] Kogiso T, Tatsumi Y, Nakano S. 1997 Trace element transport during dehydration processes in the subducted oceanic crust: 1. Experiments and implications for the origin of ocean island basalts. *Earth Planet. Sci. Lett.* **148**, 193–205.