Geophysical Research Letters

Supporting Information for

Sub-lithospheric small-scale convection tomographically imaged beneath the Pacific plate

Zachary C. Eilon[1], Lun Zhang[1], James B. Gaherty[2], Donald W. Forsyth[3], and Joshua B. Russell[3]

[1] University of California Santa Barbara, Santa Barbara, CA
[2] Northern Arizona University, Flagstaff, AZ
[3] Brown University, Providence, RI
Corresponding author: Zachary Eilon (eilon@ucsb.edu)

Contents of this file

- Text S1
- Figures S1 to S10

Introduction

This supporting information includes additional information about methods useful for reproducing (but not essential for interpreting) the main article. Specifically, it includes:

- Details of the approach and parameters used for body wave differential travel time measurements
- A discussion of combining vertical and pressure channel measurements
- The tomographic model setup, parameter space, and inverse problem details
- The framework for determining regularization parameters for the inversion
- Details (in addition to those within the main text) of various tests of the model output and the data requirements of the model space
- A description of how we translated between velocity and temperature through forward calculations that rely on (an)elastic constitutive relationships.
- Math behind our calculation of gravity-topography admittance and coherence, including error analysis
- Math behind our calculation of 1.5D gravity anomalies from a 2.5D velocity model.

This file also contains 10 figures which assist the reader in interpretation of our results, and which contain additional tests to allow the reader to make their own judgments about our models.

Text S1

Body wave travel time measurements:

For each event, we measured relative arrival times of direct *P*-waves using multi-channel cross-correlation (MCCC; Fig. S1) (VanDecar & Crosson, 1990). This method uses a least-square inversion scheme to compute relative arrival times from cross-correlation pairs, reducing multi-path effects as well as the strict requirement for waveform similarity. As a default, we filtered the vertical and pressure records to 0.3-0.6 Hz and 0.4-2 Hz respectively to avoid the effects of noise (microseisms, anomalous sensor noise, etc.). We then used an interactive GUI to adaptively adjust the time window (nominally spanning 3 seconds before to 5 seconds after the first break) and filter frequencies to maximize the prominent direct *P*-wave signal in the cross-correlation. We rejected traces based on several criteria including low signal-to-noise ratio (rejected if < 3), anomalous *P*-wave amplitude (rejected if <0.1x or >10x the event mean), and similarity with reference waveform (determined by visual check and cross-correlation coefficient, the latter of which was used to weight the MCCC inversion).

Combining pressure and vertical record travel time measurements:

We tested several approaches for combining these two related single-channel datasets (including simple averaging, simple concatenation, and least squares re-computation across multicomponent pairwise measurements). All approaches produced a travel time dataset that yielded extremely similar tomographic results, so we chose the following combination scheme: If both vertical and pressure measurements were available for a certain event, then only measurements from the channel with more usable traces were retained.

Tomographic model setup

Our differential travel time tomography used a first Fresnel zone paraxial approximation to the Born theoretical kernel (Schmandt & Humphreys, 2010), with updated normalization and voxel-volume terms (Brunsvik et al., 2021), to account for finite frequency sensitivity of travel times to 3-D slowness structure. Finite frequency kernels were constructed by interpolating the ray-normal first Fresnel zone radius at 5 km increments along the ray path. We used 1-D ray tracing through the IASP91 reference model (Kennet & Engdahl, 1991) and accounted for station elevations from multibeam depth soundings (although these elevation corrections have minor effect on the travel times).

The model space comprised a cartesian grid rotated ϕ degrees clockwise from north, with horizontal node spacing of 30 km, extending 300 km in all directions beyond the limits of the OBS array. We tested various values of ϕ for 2-D inversions (see below) obtaining a preferred value of ϕ=25˚ (this value was used for consistency with the gravity and 2.5-D inversions, but the choice

of coordinate rotation has almost zero effect on the 3-D inversions). Vertical nodes were spaced every 30-40 km between 40 and 300 km depth, with an additional shallow layer of nodes at 6 km to absorb shallow structure (in addition to the station terms). These depth bounds were chosen to approximately match station spacing (~40 km) and array aperture (~300 km considering station dropouts), respectively.

To regularize the inverse problem we applied both model norm damping and first derivative damping (*i.e.*, "flattening"), minimizing the following cost function:

$$E = \|\mathbf{W}(\mathbf{G}\mathbf{m} - \mathbf{d}_{obs})\|^2 + \gamma\|\boldsymbol{\nabla}\mathbf{m}_0\|^2 + \varepsilon\|\mathbf{m}_0\|^2 + \varepsilon_{evt}\|\mathbf{c}_e\|^2 + \varepsilon_{sta}\|\mathbf{c}_s\|^2$$

Where $\mathbf{G} = [\mathbf{G}_0\ \mathbf{G}_e\ \mathbf{G}_s]$, $\mathbf{m} = \begin{bmatrix}\mathbf{m}_0\\ \mathbf{c}_e\\ \mathbf{c}_s\end{bmatrix}$. In these expressions, $\mathbf{m}_0$ is the vector of fractional perturbations to the initial slowness model, $\mathbf{c}_\mathrm{e}$ is the N_{evts}-length vector of event static times, $\mathbf{c}_s$ is the N_{stas}- length vector of station static times, $\mathbf{d}_{obs}$ is the N_{obs}-length vector of differential travel times, and $\mathbf{G}_0$ is the data kernel matrix with $[G_0]_{ij} = \partial d_i/\partial[m_0]_j$, $\mathbf{G}_e$ is the matrix with $[G_e]_{ik} = \delta_{e(i),k}$ (where $e(i)$ represents the index of the event corresponding to i_{th} seismic ray), $\mathbf{G}_s$ is the matrix with $[G_s]_{il} = \delta_{s(i),l}$ (where $s(i)$ represents the index of the receiver station corresponding to i_{th} seismic ray), $\mathbf{W}$ is a N_{obs}-square diagonal matrix of data weights proportional to the inverse of the standard deviations (σ_d) estimated *a posteriori* during MCCC process. Since estimated differential travel time uncertainty is sometimes unreasonably small, we set a minimum standard deviation of 0.625s, equal to 1/20 of the central filter period. $\boldsymbol{\nabla}$ is the first derivative operator, γ is the smoothing parameter, ε is the damping parameter, and ε_{evt}, ε_{sta} are the damping parameters for event and station static times. Optimal regularization parameter values were determined by L-test (see below), and we fixed ε_{evt} and ε_{sta} to be 0.01 and 5 respectively. Finally, in order to avoid edge effects, damping in the shallowest (≤40 km) and deepest (300 km) layers was increased by a factor of 1.5 compared to the rest of the model volume. For our preferred model, the final RMS error was 0.23s, the RMS of $\mathbf{c}_e$ values was 0.10 s and the RMS of $\boldsymbol{c}_s$ values was 0.01 s.

We used the weighted variance reduction as another measure of the goodness of data fit, computed as $wvr = 100\left(1 - \frac{var(\mathbf{W}[\mathbf{G}\mathbf{m}-\mathbf{d}_{obs}])}{var(\mathbf{W}\mathbf{d}_{obs})}\right)$, where $var()$ is the variance operator. Hypothetically, perfectly fit data would have a wvr =100% (in practice this is impossible, due to irreducible date noise) and wholly un-fit data would have a wvr = 0%, or <0% if the (spurious) model actually worsens the data fit.

Optimizing Tomographic Regularization choices

The regularization was designed to balance data fit, model norm, and model roughness. To achieve this, we used an L-test to search for optimal values of ε and γ, grid searching in the range

0-10 for both parameters. We introduced the term "X", capturing the combination of normalized model norm and roughness:

$$X = \left(\frac{1}{3}\right)\frac{||\boldsymbol{m}||_2}{max\left(||\boldsymbol{m}||_2\right)} + \left(\frac{2}{3}\right)\frac{||\nabla \boldsymbol{m}||_2}{max\left(||\nabla \boldsymbol{m}||_2\right)}$$

where $\left|||\boldsymbol{x}|\right|_2$ represents the L2 norm, ∇ is the gradient matrix, and the 1:2 weighting was determined *ad hoc* to produce reasonable looking models. We analyzed the trade-off between model roughness and weighted data variance reduction, wvr, (Figure S3) by minimizing the penalty function $P = 100X - wvr$ (where the factor of 100 normalizes both terms to the range 0-100). These tests yielded preferred regularization parameters of $\varepsilon = \gamma = 4$ (Figure S3).

Testing the model space: 2.5D inversions, recovery tests, and squeezing tests

"2.5-D" inversions were implemented by zero model variation (through heavy smoothing) along the horizontal direction perpendicular to ϕ. We then grid searched through possible values of ϕ in increments of 5˚, seeking the direction that provided the greatest reduction in data misfit (Fig. 3b and Fig. S7).

For synthetic recovery (input-output) tests, we attempted to fit synthetic data computed by forward propagation through toy models ($\boldsymbol{d}_{test} = \boldsymbol{G}\boldsymbol{m}_{test}$) to which we added Gaussian noise using standard deviations estimated from MCCC measurements.

We used "squeezing tests" to probe the depth range of mantle anomalies required by the data. For these, we conducted a suite of inversions for which the model nodes below a deeper squeezing depth z_d or above shallower squeezing depth z_s, were very heavily damped. This yields a reformatted inversion matrix for the squeezed inverse problem $\boldsymbol{G}'\boldsymbol{m}_1 = \boldsymbol{d}$. $\boldsymbol{G}'$ includes damping that 'squeezes' any structure out of these volumes of the model, and thus forces the inversion to attempt to fit observed data with only a subset of the model nodes. For squeezed inversions, we fixed the event and station static times to values derived from the non-squeezed inversion. To quantitatively compare the squeezed models, we computed two metrics: the weighted variance reduction (wvr) and the L1 norm of $\boldsymbol{m}_2$. $\boldsymbol{m}_2$ is the model obtained through an un-squeezed inversion of $\boldsymbol{G}\boldsymbol{m}_2 = \boldsymbol{r}$ where squeezing is relaxed such that the entire model space is available to fit the data residual $\boldsymbol{r} = \boldsymbol{d} - \boldsymbol{G}'\boldsymbol{m}_1$ from the 'squeezed' inversion. This yields structure that is unable to be captured by the squeezed model but is nevertheless important for fitting the data. Higher values of wvr and lower values of $||\boldsymbol{m}_2||_1$ for a given squeezing test indicate the data more strongly require structure in the un-squeezed layers of that model.

Velocity-temperature calculation

In order to understand the implications of our tomography models for state variables, we forward model mantle velocities. We contrast the predicted seismic velocity for a parcel of

"upwelling" mantle at 1350˚C (T_a) with that for a colder parcel of "downwelling" mantle at $T = T_a - \delta T$, where we seek a δT to fit our observations. We use the database of Abers and Hacker (2016) for anharmonic velocities, assuming a pressure of 5 GPa, and a lherzolitic composition. A more depleted (harzburgitic) downwelling mantle would be ~0.02 km/s slower. We account for the effects of anelasticity using the model of Jackson and Faul (2010), assuming a 1 mm equilibrium grainsize (Behn et al., 2009), and an average seismic frequency appropriate to our travel time measurements of 0.5 Hz. With these parameters, a δT = 500˚C yields a V_P contrast of 4.8% between the downwelling (8.29 km/s) and upwelling (7.90 km/s) cells, matching tomographically observed peak to peak variations. Note that using the pre-melting anelasticity model of Yamauchi and Takei (2020) predicts a diminished contrast of 3.6%, due to less strong anelastic velocity reduction at high temperature. In the absence of experiments demonstrating a strong effect on the bulk modulus we assume that anelasticity, and the consequent physical dispersion, affects only the shear modulus. Non-negligible bulk attenuation would serve to exaggerate the velocity contrast. To explore the effect of putative melt, we adjust the anharmonic moduli according to Hammond and Humphreys (2000), and modify the anelasticity calculating by reducing the pre-factor of the diffusion creep timescale (Eilon & Abers, 2017; Holtzman, 2016) to account for the chemical and geometrical (predicted by contiguity theory (Takei & Holtzman, 2009)) effects of melt. As an example, this approach predicts that 0.2% *in situ* melt will reduce shear viscosity by a factor of 6. With these parameters, a δT = 300˚C and 0.5% *in* situ melt in the hotter (upwelling) mantle also yields the observed V_P contrast of 4.8% between the downwelling (8.13 km/s) and upwelling (7.75 km/s) cells.

Admittance, coherence, and gravity correction

We computed the free air admittance and coherence between differential gravity and bathymetry in the 2-D wavenumber domain, averaging over wavenumber annuli of width 0.017km^{-1}. We removed a planar trend from both fields before calculating the spectra. Admittance (Q) in each wavenumber band $(k)$was calculated as the weighted average spectral ratio between the Fourier transformed gravity, $G(k)$, and bathymetry $T(k)$, weighting by the bathymetry:

$$\mathrm{Q} = \frac{\sum_{\mathrm{i}}^{N} \frac{\mathrm{G(k)_i}}{\mathrm{T(k)_i}} |\mathrm{T(k)_i}|}{\sum_{\mathrm{i}}^{\mathrm{N}} |\mathrm{T(k)_i}|}$$

where i is the index among the N Fourier coefficients within the wavenumber band. Since the spectra are complex, we calculate separate admittance spectra for the sine and cosine terms, then average the two. Errors are determined as the standard errors of weighted means. We also calculate the theoretical admittance spectrum accounting for upward continuation and

assuming a 7 km-thick crust (z_c), densities of 1030 kg m^{-3}, 2750 kg m^{-3}, and 3300 kg m^{-3} for the water, crust and mantle layers, and an average water depth ($\bar{z}$) of 4600 m:

$$\mathrm{Q_T}(k) = 2\pi\mathrm{G}\{[\rho_c - \rho_w] + [\rho_m - \rho_c]\, exp(-z_c k)\}\, exp(-\bar{z}k)$$

where the non-italicized G represents the gravitational constant. The coherence is calculated as:

$$\gamma^2 = \frac{(\sum_i^N |G(k)_i\, T(k)_i^*|)^2}{(\sum_i^N T(k)_i\, T(k)_i^*)\,(\sum_i^N G(k)_i\, G(k)_i^*)}$$

With standard errors estimated from

$$\Delta\gamma^2 = (1-\gamma^2)(2\gamma^2/N)^{1/2}$$

Finally, we calculated the mantle Bouguer anomaly (MBA) by subtracting the effect of bathymetry in the spectral domain using the theoretical admittance (*i.e.*, agnostic of any true compensation; assuming constant thickness crust). Since our bathymetry coverage at long and short wavelengths is uneven, we also apply a 400-20 km cosine bandpass filter to the MBA.

Gravity modelling

We predicted 1-D gravity variations at the top of a 2.5-D dV_p tomography model as follows: We converted from velocity to temperature assuming $\frac{dT}{dV_p}$ = 54 K/%. This conservative value implicitly assumes that melt modifies velocities but has no substantial effect on density. We converted from temperature variations to density anomalies using a fixed (*i.e.*, not depth dependent) thermal expansion coefficient of 3.5e-5, and a reference density of 3250 kg/m^3. To avoid edge effects in the gravity modelling, at each depth we decomposed 1-D density variation into a series of sines and cosines with log-spaced wavelengths from 60-360 km. We used the simple relationship for upward continuation of sinusoidal vertical gravity perturbations: $\Delta g_Z(x) = 2\pi G\; dh\; \delta\rho \sin\left(\frac{2\pi x}{\lambda}\right) \exp\left(-\frac{2\pi z}{\lambda}\right)$, where G is the gravitational constant, z is layer depth and dh is layer thickness, $\delta\rho$ is density perturbation (the coefficient of the sinusoid), and λ is the wavelength of the anomaly. Cosine variations are handled analogously.

Supporting information figures

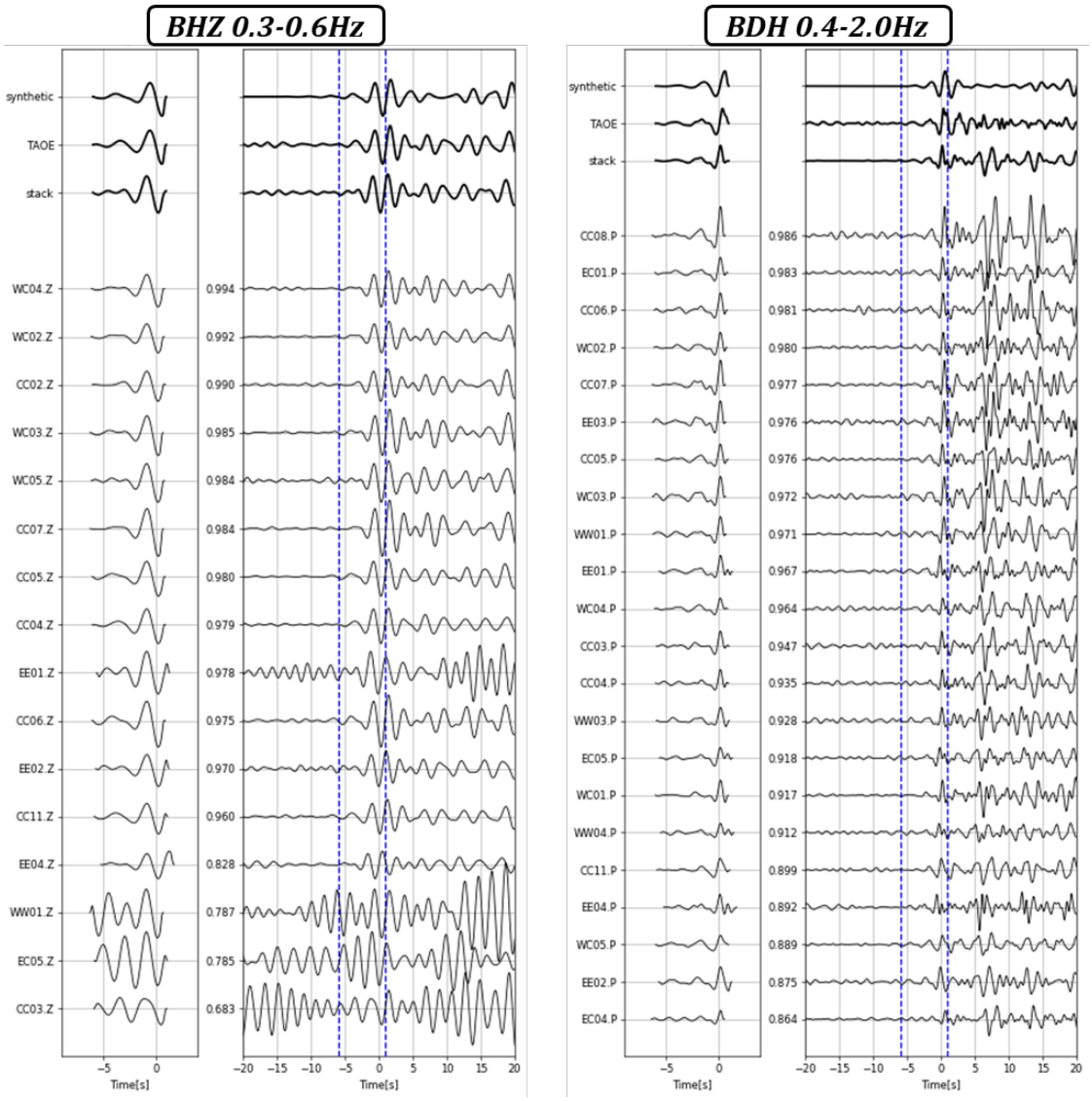

Figure S1. *MCCC differential travel time measurement example for event 2018-11-30 17:29 (Mw7.1) showing vertical channel (BHZ) and pressure channel (BDH). The distinct filter bands for the two components are given at top. For each component, the left column shows post-alignment trace segments for cross-correlation and the right column shows pre-alignment traces with the hand-selected time window indicated by blue lines. The cross-correlation coefficient between each individual trace and the stack (for that component) is given between the columns. At top, the stack of the traces (after alignment) is compared to the synthetic trace from Syngine* (Incorporated Research Institutions for Seismology, 2015)*, as well as a vertical seismic record from ~600km away land station TAOE. Note that the polarity of synthetic and TAOE waveform is flipped for comparison to the pressure channel waveform.*

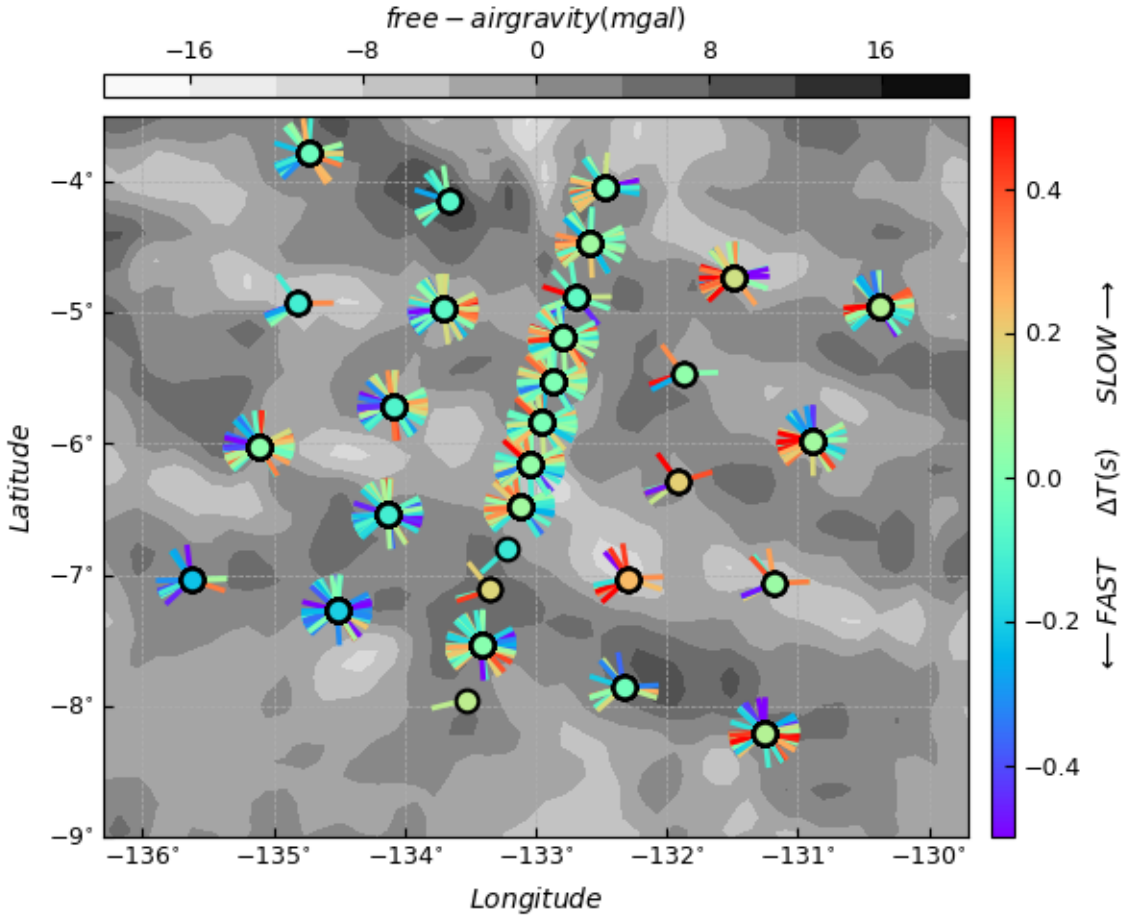

Figure S2. *P-wave differential travel time measurements. Compilation of measurements on both BHZ and BDH components, where each spoke is one measurement, coloured by relative arrival time and pointing in the station-event azimuth. Circles at the station locations indicate mean relative arrival times. The background greyscale map shows free air gravity anomalies.*

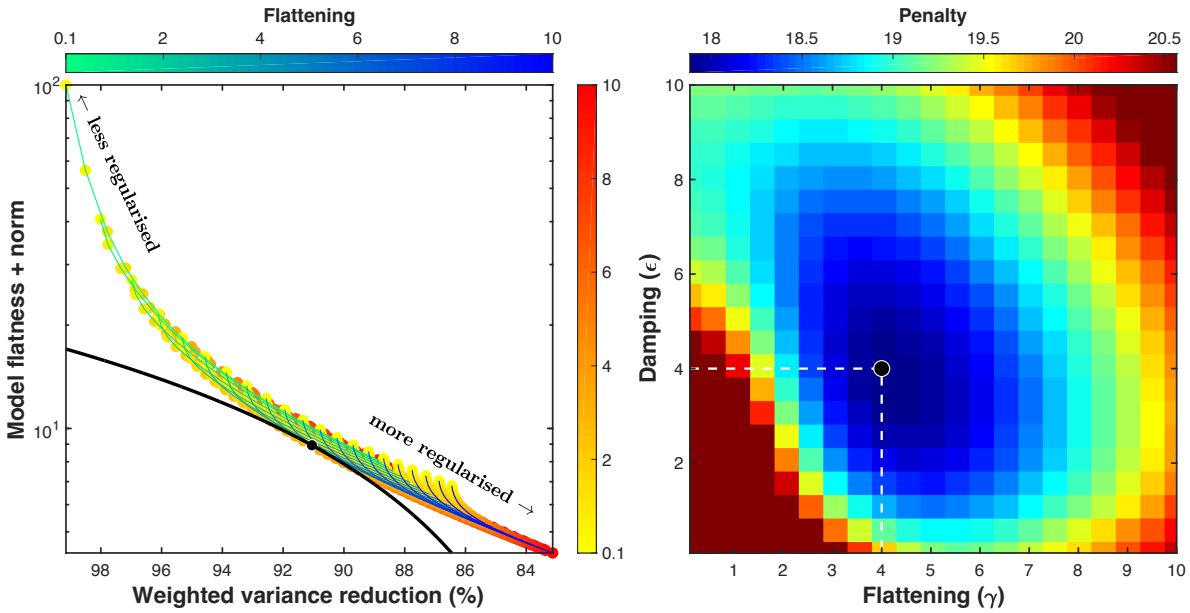

Figure S3. *Tests for optimal regularization parameters used for the inversion. Results of grid-searching for preferred weights of model norm damping (epsilon) and first-derivative damping ("flattening"; gamma) used in the weighted least squares objective function.* ***Left:*** *Trade-off between variance reduction (a measure of data fit) and model norm/roughness (computed as the sum of the norms of the model values and the first derivative values). Each dot represents a single inversion, where dot colour indicates damping weight, and line colour indicates flattening weight. Black dot indicates the preferred value from (b) and black line shows the contour of equal penalty value associated with this point.* ***Right:*** *Contour plot of penalty function computed from a weighted sum of misfit and model norm/roughness, with weighting chosen empirically. Minimum value and associated regularization weights (ε = 4, γ = 4) shown.*

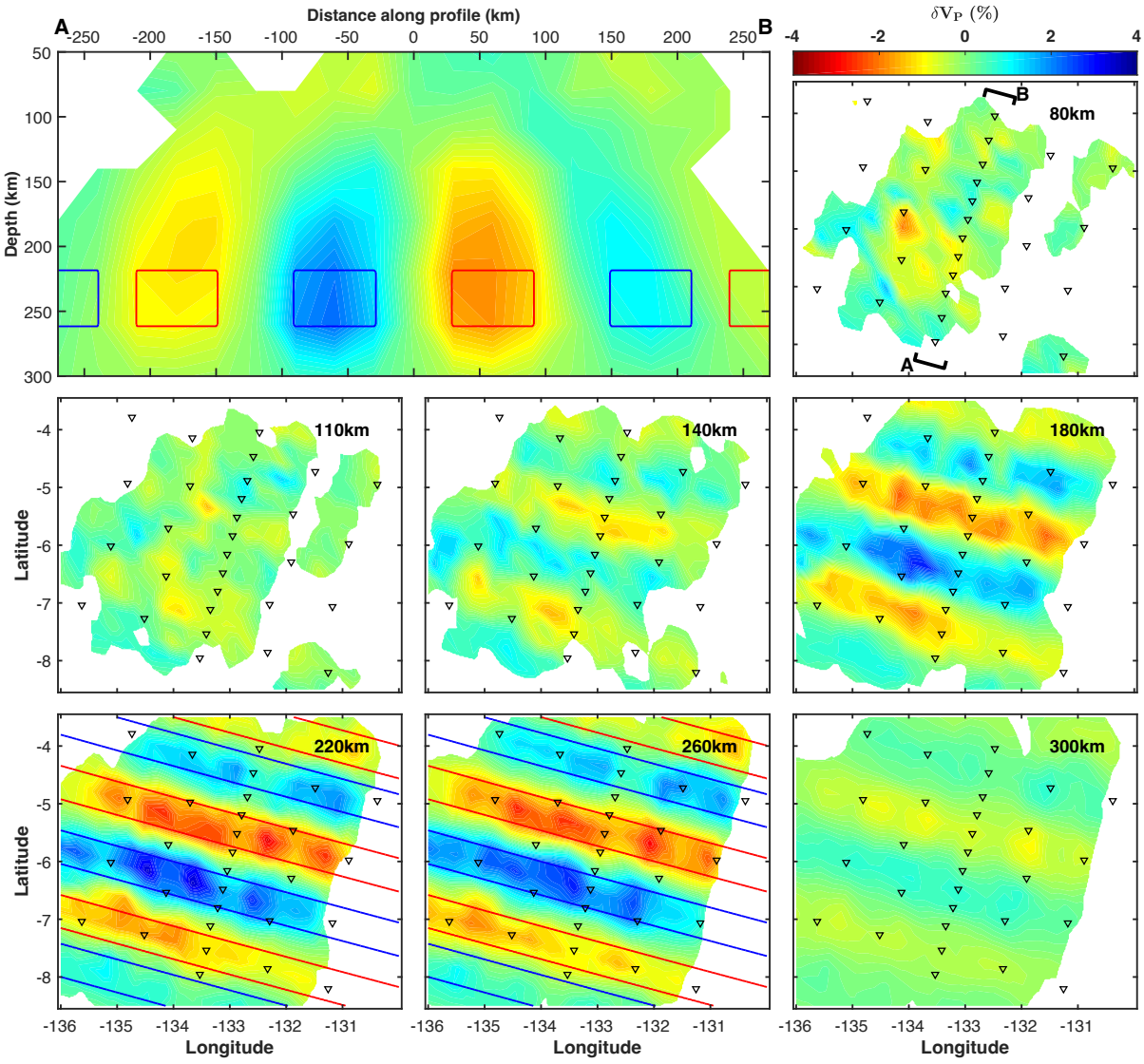

Figure S4. *Cylindrical rolls synthetic recovery test. Input-output test with buried 2.5-D bodies of alternating velocity. Outline of ±4% input structure shown by red (slow) and blue (fast) lines, superimposed on output tomographic model that uses the same regularization parameters as the true tomography. Regions of the output model with low hit quality are masked out.*

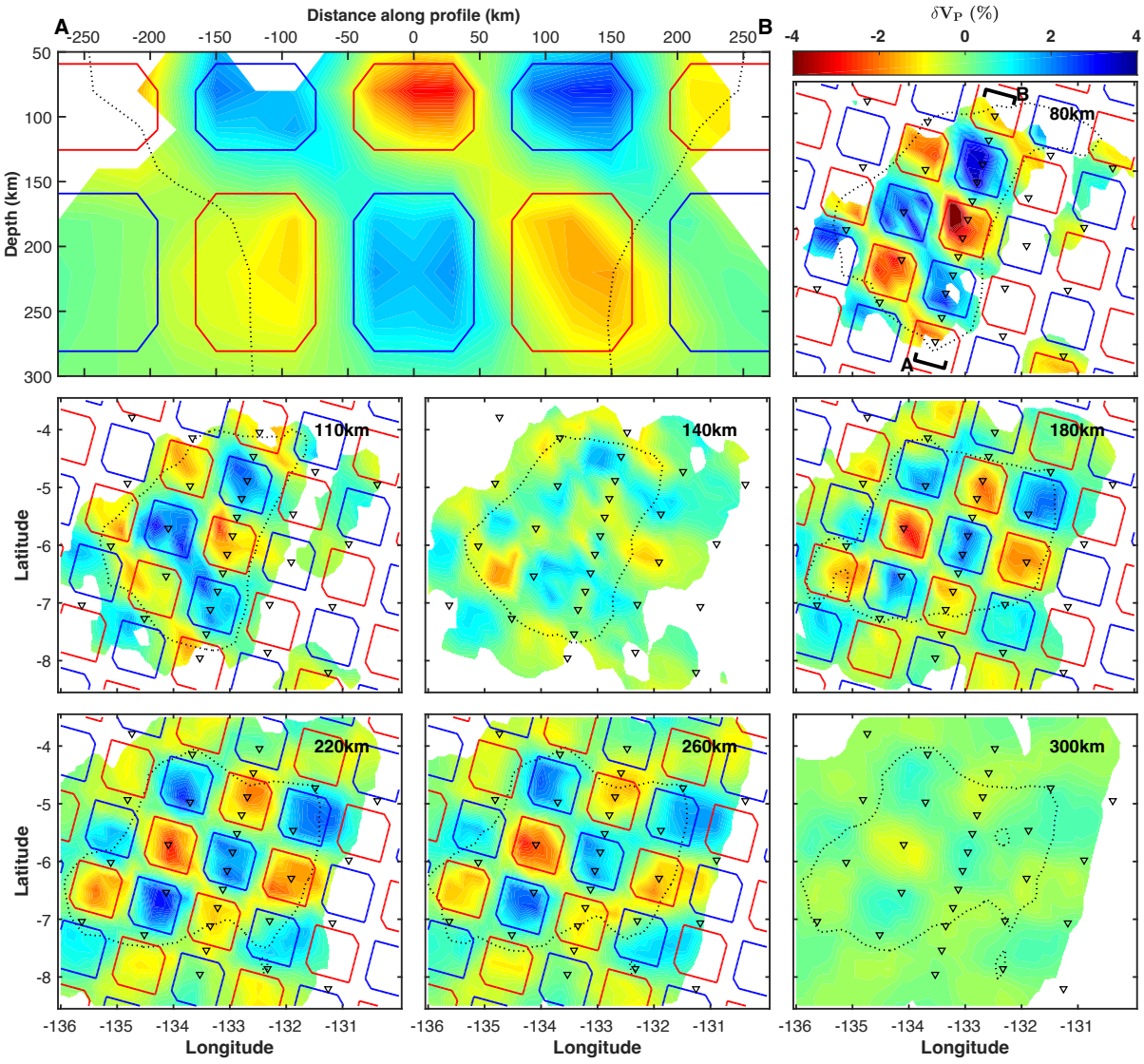

Figure S5. *Checkerboard synthetic recovery test. As for Fig. S4, but for ±4% checkerboard input structure. Semblance, a metric of spatially averaged recovery fidelity* (Zelt, 1998) *was calculated for this model using a spatial length scale equal to the checker size. The 70% semblance contour is shown by the dotted black line; this demarcates the region of very good synthetic model recovery.*

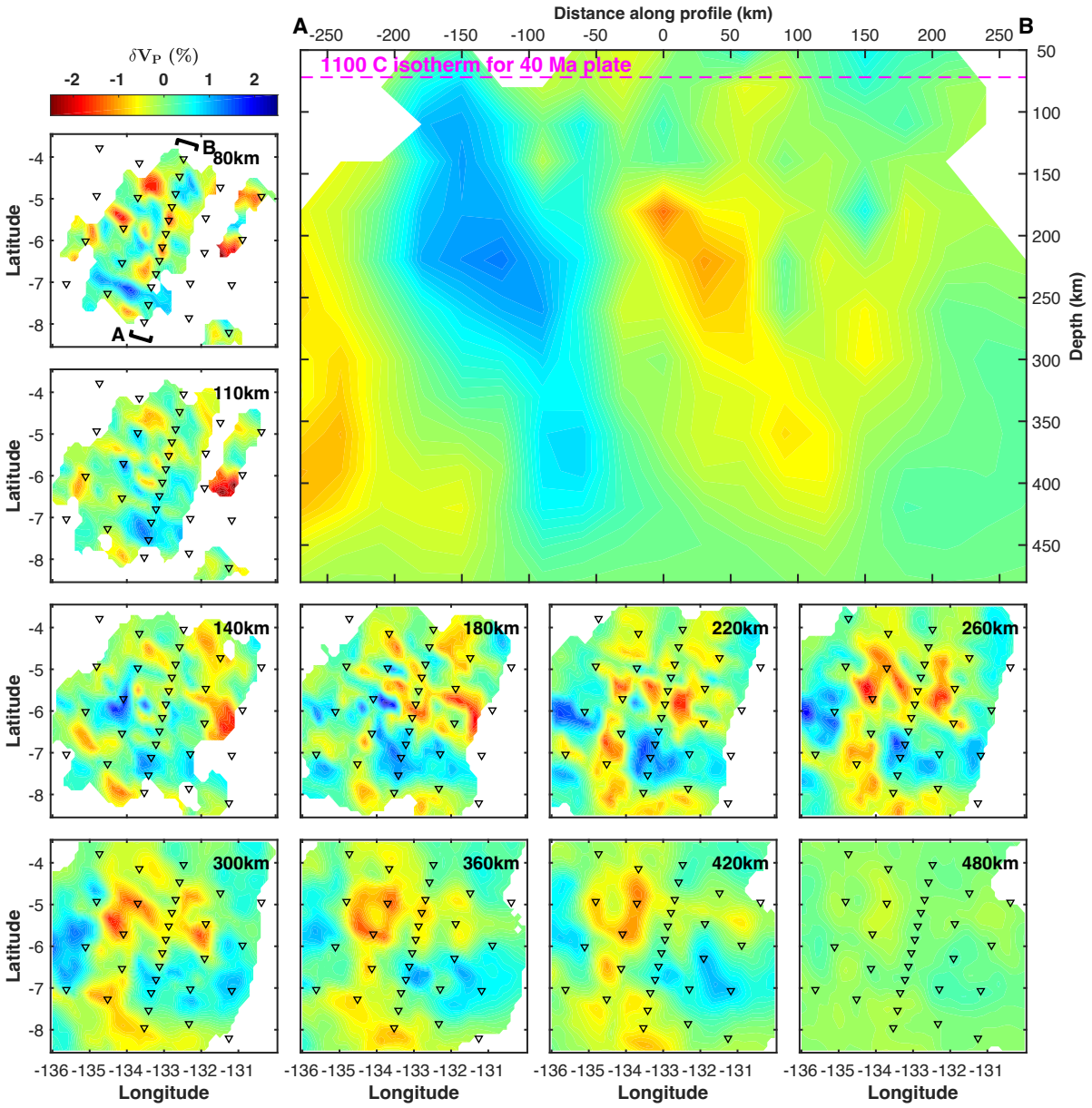

Figure S6. *Inversion with deeper base. This model was obtained by extending the bottom of the model domain to 480km depth. Although there is some smearing of features below the base of the preferred model (300 km), the majority of the structure remains in the 200-300 km depth range and the pattern of the anomalies is essentially unchanged. The weighted variance reduction for this inversion is 87%, compared to 85% for the preferred model.*

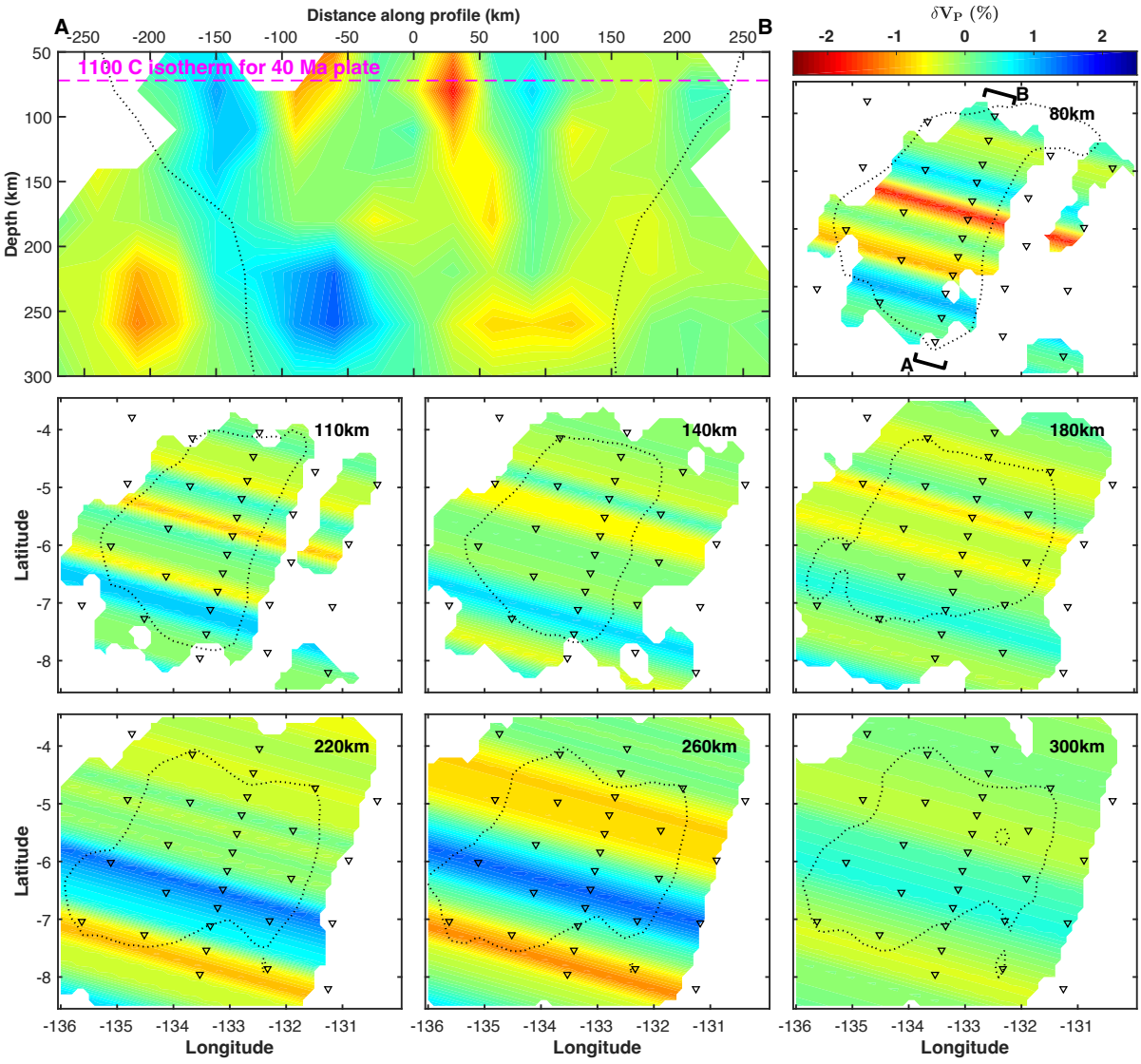

Figure S7. *2.5D data inversion. Similar to Fig. 2, but showing the model inverted with infinite smoothing in the 115˚ direction (i.e., restricting structure to vary in just two dimensions). The weighted variance reduction for this inversion is 63%, compared to 85% for the full 3D model.*

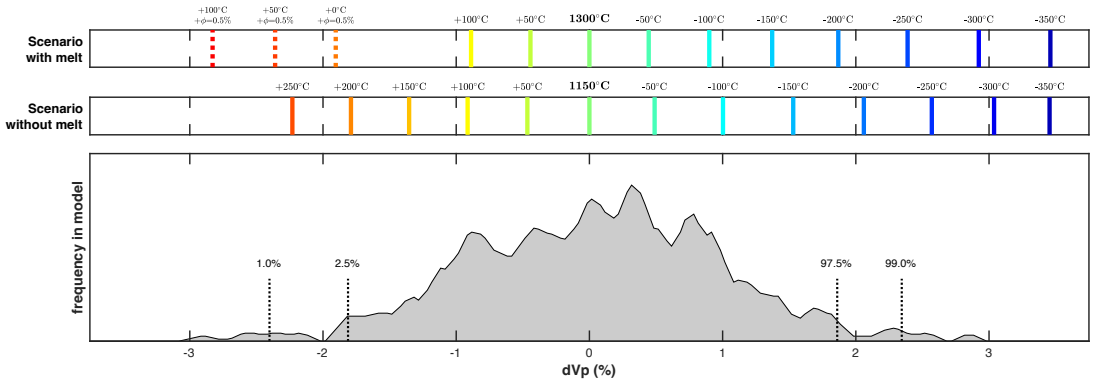

Figure S8. *Illustrative scenarios to explain observed velocity contrast. Bottom panel shows a histogram of node dVp values in the best resolved region of the model (nodes with hit-quality ≥0.3 between 120 and 280km depth, inclusive). 1%, 2.5%, 97.5% and 99% percentile dVp values are indicated. Upper panels show two scenarios for variations in dVp as a function of temperature and melt. Since the observed values provide no absolute velocity constraints, these scenarios are comparably consistent with the observations, despite having different "reference" temperatures (i.e., temperatures corresponding to the mean velocity observed). Velocities are calculated as described in the Text S1. All values shown include the effects of anelasticity, and only the dashed lines include the effects of melt (which modifies moduli both elastically and anelastically).*

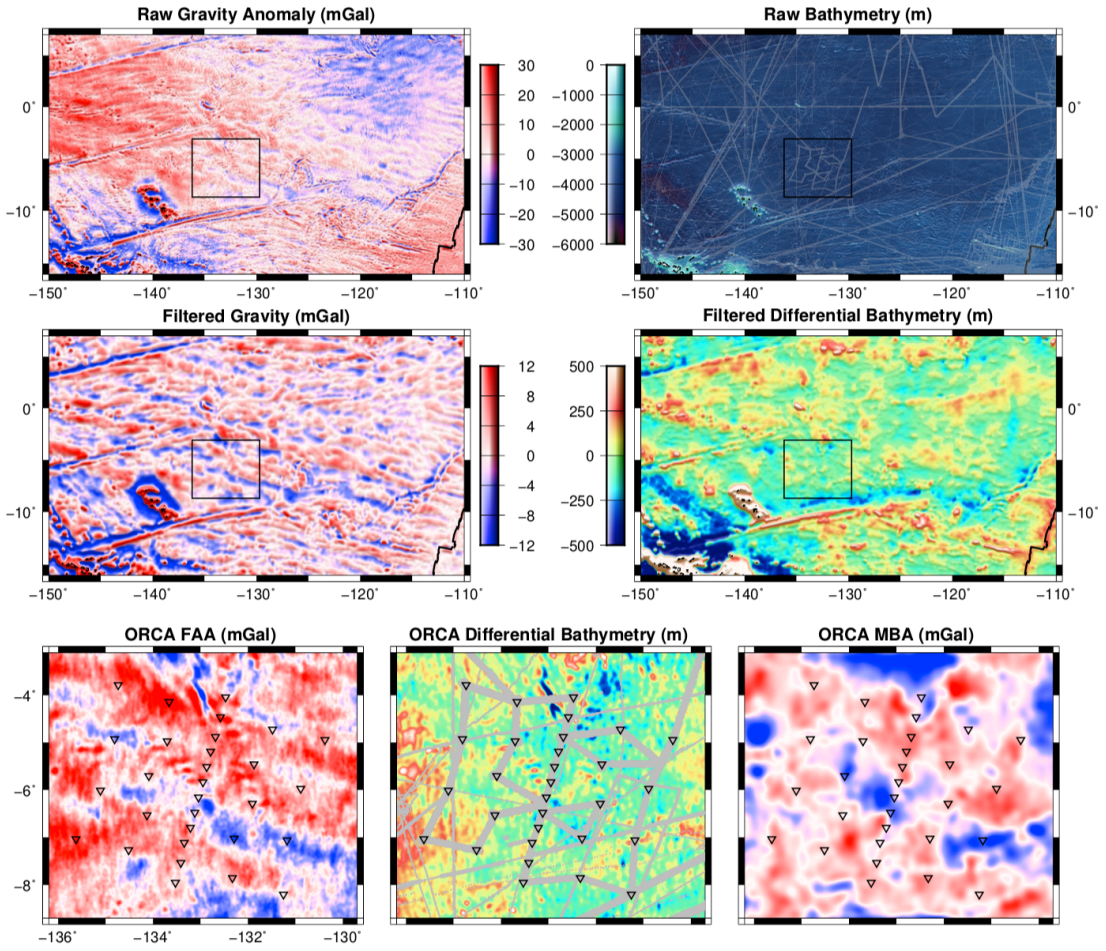

Figure S9. *Residual gravity and bathymetry. Top left: raw free air gravity anomaly [accessed at https://topex.ucsd.edu/pub/global_grav_1min/ on 11/5/21] from satellite altimetry* (Garcia et al., 2014). *Top right: bathymetry from satellite altimetry and ship soundings* (Smith & Sandwell, 1997) *[accessed at https://topex.ucsd.edu/pub/global_topo_1min on 10/14/2021]. Middle left: gravity anomaly as above but filtered in the spatial domain using a 80 km gaussian convolution filter and then a 10^6-10^1 m gaussian bandpass filter, to avoid spectral ringing. Middle right: bathymetry de-meaned and filtered identically to the gravity field. Black box in top two rows shows region in bottom row. Bottom left: Zoomed-in free air gravity anomaly (unfiltered) in the region of our OBS array (black triangles). Bottom center: Zoomed-in de-meaned (but unfiltered) bathymetry, with ship soundings shown. Bottom right: Mantle Bouguer anomaly (see Text S1) filtered from 600-10 km wavelength.*

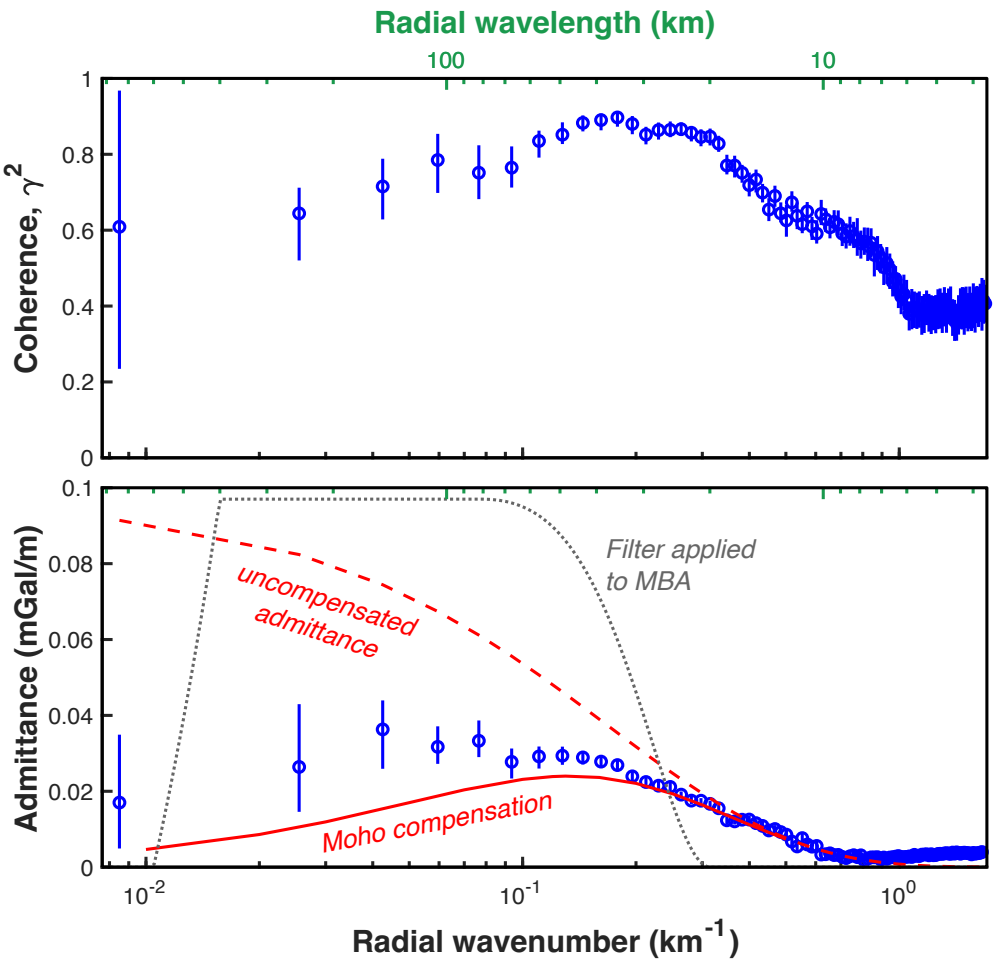

Figure S10. *Coherence and admittance in our study area. Top: Coherence between 2-D free air gravity and bathymetry, with uncertainty calculated using the approach of Bechtel et al.* (1987). *Spectra are averaged across wavenumber annuli with width 0.017 km^{-1}. Bottom: Observed (blue) and predicted (red dashed) un-compensated free air admittance values for uncompensated topography. The former are shown with 10-90th percentile ranges from bootstrap analysis (see Text S1), the latter include the effects of upward continuation and both water-crust and crust-mantle periodic density variations, assuming crustal density of 2750 kg m^{-3}, mantle density of 3300 kg m^{-3}, a constant 7 km thick crust, and average water depth of 4600 m. Solid red line is predicted admittance for topography compensated at the Moho.*