Review Details: Towards surface wave tomography with 3D resolution and uncertainty

Completed: 2024-11-11 03:55 AM
Recommendation: Revisions Required
Reviewer Comments

For author and editor

Dear editor,

Dear authors,

This manuscript describes a combination of the SOLA method (a variant of Backus-Gilbert inversion) with finite-frequency theory for surface waves. Following a detailed introduction into surface wave tomography and the need to quantify uncertainties, the authors provide a derivation of their method and a synthetic inversion at global scale.

The manuscript is well written and a pleasure to read. I have no doubt that the method, maybe after some improvements to increase computational efficiency, will be an attractive alternative to existing approaches, especially when uncertainty quantification is critical for the scientific question that one actually aims to answer. Also, the paper will surely become a valuable addition to the existing seismic tomography literature.

Most of my comments are minor and contained in the annotated manuscript. They primarily concern inaccuracies in some lines of arguments. Two more significant concerns, detailed below, are related to (1) sources of model uncertainty that may be more significant than the ones addressed by the authors, and (2) the concept and treatment of forward modelling (theory) errors.

In summary, all of my comments can be addressed with improvements of the text and maybe some additions to the discussion section. Since no additional simulations are needed, I would classify this as a minor revision.

Please find more details below, as well as in the annotated manuscript.

With kind regards

Andreas Fichtner

Major comments

Sources of model uncertainties

The authors mainly consider two sources of model uncertainties: errors in the data and forward modelling errors, i.e., inaccuracies of the underlying theory. While these two are certainly important, they are only part of the story.

Nonlinearity can have a major impact, and not taking it into account is a limitation of the method that deserves a more honest discussion. There are nonlinear versions of the Backus-Gilbert method, e.g., by Roel Snieder, that should probably be mentioned. Using a fixed, and unavoidably inaccurate, crustal model is related to nonlinearity, too. Iteratively improving tomographic models can have an enormous impact, as evidenced by recent models that use hundreds instead of just a few tens of iterations (e.g., REVEAL by Thrastarson et al., or WUS324 by Rodgers et al.).

Probably the most important source of model uncertainties are all the little choices that we make along the way. This includes the shape and size of the target resolution kernels, the code used to compute artificial data, approximations made to compute sensitivity kernels, and, again, the choice of the crustal model. Without being able to proof this, my gut feeling is that this is the zeroth-order contribution to model uncertainties. Of course, I would not request the authors to solve this issue, but it deserves to be mentioned in the discussion.

Theory errors

I can understand the authors' idea behind the estimation of forward modelling errors. Loosely speaking, everything above an rms misfit of 1 is interpreted as forward modelling errors, which are added to the observational errors. Apart from the fact that this line of arguments breaks down for non-Gaussian errors, there are a few other issues: (1) An rms significantly above 1 can also result from a parameterisation that is too coarse, or from a target resolution that is too low. This has nothing to do with forward modelling errors. Hence, by changing these subjective choices, the forward modelling error changes, thereby turning it into an arbitrary quantity. (2) Forward modelling errors are not random but deterministic. Hence, it does not make sense to add them to the data errors and to treat them as a random variable.

Going a bit further, I would even argue that there is no such thing as a forward modelling error. In fact, the notion of "error" is meaningful only in the context of a reference. Error with respect to what? For a forward model such a reference does not exist because all models are wrong from the outset, and there is no universal measure for the quality of a model. (They are mere abstractions of nature.) It follows that it does not actually make sense to introduce a forward modelling error, because it is an arbitrary quantity anyway.

Completed: 2024-12-05 01:26 PM
Recommendation: Revisions Required
Reviewer Comments

For author and editor

Dear authors and editor,

The authors have provided an interesting exploratory paper, trying to bring togheter several well-known problematic aspects in seismic imaging and possible directions to address them. I think this study could be important for people interested in mantle imaging with methods that deviate from the common assumptions adopted (damped least squares or probabilistic sampling). The manuscript is well written, even if I think some additional details could be beneficial to it (see comments below), the title and the abstract fits well the topics discussed. All the pictures are relevant in the context they are presented in.

I had to familiarize a little with the methodology used, reading the work of Latallerie et al. 2022 and the precedent works of Zaroli et al. 2016 and Zaroli et al. 2017. Considering this, my first comment is that this study works well following up the precedent papers, but it could be not enough self-explanatory to be fully understandable by itself. For this reason, I suggest expanding the methodology section to provide some additional information about the SOLA algorithm (eventually, this could be even presented in the Appendix). At the end, this study revolves mostly around methods, not results, so I suppose the methodology should be outlined with some additional details.

These are the aspects of the inversion scheme I would like to see investigated further by the authors (lines in the manuscript are referred to as Lx, where x is the number/range):

L275-279: I suppose that the idea of building the generalized inverse matrix assuming uncorrelated equal errors in all the observations (data covariance matrix = identity matrix) rises from the difficulty in assessing the true uncertainty (measurement + theorical) and the influence this can have in the inversion process. However, since you are working with a wide range of surface waves' periods, I would expect the noise in the measurements to manifest some kind of correlation with the frequency. So, either - by using the estimated uncertainties you consider more "robust" - you justify assuming the same error (i.e., weight) at all periods in the first step, or I would like to see discussed the impact different weights attributed to the data can have on the generalized inverse and on the final solution. I suppose fixing the data covariance matrix with reasonable uncertainty estimates and building multiple solutions can illuminate this. Another reason why I consider this relatively important is because of the strong – and unexplained (not undescribed) - deviations your solution model manifests from the input filtered model. I wonder if one possible cause could be in the way the generalized matrix is built. The deviations between the solution model and the filtered input model seems to be largely due to fine scale structures, I would not be surprised if the average deviation could be reduced by a more appropriate weighting scheme.

L303-304: the part about resolved deep kernels representing an average of the shallower portions of the model is particularly interesting. I suppose this is dependent on the choice of the trade-off parameter “eta” that prioritizes low covariances for the model increasing the deviation of the resolved kernels from the target ones. However, this makes me wonder if the trade-off parameter could be selected to partially prevent these cases (?). If I understood correctly, error propagation factor decreases around 200 km depth (fig. 5) because resolved kernels represent the average of large shallower regions of space (fig. 4), enclosing many finite frequency kernels, resulting in low resolution and low model’s uncertainty. Is this correct? I would like to see a comparison between target and resolved kernels toward 400 km depth, where the uncertainty propagation factor seems to drop to 0.
L331-341: since your method uses a subset of the dataset to reduce the storage required, it is my opinion that an approach more robust than randomly selecting rays excluding similar paths would be i) using a binning technique to merge similar-paths, extracting more robust estimates for the observables (and also uncertainties…), or ii) using a procedure similar to bootstrapping, where the dataset is randomly subsampled multiple times, and the solution models resulting from each subset are brought together in an ensemble of solutions. This allows, using metrics like average and standard deviation, to explore central features and poorly constrained anomalies highly dependent on the specific subset. Considering that many of your measurements have the uncertainties potentially underestimated, method ii) could be interesting to compare the uncertainties in your model coming from bootstrapping with the error propagation you use to compute model’s covariance matrix. I will not explicitly ask to add these to the study, but if you agree that these approaches could work, maybe this discussion should be added to the text, or considered for future experiments.
L359-366: I find a little concerning that your forward modeling strategy can describe only 30% of the selected dataset. Considering that the - relatively accurate - approximations adopted in this study should account for the primary elements controlling surface wave phase observables (especially if simulated), I cannot fully understand why almost 70% of the measurements deviate so much from the predictions of the true model. I suppose phenomena like cycle-skipping could be identified by comparing the same event-station phase measurement at different periods and checking if one breaks an expected trend (since it should be wavelength dependent…). If a wide range of frequencies gives rise to cycle-skips, like in fig. 6, I wonder if alternative measurement methods based on nearby stations cross-correlations (Jin & Gaherty, 2015), could provide more reliable observations, at least for some problematic cases. In the pictures you show in Fig. 6 (top), the case representing a class III measurement is shown as a surface wave traveling across North America, while in the other two – non problematic – cases the propagation is across the Pacific. Could you plot a map with the “rays” connecting events to stations (like Latallerie et al. 2022, fig.1a) colored by the deviation (measurement – input prediction)? Just to check if the deviations manifest some sort of correlation with the waves’ paths. Does using finite frequency sensitivity kernels improve data-misfit with respect to ray-theory predictions (considering the true model)? I would like to better understand if this is mostly a problem with the measurement algorithm, or with the forward modeling you adopt in the code. Either way, this could affect real data applications of the inference code, where no true model predictions exist to compare with the measurements.

L438: since many deviations from the input filtered model are small scale features, I wonder if varying the trade-off parameter “eta” could have any effect. In this study I don’t see any trade-off curve, or a similar criterion, for the selection of the trade-off value. How did you choose this parameter? Is there trade-off value that could decrease significantly chi^2? Would that be far from your choice (50)?

This is a general comment about the discussion section. I agree that SOLA is an interesting option to consider when compared with damped least squares methods, even if lack of smoothing constraints in SOLA seems to give rise to relatively coarse solution models. However, it remains non-trivial to interpret the robustness of anomalies bringing together resolution and uncertainty propagation (especially in an unambiguous quantitative way). Moreover, the method remains limited to linear(ised) problems, because any sort of iterative linearization would make error propagation non-trivial (due to the iterative perturbative approach). In Bayesian Monte Carlo inference, I agree that prior probability functions have their influence on the posterior density, but i) non-informative priors (like uniforms) can be chosen for the fields directly investigated in the inversion, ii) even if informative priors were chosen, Bayesian inference at the end is all about comparing priors with posteriors, to see where information deviates from prior knowledge. These solvers provide uncertainty estimation directly sampling from the posterior models that fit the data within the uncertainty. Moreover, in the case of the trans-dimensional (t-dim) methods you mention (Bodin and Sambridge, 2009), many parameters that typically need to be selected a priori for the model’s parametrization are removed, resulting in a self-parametrized multi-resolution approach. I have not fully understood your claim that t-dim solvers are approximate (??) and provide partial information about the resolution (L475-477), because this is the entire point behind trans-dimensional imaging. If data strongly requires a feature, small or large scale, that will be manifested in many models in the ensemble, resulting in the average and uncertainty maps. I have also some thoughts about your claim that hierarchical methods, where noise is a random variable in the posterior, use a single noise parameter for all data (L568-572). This is a partial consideration because of the following reasons: i) data can be divided in “classes”, or different observables, and independent noise parameters can be introduced for each one (Zhang et al., 2018; surface-waves phase and group travel times; Del Piccolo et al. 2024; body-waves P&S delays and splitting intensity). The “classes” of observables can be chosen arbitrarily; hypothetically, I could also separate different ranges of periods for surface waves and invert for independent noise parameters, or different classes based on a “measurement quality metric”… Another option is ii) to use more complicated parametrizations to represent noise depending on physical quantities, for example introducing polynomial functions (or other basis functions) depending on the wave-paths’ ranges (Bodin et al., 2012; Zhang et al., 2020), on the periods, etc… defining the coefficients of the polynomials as posterior’s random variables to be inferred during the inversion. In this process, noise enters directly the posterior definition with the possibility to be inferred, but also directly controls model’s total uncertainty acting on the width of the likelihood functions.

These were my main comments about the manuscript submitted. Again, since this study in mostly about methods and uncertainty assessment, I think that the additional discussion and general aspects mentioned above could really help to better frame this work and its relationship with already existing methodologies to address the problem of uncertainty estimation in seismic imaging.

Here, some additional minor points follow.

L56-59: I would say that the reason why data-misfit based methods do not directly account for resolution is because this metric is influenced by the grid choice (model's parametrization) and smoothing constraints in a non-trivial combination. It is not clear what you mean by "computational reasons". Are you referring to full-waveform inversions? Because otherwise these solvers are typically extremely efficient when working with secondary observables (like travel-times).
L60: what do you mean exactly by "crude approximation" of the resolution?
L161-163: does the model's covariance matrix also show non-zero elements outside the diagonal representing covariances between parameters? Would accounting for covariances change significantly the estimated model's uncertainty?
L185-187: it is my understanding that, if we refer to linear problems, matrix G (and the generalized inverse) does not depend on the values of data d (also for "data-fitting" methods). A typical example is a body wave imaging experiment with fixed rays and a grid parametrization for the velocity model; the element Gij would represent the length of the segment of the i-th ray in the j-th voxel, so also the generalized inverse – as a manipulation of G, would not depend on data d (travel times). Probably you were referring to iterative linearized solvers, where G represents the Fréchet derivatives, and it is in general a function on the data d, but this would make the comparison with SOLA (only linear) unfair I suppose. Could you elaborate this point further?
L193: please, specify what kind of data delta_phi_l in eq. (6) represents.
L230-231: do you expect that an alternative 3D parametrization in voxels different from 2°x2°x25 km would significantly impact your conclusions about resolution? Or is this strategy robust for even smaller voxels?
2(b): correct title "data sensitivity".
L267-270: do you think that a different choice for the target resolution, for example depending on the coverage, could lead to an improved data-misfit with respect to the chi_2 = 33 you achieved in this study?
L275: I would think that data uncertainty always (not potentially) influences solution models in multi-observables (like multi-period) inversions.
L285-287: was the L-curve used to select a trade-off parameter value equal to 50 in this study?
L336: this is where I suppose something like bootstrapping could produce more informative models in terms of uncertainty.
L361-362: is the undetected cycle-skip at 8mHz responsible for all the discrepancies between analytical predictions and measurements in that 70% fraction of the dataset? It is not completely clear from the manuscript how you justify the deviations of such a significant fraction of the dataset.
Looking at figure 7(a), it looks like there is weak tilt in the relationship between the measured and the analytical phase delays. Would applying a rotation (meaning that either measured delays are systematically overestimated or predictions are systematically underestimated) lead to the analytical delays describing a wider range of observations (more than the current 30%)?

L397-400: I wonder if there could be a correlation between the rays corresponding to class III measurements (poor measurement unrealistically low uncertainty) and regions where you recover stronger anomalies with respect to the filtered input. This would also follow up the comment above about the possibility of predictions systematically underestimated due to the local average of resolved kernels. Would this mean that resolved kernels are too wide? Or the trade-off parameter not appropriate?.
Fig. 8: I would add a map showing the deviation (filtered input – model solution). It can be difficult to directly compare the two solution maps. Less important and not necessary to add, but I would really like to see what the input model looks like applying a filter with the target kernels.
L475-479: as mentioned in the main comments, some words here like “approximate” and “partial” probably needs some additional context. Moreover, some recent studies using Monte Carlo inference adopt a single-step scheme for the inversion of surface waves (Zhang et al., 2018; Zhang et al., 2020), providing more reasonable estimates for uncertainty compared to the 2-steps inversions.
L487-488: I agree that no explicit global constraints are applied to the model’s parameters. However, it is my opinion that limiting the model’s covariance acts similarly to an implicit “smoothing” constraint on the parameters, since a smooth model is typically expected to be more “stable” than a model with very rough (i.e., small-scale) features. I would expect that by changing the trade-off parameter, and promoting lower covariance models, they would also look smoother than the one in this study (because of the increasing size of the resolving kernels).
L479-480: I suppose the large-scale features are also the ones on which almost all the surface waves tomography models agree on…(?)
L504-510: this point is extremely interesting. I suppose that this would manifest as well in teleseismic imaging with body waves using SOLA algorithm. Anomalies tend to be smeared up to the surface, meaning the shallow parameters’ kernels are likely to represent the average of deeper sections. Again, interesting how the trade-off parameter could control these biases, but probably there is not an intuitive way to tune it for this purpose.
Thank you for submitting such an interesting study about uncertainty and resolution in mantle imaging, hopefully my comments can contribute to this manuscript. It was really interesting to go over the methodology you have adopted as alternative to dampled least squares.

Regards

References:

Franck Latallerie, Christophe Zaroli, Sophie Lambotte, Alessia Maggi, Analysis of tomographic models using resolution and uncertainties: a surface wave example from the Pacific, Geophysical Journal International, Volume 230, Issue 2, August 2022, Pages 893–907, https://doi.org/10.1093/gji/ggac095
Christophe Zaroli, Global seismic tomography using Backus–Gilbert inversion, Geophysical Journal International, Volume 207, Issue 2, November 2016, Pages 876–888, https://doi.org/10.1093/gji/ggw315

Zaroli, C., Koelemeijer, P., & Lambotte, S. (2017). Toward seeing the Earth's interior through unbiased tomographic lenses. Geophysical Research Letters, 44, 11,399–11,408. https://doi.org/10.1002/2017GL074996

Ge Jin, James B. Gaherty, Surface wave phase-velocity tomography based on multichannel cross-correlation, Geophysical Journal International, Volume 201, Issue 3, June 2015, Pages 1383–1398, https://doi.org/10.1093/gji/ggv079

Thomas Bodin, Malcolm Sambridge, Seismic tomography with the reversible jump algorithm, Geophysical Journal International, Volume 178, Issue 3, September 2009, Pages 1411–1436, https://doi.org/10.1111/j.1365-246X.2009.04226.x

Xin Zhang, Andrew Curtis, Erica Galetti, Sjoerd de Ridder, 3-D Monte Carlo surface wave tomography, Geophysical Journal International, Volume 215, Issue 3, December 2018, Pages 1644–1658, https://doi.org/10.1093/gji/ggy362

Gianmarco Del Piccolo, Brandon P. VanderBeek, Manuele Faccenda, Andrea Morelli, Joseph S. Byrnes; Imaging Upper-Mantle Anisotropy with Transdimensional Bayesian Monte Carlo Sampling, Bulletin of the Seismological Society of America, 2024; 114(3): Pages 1214–1226, https://doi.org/10.1785/0120230233

Thomas Bodin, Malcolm Sambridge, Nick Rawlinson, Pierre Arroucau, Transdimensional tomography with unknown data noise, Geophysical Journal International, Volume 189, Issue 3, June 2012, Pages 1536–1556, https://doi.org/10.1111/j.1365-246X.2012.05414.x

Xin Zhang, Corinna Roy, Andrew Curtis, Andy Nowacki, Brian Baptie, Imaging the subsurface using induced seismicity and ambient noise: 3-D tomographic Monte Carlo joint inversion of earthquake body wave traveltimes and surface wave dispersion, Geophysical Journal International, Volume 222, Issue 3, September 2020, Pages 1639–1655, https://doi.org/10.1093/gji/ggaa230

Towards surface-wave tomography with 3D resolution and uncertainty

Latallerie Franck * [1], **Christophe Zaroli** [2], **Sophie Lambotte** [2], **Alessia Maggi** [2], **Andrew Walker** [1], **Paula Koelemeijer** [1]

[1]Department of Earth Sciences, University of Oxford, Oxford, United Kingdom, [2]Institut Terre et Environement de Strasbourg, UMR7063, Université de Strasbourg, EOST/CNRS, 67084, Strasbourg CEDEX, France

Author contributions: *Conceptualization*: FL, CZ, SL, AM, AW, PK. *Methodology*: FL, CZ, SL. *Software*: FL, CZ, SL. *Data curation*: FL. *Formal Analysis*: FL. *Investigation*: FL, CZ, PK. *Writing - Original draft*: FL. *Writing - Review & Editing*: FL, CZ, SL, AM, AW, PK. *Visualization*: FL. *Supervision*: CZ, SL, AM, AW, PK. *Project administration*: AW, PK. *Funding acquisition*: AM, AW, PK.

Abstract Surface-wave tomography is crucial for mapping upper-mantle structure in poorly instrumented regions such as the oceans. However, data sparsity and errors lead to tomographic models with complex resolution and uncertainty, which can impede meaningful physical interpretations. Accounting for the full 3D resolution and robustly estimating model uncertainty remains challenging in surface-wave tomography. Here, we propose an approach to control and produce resolution and uncertainty in a fully three-dimensional framework by combining the Backus-Gilbert-based SOLA method with finite-frequency theory. Using a synthetic setup, we demonstrate the reliability of our approach and illustrate the artefacts arising in surface-wave tomography due to limited resolution. We also indicate how our synthetic setup enables us to assess the theoretical model uncertainty (arising due to assumptions in the forward theory), which is often overlooked due to the difficulty in assessing it. We show that in the current setup the theoretical uncertainty components may be much larger than the measurement uncertainty, thus dominating the overall uncertainty. Our study paves the way for more robust and quantitative interpretations in surface-wave tomography.

Non-technical summary In the oceans, several surface features such as isolated volcanic islands or variations in the depth of the seafloor, result from dynamic processes in the underlying mantle. To understand these processes, we need to image the three-dimensional structures present in the subsurface. While long-period surface waves can be utilised, the data are typically noisy and provide poor data coverage of the oceans. This limits the quality of our images and therefore the interpretations that can be drawn from them. In addition, limitations of our images are difficult to quantify with current methods, which makes interpretations even more difficult. In

---

*Corresponding author: franck.latallerie@earth.ox.ac.uk

this study, we propose an approach that uses elaborate computational methods to produce high-quality maps of 3D structures in the upper mantle, at the same time informing on the quality of our images. As a proof of concept, we present the method in a synthetic framework, which serves to demonstrate our ability to retrieve an input Earth model and enables us to estimate theoretical model uncertainties. Our approach will enable more robust interpretations of surface-wave tomography models in future.

1 Introduction

Many important geological processes (e.g. melting at mid-ocean ridges, spreading, subduction and hotspot volcanism) occur in oceanic regions. To improve our understanding of these processes, we need to robustly image the structure of the upper mantle. In poorly instrumented oceanic regions, this imaging relies heavily on surface-wave tomography. However, surface-wave data have poor spatial coverage, both laterally due to the uneven distribution of earthquakes (sources) and seismic stations (receivers), and vertically due to how their sensitivity varies with depth. Surface-wave data also contain errors due to imperfect measurement and physical theory. Poor data coverage renders the inverse problem ill-posed and together with data errors leads to complex model resolution and model uncertainty (e.g. Parker, 1977; Menke, 1989; Tarantola, 2005). This complex model resolution and uncertainty explain the strong discrepancies between published tomography models (e.g. Hosseini et al., 2018; Marignier et al., 2020; De Viron et al., 2021). With time, seismic tomography is moving towards more detailed imaging, while it is also increasingly utilised in other fields. However, to guarantee the usefulness of surface-wave tomographic images, we need to account for their full 3D resolution and uncertainty (e.g. Ritsema et al., 2004; Foulger et al., 2013; Rawlinson et al., 2014). Equipped with these, we will be able to avoid interpreting ~~non significant~~ anomalies (e.g. Latallerie et al., 2022), set up meaningful comparisons with theoretical predictions (e.g. Freissler et al., 2020), or include tomography models in further studies such as earthquake hazard assessments (e.g. Boaga et al., 2011; Socco et al., 2012; Boaga et al., 2012).

Many approaches have been proposed to solve ill-posed inverse problems in seismology (e.g. Wiggins, 1972; Parker, 1977; Tarantola and Valette, 1982; Nolet, 1985; Scales and Snieder, 1997; Trampert, 1998; Nolet, 2008). Most take a data-misfit point of view and search for a model solution whose predictions are 'close enough' to observations. However, such approaches usually do not account directly for model resolution and uncertainty, mainly for computational reasons. Several methods have been proposed to estimate resolution once a model solution is obtained, but they are usually computationally expensive or provide only crude approximations to the resolution (Nolet et al., 1999; Barmin et al., 2001; Ritsema et al., 2004; Shapiro et al., 2005; Ritsema et al., 2007; Fichtner and Trampert, 2011; An, 2012; Fichtner and Zunino, 2019; Simmons et al., 2019; Bonadio et al., 2021). Synthetic tests, sometimes in the form of checkerboard tests, can be useful to assess resolution, but these have been shown to be potentially misleading (e.g. Lévêque et al., 1993; Rawlinson and Spakman, 2016).

Other approaches for solving ill-posed inverse problems move away from the data-misfit point of view and instead concentrate on directly optimising model resolution and uncertainty. These approaches are typically based on Backus–Gilbert theory (Backus and Gilbert, 1967, 1968, 1970). One such approach, the SOLA (Subtractive Optimally

Localized Averages) formulation, was derived for helioseismology by Pijpers and Thompson (1992, 1994) before being introduced and adapted to linear body-wave tomographic inversions by Zaroli (2016) and Zaroli (2019). Besides body waves, the method has been successfully applied to normal-mode splitting data to constrain ratios between seismic velocities (Restelli et al., 2024) and to surface-~~s~~waves dispersion data to build group-velocity maps (Ouattara et al., 2019; Amiri et al., 2023) or 2D maps of the vertically polarised shear-wave velocity V_{SV} (Latallerie et al., 2022). Although SOLA can be applied only to linear problems, it requires no prior on the model solution, provides direct control on model resolution and uncertainty, and produces solutions free of averaging bias as shown by Zaroli et al. (2017).

Traditionally, surface-wave tomography studies are based on ray-theory. This infinite-frequency approximation requires a two-step procedure that can be done in either order. One way is to first solve the inverse problem laterally (to produce 2D phase or group-velocity maps) and to subsequently solve for velocity structure with depth (to produce 1D velocity profiles) (e.g. Ekström et al., 1997; Montagner, 2002; Yoshizawa and Kennett, 2004; Ekström, 2011; Ouattara et al., 2019; Seredkina, 2019; Isse et al., 2019; Magrini et al., 2022; Greenfield et al., 2022). The other approach is to first solve for velocity structure with depth for independent source-receiver pairs (to produce 1D path-averaged velocity profiles) and to subsequently solve for lateral variations (to produce 2D velocity maps) (e.g. Debayle and Lévêque, 1997; Lévêque et al., 1998; Debayle, 1999; Debayle and Kennett, 2000; Simons et al., 2002; Lebedev and Nolet, 2003; Priestley, 2003; Debayle and Sambridge, 2004; Maggi et al., 2006b,a; Priestley and Mckenzie, 2006). This second approach was adopted by Latallerie et al. (2022) who applied the SOLA method to the second step (lateral inversion) to produce 2D lateral resolution and uncertainty information, together with their tomography model. Because the first step is a non-linear depth inversion, it could not be performed using SOLA – a purely linear method. Therefore, this study was not able to provide high-quality information about vertical resolution, a significant drawback given the complex depth sensitivity of surface-waves.

In this study, we extend the approach of Latallerie et al. (2022) to 3D using the framework of finite-frequency theory (e.g. Snieder, 1986; Snieder and Nolet, 1987; Marquering et al., 1998; Dahlen and Tromp, 1999; Yoshizawa and Kennett, 2004; Zhou et al., 2004, 2005; Yoshizawa and Kennett, 2005; Zhou, 2009a,b; Ruan and Zhou, 2010; Tian et al., 2011; Zhou et al., 2006; Liu and Zhou, 2016b,a). In this framework, surface-wave dispersion data are linearly related to perturbations in the 3D upper-mantle velocity structure. This makes it possible to perform a one-step inversion and thus to obtain 3D resolution information using SOLA. Finite-frequency inversions come with higher memory costs because the sensitivity kernels are volumetric (with both a lateral and depth extent) and the whole 3D model must be stored all at once (large number of model parameters). However, with smart data selection and ever increasing computational power, this memory cost is becoming less of an issue.

SOLA offers a way to propagate data uncertainty into model uncertainty. However, the robustness of model uncertainty in turn relies on the quality of data uncertainty, which is challenging to estimate. It is often estimated by comparing the dispersion of measurements for nearby rays (e.g. Maggi et al., 2006b). However, this approach dramatically underestimates the data uncertainty and poorly accounts for systematic biases (e.g. Latallerie et al., 2022). This is less of an issue if we are only interested in the relative uncertainty between individual data (e.g. when we weigh data contributions in a data-driven inversion). However, it is not sufficient if we want to interpret the true magnitude of the model uncertainty. It therefore becomes important to estimate data uncertainties carefully. Since data errors stem from imperfect measurements and inaccurate forward theory, it is natural to split them into two components: measurement and theoretical. Measurement uncertainty is estimated during the dispersion measurement and accounts for imperfections in the measurement algorithm (including cycle-skipping and mode contamination). Theoretical uncertainty is defined in a broad sense and accounts for errors not captured by the measurement algorithm. In particular, it includes assumptions in the forward problem, where we identify several main contributions, such as: single-scattering, discretisation and the sensitivity of the data to multiple physical parameters. The theoretical component is often missing in uncertainty estimates based on measurement uncertainty only, which partly explains why model uncertainty appears to be dramatically underestimated.

In this study, we show that it is possible to obtain detailed 3D resolution and robust uncertainty information using surface waves with SOLA within a finite-frequency framework, thus extending the approach of Latallerie et al. (2022) to 3D. By working in a synthetic setup, we demonstrate the feasibility of our approach, and quantitatively discuss statistical estimates of theoretical uncertainty. To achieve these aims, we develop a complete workflow from dispersion measurements on the waveforms to analyses of the resulting 3D model, its resolution and uncertainty. In Section 2, we introduce the SOLA method and the forward modelling approach. Section 3 details the tomography setup, including the data geometry, target resolution and generalised inverse. Subsequently, we discuss the data and their uncertainty in detail in Section 4, before presenting our tomographic results, both qualitatively and quantitatively in Section 5. Finally, we discuss the 3D resolution and uncertainty estimates of our model in Section 6 and indicate possible future directions.

2 Theory

We present here the main building blocks of our approach. Firstly, we briefly introduce the general forward problem. We then discuss the inverse problem, introducing the discrete linear SOLA inverse method (Zaroli, 2016) that provides control and produces full resolution and uncertainty information together with the tomographic model. Finally, we present the finite-frequency theory that allows the surface-wave inverse problem to be expressed in a linear and fully three-dimensional framework.

2.1 General forward theory

Let $\boldsymbol{d} \in \mathcal{R}^N$ be a data vector and let $\boldsymbol{m} \in \mathcal{R}^M$ be a model vector containing model parameters given a pre-defined parameterisation. Let $\boldsymbol{G} \in \mathcal{M}(N \times M)$ be the sensitivity matrix (in the set of matrices of size $N \times M$), describing a linear relationship between model parameters and data. We can then write the forward problem as:

$$\boldsymbol{d} = \boldsymbol{G}\boldsymbol{m} \tag{1}$$

Rows of $\boldsymbol{G}$ are the sensitivity kernels and $\boldsymbol{G}$ thus contains all the information regarding the sensitivity of the entire dataset to all model parameters; this is what we refer to as the data geometry.

To account for data errors, we treat $\boldsymbol{d}$ as a normally distributed multi-variate random variable with data covariance matrix $\boldsymbol{C_d} \in \mathcal{M}(N \times N)$. We assume uncorrelated noise, thus the data covariance matrix is diagonal and we can write $\boldsymbol{C_d} = \text{diag}(\sigma_{d_i}^2), i \in [|1, N|]$, where σ_{d_i} is the standard deviation of the error on the i^{th} datum, i.e. the data uncertainty. Note that under the Gaussian hypothesis both theoretical errors (due to imperfect forward theory) and measurement errors (due to imperfect measurements) are included in $\sigma_{d_i}^2$ (see e.g. Tarantola, 2005).

2.2 SOLA inverse method

Poor data geometry in seismic tomography makes the inverse problem ill-constrained: the sensitivity matrix $\boldsymbol{G}$ is not invertible. This justifies the use of various methods for obtaining model solutions (see e.g. Parker, 1977; Trampert, 1998; Scales and Snieder, 1997; Nolet, 1985; Tarantola and Valette, 1982; Wiggins, 1972; Nolet, 2008). Let $\boldsymbol{G}^\dagger$ be a 'generalised inverse' such that the model solution is expressed as linear combinations of the data:

$$\widetilde{\boldsymbol{m}} = \boldsymbol{G}^\dagger \boldsymbol{d}. \tag{2}$$

Using Equation 1, we obtain a relation between the model solution and the 'true' model:

$$\widetilde{\boldsymbol{m}} = \boldsymbol{G}^\dagger \boldsymbol{G} \boldsymbol{m}. \tag{3}$$

Each parameter in the model solution is a linear combination of the 'true' model parameters linked by the resolution matrix $\boldsymbol{R} = \boldsymbol{G}^\dagger$~~$\boldsymbol{G}$. In other words,~~ this means that the value of a model parameter in the model solution represents a spatial weighted average of the whole true model (plus some errors propagated from data noise). The resolution for one model parameter is determined by one such averaging and is referred to as 'resolving' or 'averaging kernel'. In general, we will want the averaging for a model parameter to be focused around that parameter location. The full resolution matrix thus acts as a 'tomographic filter' (e.g. Ritsema et al., 2007; Schuberth et al., 2009; Zaroli et al., 2017). Note that in the hypothetical case where the data geometry constrains all model parameters perfectly, the sensitivity matrix is invertible, the generalised inverse is the exact inverse, the resolution matrix is the identity matrix, and, in the case of error-free data, the model solution is exactly the true model.

The model uncertainty is propagated from the data uncertainty using the diagonal elements of~~:~~

$$\boldsymbol{C}_{\widetilde{\boldsymbol{m}}} = (\boldsymbol{G}^\dagger)^T \boldsymbol{C_d} \boldsymbol{G}^\dagger, \tag{4}$$

where ^{T} denotes the matrix transpose. We define the model uncertainty as the square root of the diagonal of the model covariance matrix, i.e. $\sigma_{\widetilde{m}_k} = \sqrt{C_{\widetilde{m}_{kk}}}$ is defined as the model uncertainty for model parameter k. In summary, the generalised inverse $\boldsymbol{G}^\dagger$ determines the model solution, model resolution and model uncertainty.

Most inverse methods are based on a data-misfit point of view. They solely consider the forward problem (Equation 1) and seek a model solution that minimises the distance between predicted and observed data. These approaches do not directly control the resolution and uncertainty of the solution and estimating these can be challenging depending on the inverse method used. To overcome this issue, we use the SOLA method, which is based on Backus-Gilbert theory (Backus and Gilbert, 1967, 1968, 1970; Pijpers and Thompson, 1992, 1994; Zaroli, 2016). With SOLA, we explicitly design $\boldsymbol{G}^\dagger$ to achieve certain objectives for the resolution and model uncertainty. In particular, we design a target resolution $\boldsymbol{T}$ and seek a generalised inverse that leads to a resolution close to the target. At the

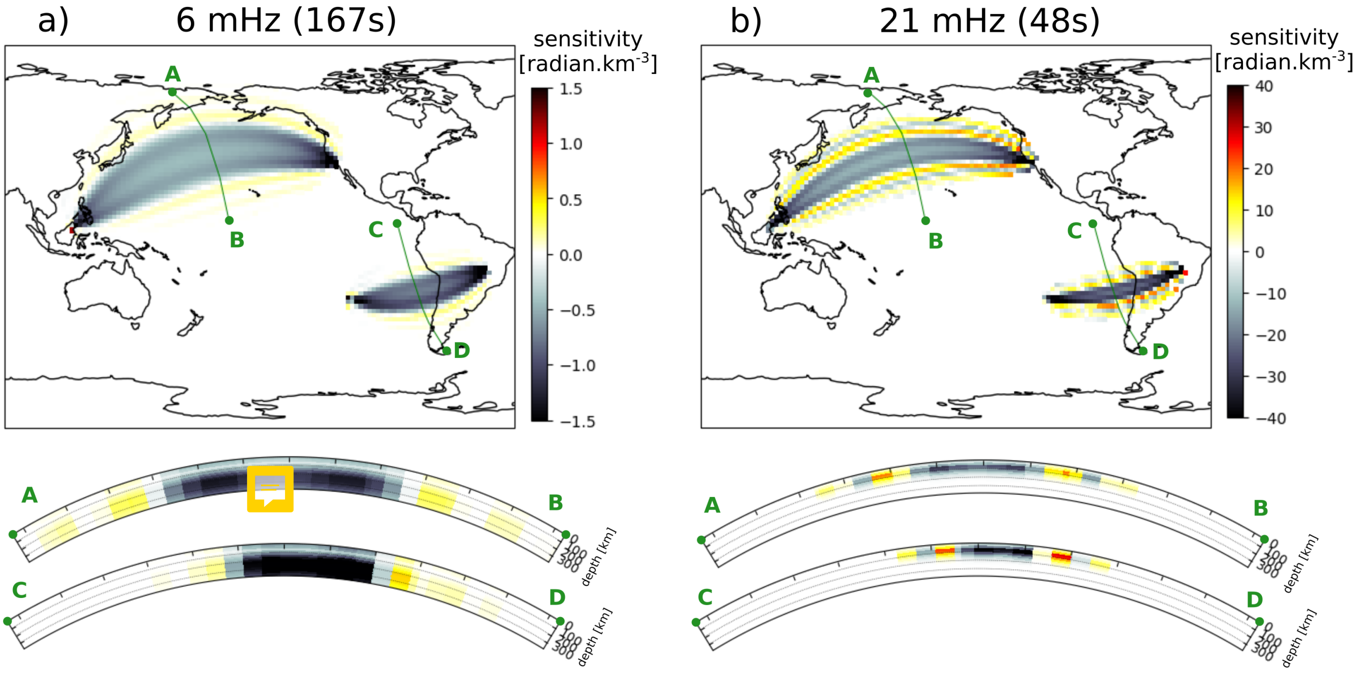


Figure 1 Examples of sensitivity kernels at (a) 6 mHz and (b) 21 mHz for two source-receiver pairs. The maps are plotted at depths of 87 km and 237 km depth respectively, which are the depths where the kernels reach their respective maximum amplitudes. Below each map, we also show a vertical cross-section through each kernel, as indicated on the maps. The northern kernel is for a Mw=6.1 earthquake in Borneo (2015) recorded by station DSN5. The southern kernel is for a Mw=6.1 earthquake in the Easter Island region (2011) recorded by station BDFB. Note the difference in amplitude between the two frequencies.

same time, we aim to minimise model uncertainty. These are two contradictory objectives that are balanced in an optimisation problem:

$$\underset{\boldsymbol{G}^{\dagger k}}{\arg\min} \sum_j [A_j^k - T_j^k]^2 \mathcal{V}_j + {\eta^k}^2 \sigma^2_{\widetilde{m_k}}, \quad \text{s.t.} \sum_j R_j^k = 1, \tag{5}$$

where k is the index of the model parameter we are solving for (the target), j is a dummy index that iterates over model parameters, $\mathcal{V}_j$ is the volume of cell j, $A_j^k = R_j^k / \mathcal{V}_j$ is the averaging (or resolving) kernel (normalised by the cell volumes), and η^k is a trade-off parameter that balances the fit to the target resolution with the minimisation of model uncertainty. The constraint $\sum_j R_j^k = 1$ guarantees that local averages are unbiased, another ~~striking~~ difference with data-fitting approaches as demonstrated by Zaroli et al. (2017). The optimisation problem leads to a set of equations (see Appendix A1 from Zaroli, 2016) that we solve for each model parameter using the LSQR algorithm of Paige and Saunders (1982), as suggested by Nolet (1985).

The SOLA inversion is point-wise, i.e. the minimisation problem is solved for each parameter independently from the others. This makes SOLA inversions straightforward to solve in parallel. Note that we do not need to solve for all model parameters nor do we need to solve for the whole region to which the data are sensitive (a necessity in data-fitting inversions): we have the possibility to solve only for model parameters of particular interest (the targets). We provide information on the computational costs of this study in Appendix B. ~~Also note that~~ the solution of the SOLA optimisation problem, $\boldsymbol{G}^{\dagger}$, does not depend on the data values themselves $\boldsymbol{d}$, which is an important difference with data-fitting methods.

2.3 Finite-frequency forward theory

In order to make the implementation of SOLA for surface-wave tomography fully three-dimensional, we need a linear relation between surface-wave data and 3D physical properties of the Earth mantle. Here, we consider vertical-component Rayleigh-wave phase delays measured at given frequencies ω for particular source-receiver pairs l. If we assume these are primarily sensitive to perturbations in the vertically polarized S-wave velocity δV_{SV} in the 3D mantle $\bigoplus$, we have the following relationship between data $\delta\phi_l(\omega)$ and model $\delta \ln V_{SV}(\boldsymbol{x})$:

$$\delta\phi_l(\omega) = \iiint_{\bigoplus} K_l(\omega;\boldsymbol{x})\delta \ln V_{SV}(\boldsymbol{x}) d^3\boldsymbol{x}, \tag{6}$$

where $\boldsymbol{x}$ indicates the ~~physical~~ location, and $K_l(\omega;\boldsymbol{x})$ is the sensitivity kernel.

Analytical expressions of surface-wave sensitivity kernels have been derived based on the scattering principle in the framework of normal mode theory. Here, we use formulations from Zhou et al. (2004), later extended to multi-mode surface waves and anisotropy by Zhou (2009b). These assume far-field propagation, single forward scattering, and use a paraxial approximation. Thanks to the single-scattering assumption, also known as Born approximation, the resulting relationship between data and model is linear, which makes it tractable with SOLA. The sensitivity kernels can be expressed as:

$$K(\omega;\boldsymbol{x}) = \mathrm{Im}\left(\sum_{n'}\sum_{n''} \frac{S'{}_{n'}\Omega_{n''}\, R'' e^{-i[k'\Delta' + k''\Delta'' - k\Delta + (s'+s''-s)\frac{\pi}{2} + \frac{\pi}{4}]}}{S\, R\sqrt{8\pi(\frac{k'k''}{k})(\frac{sin|\Delta'||sin|\Delta''|}{|sin\Delta|}}} \right). \tag{7}$$

Symbols with prime $'$ refer to the source-scatterer path, ones with double prime $''$ to the scatterer-station path, and those without prime to the great-circle source-station path. n is the overtone number (here we consider only fundamentals, so $n' = n'' = 0$), k the wave-number and s the Maslov index (here $s = 0$ or $s = 1$, i.e. single orbit). Δ is the path length, S the source radiation in the direction of the path, and R the projection of the polarisation onto the receiver orientation. The exponent term indicates the phase delay due to the detour by the scatterer, while the other terms express the relative amplitude of the scattered wave relative to the initial unperturbed wavefield. This relative strength depends on the source and receiver terms (the scattered wave leaves the source and arrives at the receiver with some angle compared to the unperturbed wave), on the geometrical spreading (the scattered wave makes a detour compared to the unperturbed wave), and on the scattering coefficient Ω. The scattering coefficient depends linearly on physical model properties, for which detailed expressions can be found in Zhou (2009a). In practice, we use a slightly different form of Equation 7 to include the effect of waveform tapering in the measurement algorithm (see Zhou et al., 2004, for more details).

We use routines from Zhou (2009b) to compute the sensitivity kernels for the fundamental mode, assuming self-coupling. We only compute these in the top 400 km of the mantle as their amplitude decreases sharply with depth. We consider the first two Fresnel zones laterally as their side-lobes become negligible further away. Examples of sensitivity kernels are given in Figure 1, where they are projected onto the tomographic grid. The kernels have particularly strong amplitude at the source and station. This is caused by a combination of natural high sensitivity at end-points of a path and the far-field approximation (e.g. Liu and Zhou, 2016b). Low-frequency kernels peak at deeper depths, have a broader lateral and vertical extent, and have weaker amplitudes than high-frequency kernels. Although the projection onto the tomographic grid degrades the shape and amplitude of the sensitivity kernels, their main properties are retained on a sufficiently-fine tomographic grid as is the case here.

3 Tomography setup

In this section, we present the construction of the forward problem (the sensitivity matrix) and the inverse solution (the generalised inverse) that determines the resolution, the propagation of data uncertainty into model uncertainty and data values into model estimate. We will describe the data and data uncertainty in the next section. These will feed into the inverse solution to produce the tomography model and the measurement model uncertainty.

3.1 Parameterisation

We use a local model parameterisation and split the 3D spatial domain into voxels of size $2^\circ \times 2^\circ$ laterally (latitude and longitude) and 25 km depth vertically. We parameterise only the top 400 km depth, since the sensitivity of fundamental mode surface waves to V_{SV} becomes negligible at greater depths. This leads to $M = 259\,200$ voxels to parameterise the top 400 km depth of the mantle globally. It is worth recalling that with SOLA we do not need to solve for all M model parameters nor for the whole region to which the data are sensitive. For example, we could solve only for cells where the data sensitivity is sufficiently high or only for a particular region of interest.

3.2 Data geometry

We select 312 earthquakes with M_w between ~6.0 and 7.7 and a depth between ~12 and 87 km, all located in the Pacific region, occurring between July 2004 and December 2020. We consider 1228 stations, also located in the Pacific region (see Fig. 2). Sources and stations are both selected in a way to avoid strong spatial redundancy. For all paths, we consider 16 frequencies ranging from 6 to 21 mHz (48-167s), in steps of 1 mHz.

Compared to ray-theory, finite-frequency theory is fully three-dimensional. This makes the sensitivity matrix larger because we need to consider the whole 3D spatial extent of the model domain all at once, and less sparse because finite-frequency sensitivity kernels have a volumetric extent. This is a challenging issue that limits the number of data we can take into account in the inversion. For a computational node with 254 GB of RAM, and our current strategy for storing matrices in RAM, we estimate that we can incorporate at most $N = 300\,000$ measurements (more information on the computational costs of this study is given in Appendix B). Here, we restrict ourselves to $N \approx 50\,000$ measurements, making it possible to expand our work to overtones in the future. To achieve $N \approx 50\,000$ data, we carefully select our data with the aim to homogenise the lateral distribution of rays (see Section 4). We end up with 47,700 data in total, with approximately 3,000 data per frequency (figure 2).

For each selected measurement, we compute the corresponding 3D finite-frequency sensitivity kernel to build the sensitivity matrix $\boldsymbol{G}$, with examples shown in Figure 1. As a measure of the constraint offered by the data on the structure of the 3D upper mantle, we compute the decimal logarithm of the data sensitivity, $\log_{10} \sum_i |G_{ij}|$, where i and j designate a particular datum and model parameter respectively (see figure 2, lower right).

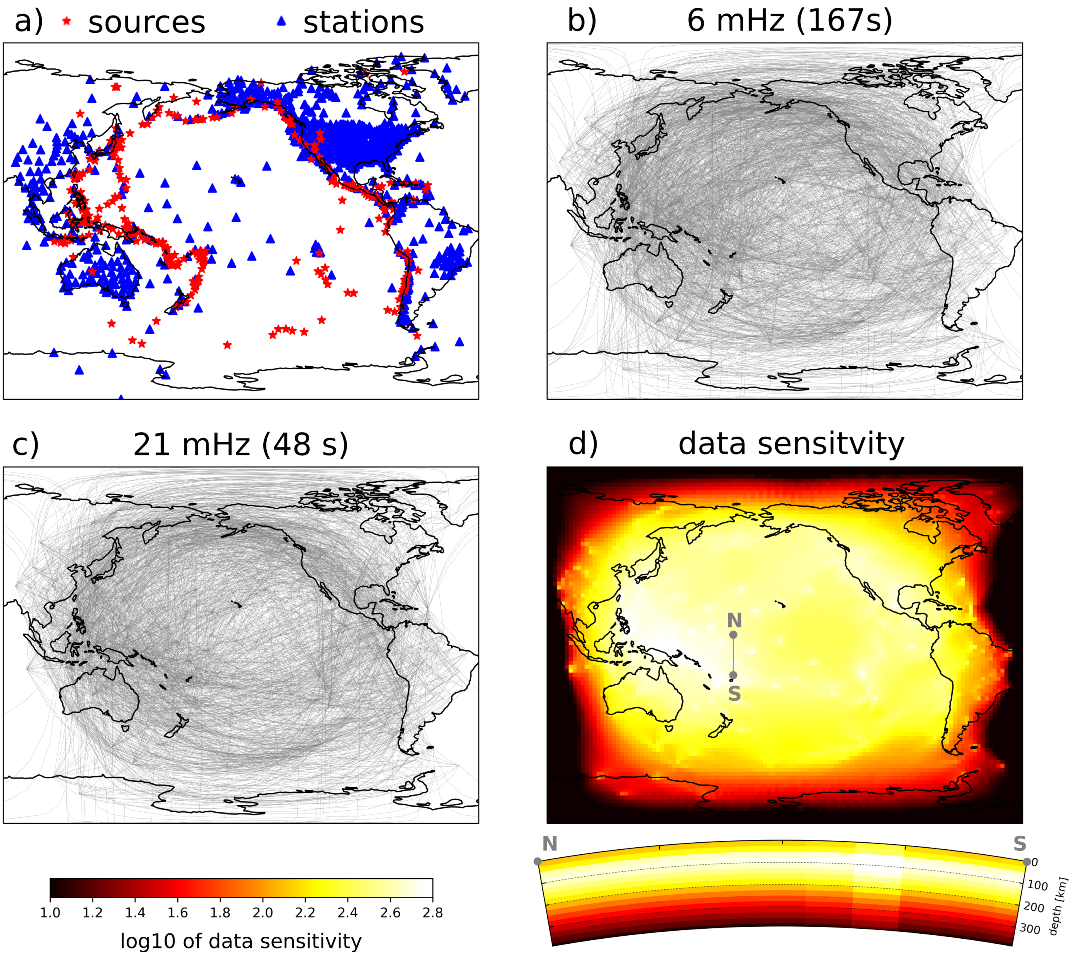


Figure 2 Data geometry of our tomography, showing (a) the distribution of sources and receivers, (b) the selected ray paths at 6 mHz and (c) at 21 mHz, and (d) the decimal logarithm of the data sensitivity, $\log_{10} \sum_i |G_{ij}|$. The data sensitivity is plotted at 112 km depth, with a N-S oriented vertical cross-section below it, indicated by the grey line on the map view.

3.3 Target resolution, uncertainty propagation, and tradeoff

The shape of the target kernels used in the SOLA inversion is arbitrary. Ideally, it is chosen such as to produce results oriented towards addressing a specific ~~key~~ question. In this study, we wish for the resolution to represent simple, easy-to-interpret 3D local averages. For a given model parameter, we therefore choose the target kernel to be a 3D ellipsoid. The lateral resolution we can achieve with surface-wave data is controlled by the distribution of sources and receivers (and, to some extent, frequency). Our experience shows that it is rarely better than a few hundreds of kilometers for the frequency range used here. The vertical resolution is mostly controlled by the frequency content of the signal and it is typically on the order of tens to hundreds of kilometers. Therefore, a reasonable target kernel at a given point in the 3D grid would resemble a flat pancake centered at the query point. More formally, we design the target kernel of a model parameter as an ellipsoid whose major and semi-major axes are equal and aligned with the north-south and east-west directions at the location of the model parameter, and whose minor axis is vertical. The resulting target kernels are thick versions of the 2D kernels of Latallerie et al. (2022) and Amiri et al. (2023) and they represent a horizontally isotropic target resolution.

With SOLA, it is possible to adapt the size of the target kernels for each model parameter (i.e. for each location). For example, we could choose to achieve the best resolution possible at each location in the model given the data coverage, or we may prefer a homogeneous resolution or constant uncertainty across the spatial domain (see Freissler et al., 2024). This freedom illustrates the typical non-uniqueness of tomographic inversions. Any choice that fits the purpose of the study can be considered 'good', so long as the tomographic model is analysed together with its resolution and uncertainty. In this study, for simplicity, we make all target kernels the same, with 200 km long horizontal major and semi-major axes and 25 km long vertical minor axis. Figures 3 and 4 illustrate the extent of our target kernels for a selection of 10 model parameters (blue ellipses).

The data uncertainty ~~potentially~~ influences the solution to the inverse problem (second term of Equation 5). However, as we aim to study the robustness of the data uncertainty itself in this study, we decide not to take them into account in solving the inverse problem. Thus, we initially set $\boldsymbol{C_d} = \boldsymbol{I}$ and therefore $\boldsymbol{C_{\widetilde{m}}} = (\boldsymbol{G}^\dagger)^T\boldsymbol{G}$. Note that this is only a choice for solving the optimisation problem: once the generalised inverse has been computed, we still consider non-unitary data uncertainty and propagate it into model uncertainty through $\boldsymbol{C_{\widetilde{m}}} = (\boldsymbol{G}^\dagger)^T\boldsymbol{C_d}\boldsymbol{G}$.

The optimisation problem involves the minimisation of the difference between target and actual resolution on the one hand, and the magnitude of model uncertainty on the other hand. These two terms are balanced by the trade-off parameter η, which we set equal to 50 for all parameters. Again, it is possible to choose different values of η for different model parameters, but in practice it is computationally easier to keep η constant (see Appendix A1 of Zaroli, 2016). If, for example, one wants to give more weight to the resolution of a particular model parameter, this can also be obtained by designing a smaller size target kernel. If we vary the trade-off parameter, we obtain a typical L-shaped trade-off curve for resolution versus model uncertainty (Latallerie et al., 2022; Restelli et al., 2024), which could be used to pick the appropriate η value for the study at hand.

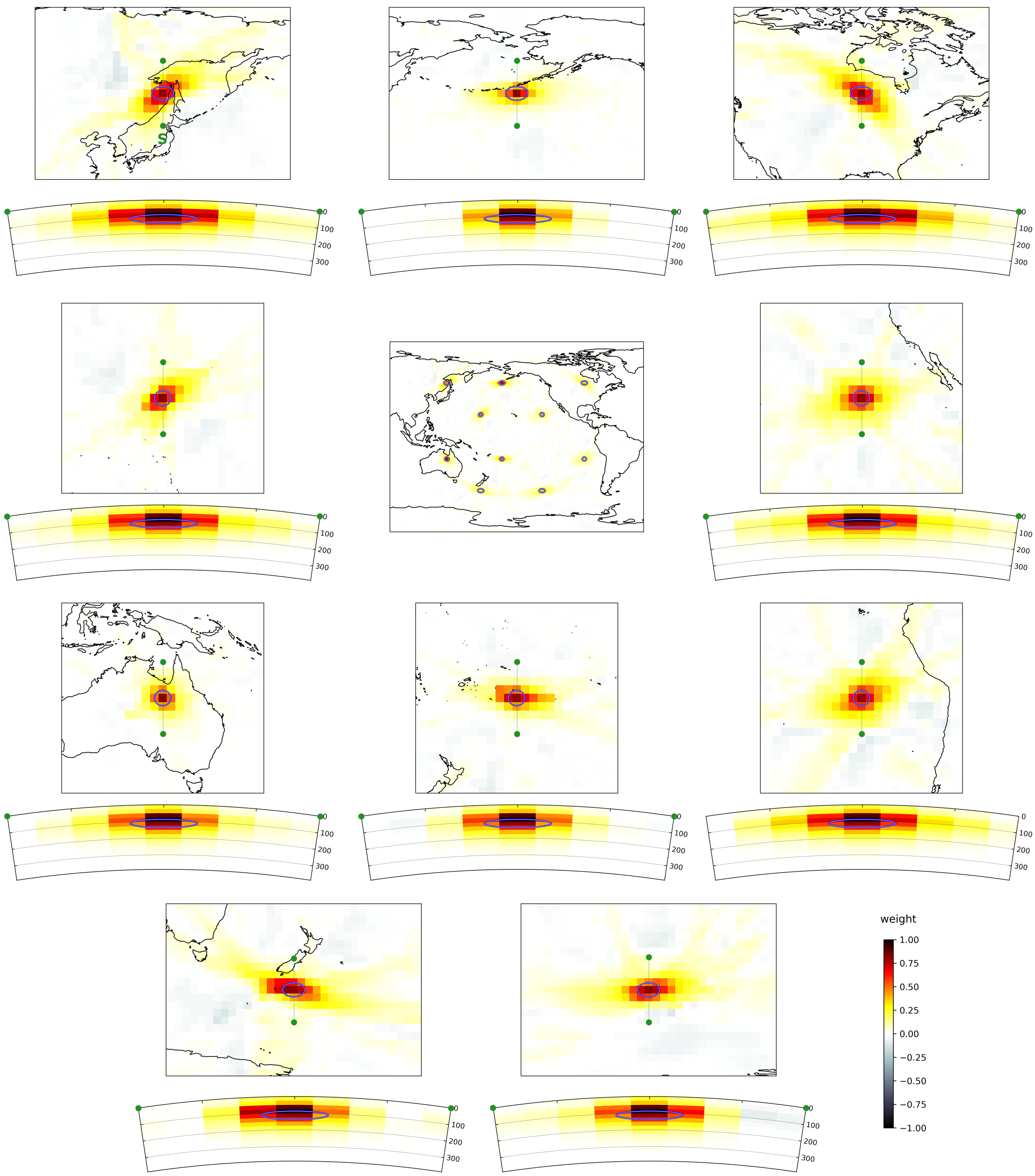


Figure 3 Resolution at 112 km depth illustrated for a selection of 10 model parameters. The centre map shows the locations of the 10 target and resolving kernels. This is shown as a sum, which may exaggerate the apparent strength of the tails. The surrounding panels are close-ups on individual kernels, both in map-view and as cross-section. All maps represent depth slices at 112 km depth and below each map is a $\sim$3100 km long, N-S oriented (left to right) cross-section as indicated in green in the maps. The depth in km is indicated on the right of each cross-section. Blue ellipses show the lateral extent of the target kernels. All averaging kernels are normalised by their maximum, and the color scale indicated in the lower right applies to all panels.

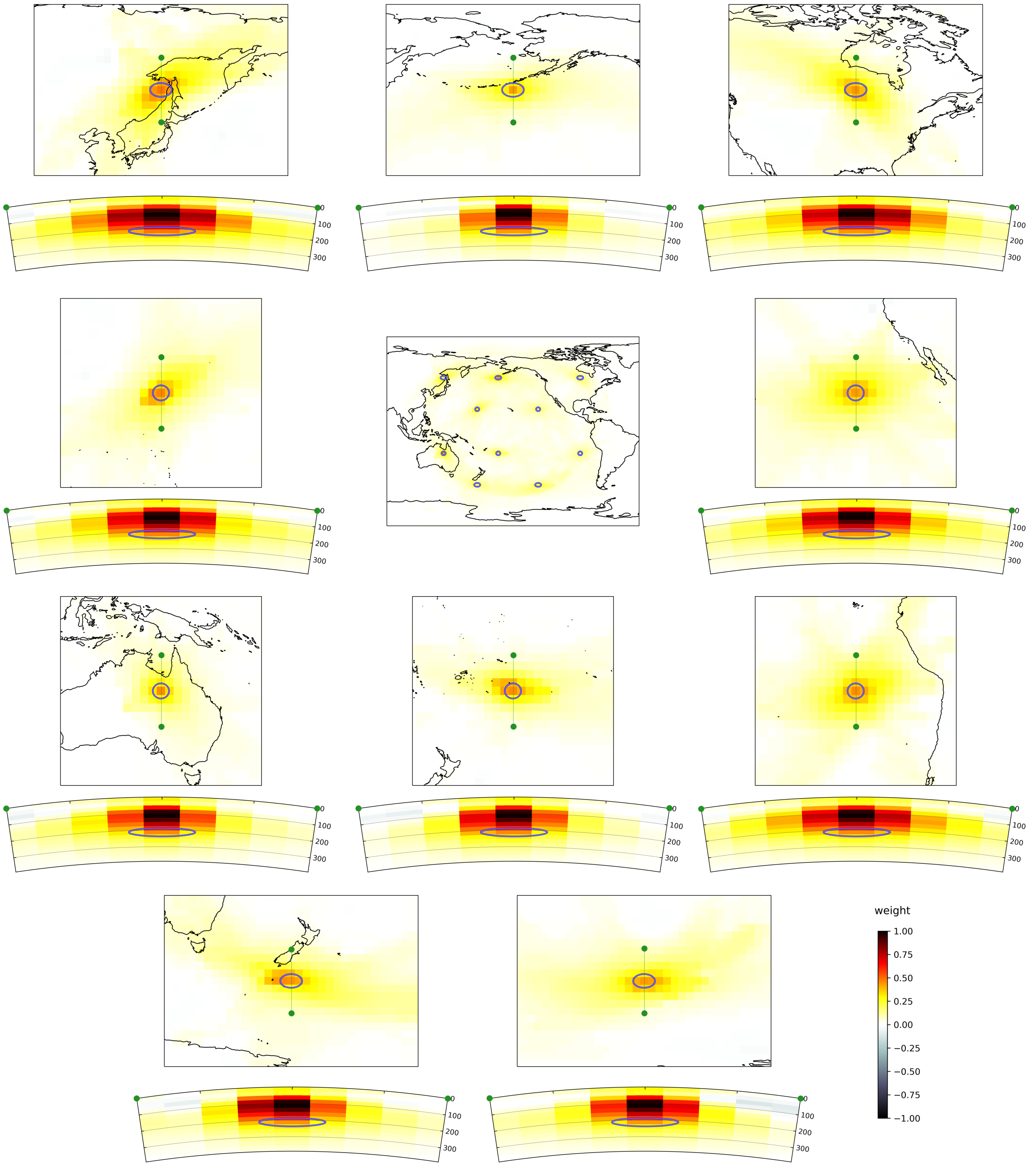


Figure 4 Same as figure 3, but for target locations at 212 km depth.

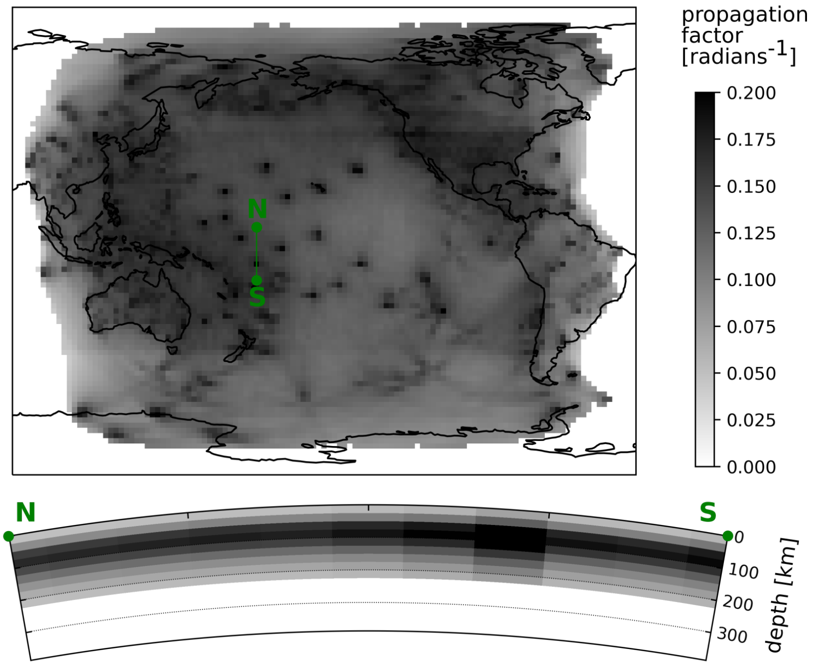


Figure 5 Illustration of the propagation of data uncertainty into model uncertainty. The map shows the 'propagation factor' at 112 km depth, defined as the model uncertainty given unit data uncertainty. The cross-section below the map indicates the depth dependence of the propagation factor along a vertical 2500-km long N-S oriented profile as indicated by the green line on the map.

3.4 Generalised inverse: Resolution and uncertainty propagation

The seismic tomography inversion is fully characterised by the generalised inverse $\boldsymbol{G}^\dagger$: it determines the resolution (from $\boldsymbol{R} = \boldsymbol{G}^\dagger \boldsymbol{G}$) as well as the propagation of data uncertainty into model uncertainty (from $\boldsymbol{C}_{\widetilde{\boldsymbol{m}}} = (\boldsymbol{G}^\dagger)^T \boldsymbol{C}_d \boldsymbol{G}^\dagger$). Lastly, it determines the propagation of data into model solution (from $\widetilde{\boldsymbol{m}} = \boldsymbol{G}^\dagger \boldsymbol{d}$).

It is difficult to represent the full 3D resolution as it is most easily understood in terms of an extended 3D resolving kernel associated with each model parameter. A detailed analysis thus requires 3D rendering software or the production of simple proxies, for example those proposed by Freissler et al. (2024). Here, we instead illustrate the resolution by selecting example resolving kernels. At 112 km depth (Figure 3), the resolving kernels match the target location well laterally. Their lateral size is roughly 250-450 km (if we take the radii of a circle containing 68% of the kernel). This can be compared to the length of the major and semi-major axes of the target kernels of 200 km. Some averaging kernels are significantly anisotropic, indicating lateral smearing due to the heterogeneous ray path distribution. Vertically, the resolving kernels appear also to be focused with a half-thickness of roughly 50 km. This can be compared to the length of the minor axis of the target kernels of 25 km. However, they appear slightly shifted upward from the target. Deeper down, at 212 km depth (Figure 4), the resolving kernels still match the target locations laterally, but they appear broader (300-700 km). They now also poorly match the target kernel depth-wise. Instead of peaking at 212 km depth, the resolving kernels peak at 112 km depth and tail off deeper down. This implies that what we observe in the tomographic model at 212 km depth is actually an average of the 'true model' at shallower depth.

We show the 'error propagation factor' in Figure 5. This can be interpreted as the model uncertainty for unit data uncertainty ($\boldsymbol{C}_d = \boldsymbol{I}$), obtained from $(\boldsymbol{G}^\dagger)^T \boldsymbol{G}^\dagger$. We observe a positive correlation between data coverage and error propagation factor: the error propagation tends to be high where data coverage is high (e.g. North America, South-East Asia). We also clearly see patches of high error propagation in the Pacific Ocean at locations of isolated stations. This is due to the high data sensitivity at stations where many oscillatory sensitivity kernels add together. Furthermore, we note linear features with high error propagation that follow great-circle paths radiating away from some isolated stations. These probably outline sensitivity kernels that repeatedly sample similar regions. With depth, we find that the propagation factor increases down to 87 km depth and then decreases again deeper down. While this decrease may be surprising, it is balanced by poor resolution at greater depth. In general, SOLA tends to produce models with better resolution where data sensitivity is high, at the cost of a larger error propagation factor. By choosing different sizes for the target kernels, this can be balanced (Freissler et al., 2024).

4 Input data and measurement uncertainty

We measure phase delays between 'observed' and 'reference' seismograms for 16 different frequencies ranging from 6 to 21 mHz (48-167s), in steps of 1 mHz. In this synthetic study, we use as 'observed seismograms' waveforms computed using SPECFEM3D_GLOBE (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 2002) for the 3D input model S362ANI (Kustowski et al., 2008) combined with CRUST2.0 on top (Bassin et al., 2000). Hereafter, we refer to these as SEM seismograms or SEM measurements. These were obtained from the GlobalShakeMovie project data base (Tromp et al., 2010) and downloaded from Earthscope, formerly IRIS (IRIS DMC, 2012; Hutko et al., 2017). Reference seismograms were computed using normal-mode summation with the Mineos software (Masters et al., 2011) for the 1D radial model stw105 (Kustowski et al., 2008), consistent with S362ANI. For both sets of seismograms, we use source solutions obtained from the Global-CMT project (Ekström et al., 2012) and station metadata from Earthscope. To measure the phase delay between the two sets of seismograms, we use a multi-taper measurement algorithm as suggested by Zhou et al. (2005) and detailed in appendix A. The multi-taper technique has the advantage of providing an estimate for the measurement data uncertainty as the standard deviation of the measurements across all tapers. This uncertainty estimate is particularly sensitive to cycle-skipping and contamination by higher modes and other phases.

Considering only source-receiver combinations for which the measurement time window does not include the event origin time (150 s before to 650 s after the predicted group arrival time), we obtain 2,414,515 measurements of Rayleigh wave phase delays. We select a subset of these measurements based on the following criteria: similarity between the seismograms (cross-correlation > 0.8), source radiation in the direction of the station (> 80% of maximum radiation), measurement uncertainty (< 1.9 radians), outliers removal (1% of the dataset). This leads to 564,940 potential measurements. Due to memory limitations (as explained in section 3.2), we select a subset of $N = 47,700$ data to reduce the size of $\boldsymbol{G}$. This is achieved by randomly selecting one ray, then removing all rays whose endpoints are within 800 km radius of the endpoints of the selected ray, and repeating this process until we reach the desired number of measurements, at the frequency of interest. This gives the vector of measured data that we denote $\boldsymbol{d}^{\text{measured}}$. As a check, we also compute the corresponding analytical data $\boldsymbol{d}^{\text{analytical}}$ by applying our forward theory $\boldsymbol{G}$ to the 3D input model S362ANI ($\boldsymbol{m}^{\text{input}}$), i.e. $\boldsymbol{d}^{\text{analytical}} = \boldsymbol{G}\boldsymbol{m}^{\text{input}}$.

As we do not invert for the crustal structure, we need to apply a crustal correction to our measurements (e.g.

Marone and Romanowicz, 2007; Bozdağ and Trampert, 2008; Panning et al., 2010; Liu and Zhou, 2013; Chen and Romanowicz, 2024). For consistency with the synthetic 'observed' waveforms, we also use CRUST2.0 to compute the crustal correction (Bassin et al., 2000). We first construct 1D radial models for a combination of stw105 and CRUST2.0 at every location in a 2°x2° grid. For each grid point, we then solve a normal-mode eigenvalue problem using Mineos (Masters et al., 2011) to obtain the local phase velocity, thus building phase velocity maps for the reference model with the added crustal structure. For each source-receiver path and all frequencies in our dataset, we subsequently compute the phase accumulated in this model $\phi^{\text{ref+crust}}$ as well as in the reference model ϕ^{ref}, assuming ray-theory (i.e. great-circle approximation). The difference in phase due to the crustal structure $\delta\phi^{\text{crust}} = \delta\phi^{\text{ref}} - \delta\phi^{\text{ref+crust}}$ is then used to correct the measured data: $d^{\text{corrected}} = d^{\text{measured}} - \delta\phi^{\text{crust}}$.

Examples of our dispersion measurement procedure and results are given in Figure 6 and used to illustrate three typical cases. Some of our measurements agree well with the analytical predictions and have low uncertainty (left column). This case is representative of 19% of the final dataset, with a difference of less than 1 radian between measured and analytical data. This difference is also less than 3 times the measurement data uncertainty. Other measurements do not agree well with the analytical predictions (middle column), but this is compensated by high data uncertainty. This case is representative of 10% of the final dataset, with a difference of more than 1 radian between measured and analytical data. This difference is still within 3 times the measurement data uncertainty. The last column shows a more problematic case: the measurement has low uncertainty, but does not match the analytical prediction. It appears that the measurement algorithm has failed to detect a cycle-skip around 8 mHz. Since the measurements are consistent for all tapers, the uncertainty remains low in this case. Therefore, the final measurement includes a cycle-skip difference with the analytical data above 8 mHz that is not reflected in the uncertainty. This case is representative of 67% of the final dataset, with a difference of more than 1 radian between measured and analytical data. This difference is greater than 3 times the measurement data uncertainty. Note that these discrepancies are due both to errors in the measurement (poorly measured data), that may be underestimated, but also to errors in the forward theory (poor analytical data), which we ignore at this stage.

Figure 7 presents statistics summarising our measurements and associated uncertainty. Our measured phase delays are typically larger than the analytical predictions ($d^{\text{analytical}} = Gm^{\text{input}}$) for both positive and negative delays, possibly due to non-linear effects. We may therefore expect increased positive and negative anomalies in our resulting tomographic model. We also observe a parallel branch of negative measured phase-delays with respect to the analytical predictions, likely due to non-detected cycle-skips. Our measurement uncertainty peaks around 0.3-0.5 radians, with the peak uncertainty shifting to higher values (to the right) for higher frequencies (darker colours). The effect of this shift on the resulting model uncertainty is not easy to predict as different frequencies impact the model solution in different ways (e.g. low frequency data have overall lower sensitivity). We also observe two additional peaks for higher uncertainty values, probably due to cycle-skipping and contamination with higher modes. However, measurements with these uncertainty values are not included as we apply a cutoff of 1.9 radians in our data selection.

We now have a dispersion data set with an estimate of the measurement uncertainty. As described above, the measurement uncertainty provided by the measurement algorithm accounts for cycle-skips and contamination by other phases or higher modes, but not fully. Moreover, it does not capture the theoretical errors. We aim to estimate

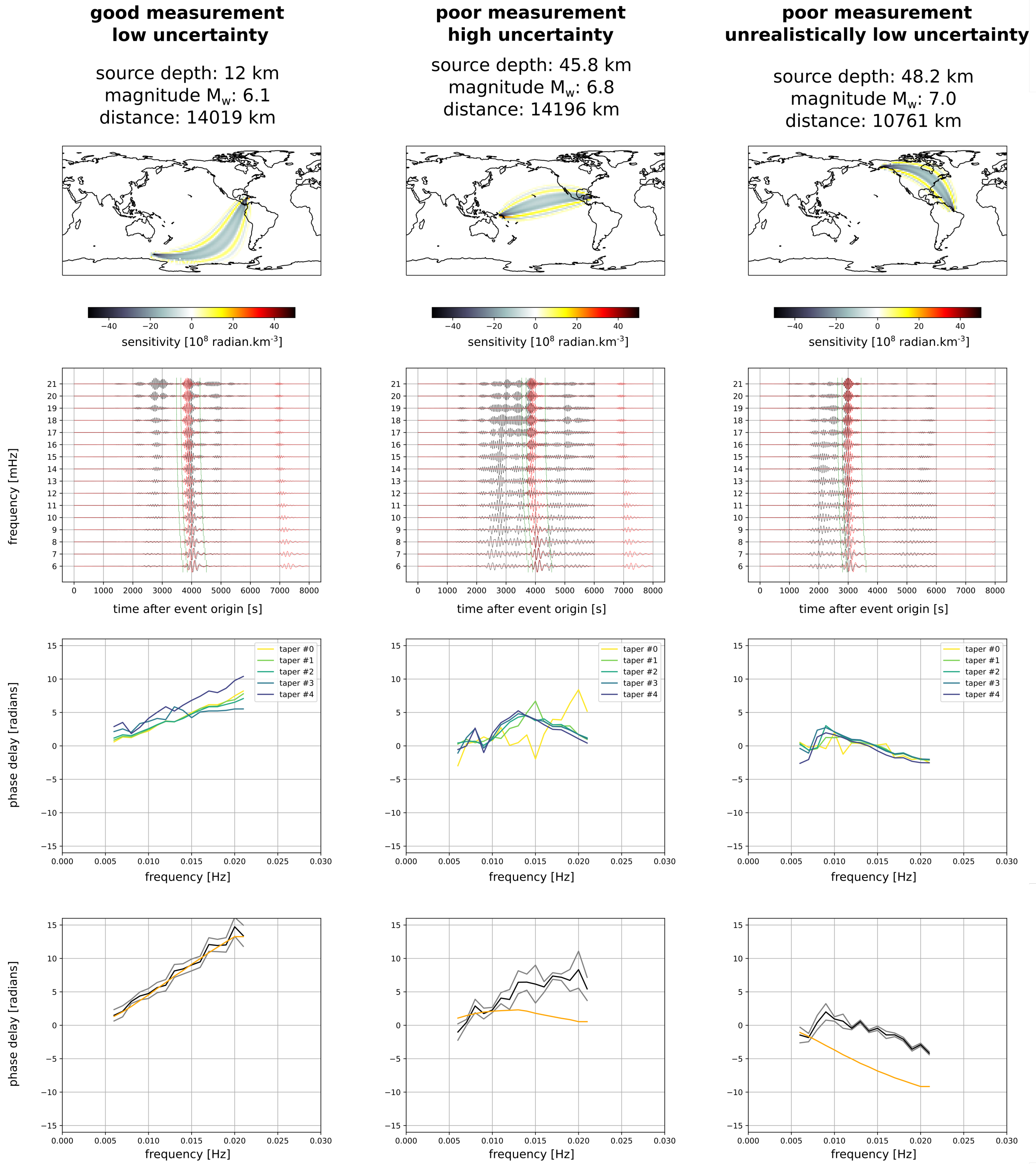


Figure 6 Example dispersion measurements, showcasing three typical cases. For each case, we include the sensitivity kernel at 16 mHz, plotted at 112 km depth (top row); the seismic traces (second row) for 8000 s after the event origin time (reference in black, SEM in red), filtered around each measurement frequency, and the green vertical lines indicate the start and end times of the applied tapers, around the predicted group arrival time; the measured dispersion for each taper (third row); and the final dispersion measurement (bottom row) averaged over all tapers (black) with the estimated uncertainty (grey), compared with the analytical prediction (orange). In the last row, the crustal correction is also applied to the measurements

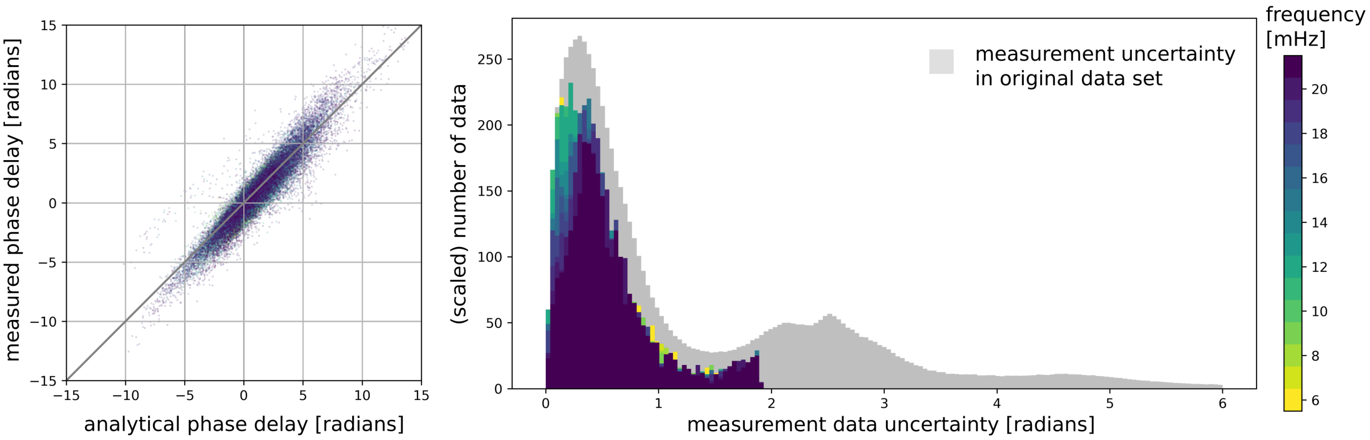


Figure 7 Summary of data and measurement uncertainty. Left: Cross-plot of the measured phase delay (after crustal correction) versus the analytical phase delay prediction, coloured by frequency. Positive phase-delays typically indicate slow velocity anomalies. Right: Distribution of measurement data uncertainty (coloured by frequency) before (grey) and after applying several selection criteria. Our selection criteria include a threshold for the data uncertainty (lower than 1.9 radians, as visible in the plot). The distribution of the measurement uncertainty before applying the selection criteria is scaled by 0.003 to enhance its visibility.

these in the following section.

5 Results

In the perfect case of error-free analytical data $\boldsymbol{d}^{\text{analytical}}$, an inversion should produce a model solution that is exactly the same as the filtered input. We confirm that by comparing the analytical model solution $\widetilde{\boldsymbol{m}}^{\text{analytical}} = \boldsymbol{G}^{\dagger}\boldsymbol{d}^{\text{analytical}}$ to the filtered input $\boldsymbol{R}\boldsymbol{m}^{\text{input}}$. When we instead use the measurements on SEM waveforms $\boldsymbol{d}^{\text{corrected}}$, differences between the filtered input model $\boldsymbol{R}\boldsymbol{m}^{\text{input}}$ (Figure 8b) and the obtained model solution $\widetilde{\boldsymbol{m}}^{\text{output}}$ (Figure 8d) arise due to data errors. These errors are a combination of both measurement and theoretical errors. Only the former have been taken into account in the model uncertainty map shown in (Figure 8c). Note how the edges of the model solution appear rough. This is because we invert only for model parameters where the data sensitivity is higher than a certain threshold (depending on depth); this is possible due to the point-wise nature of the SOLA inversion.

5.1 Qualitative proof of concept: velocity models

The features in the input model (Figure 8a) are also mostly present in the filtered model (Figure 8b). This indicates that the model resolution is good, at least at 112 km depth. For example, we retrieve mid-ocean ridges (low velocities at the East-Pacific rise, Pacific-Antarctic ridge, the edges of the Nazca plate), the lithosphere cooling effect (increasing velocity with distance from the ridge), the ring of fire (low velocity in the back-arc regions behind subduction zones such as the Aleutian trench, Okhotsk trench, edges of the Philippine sea plate and the Tonga-Kermadec trench), and cratons (fast velocities within the Australian and North American continents). The amplitudes of the velocity anomalies in the filtered model are lower than in the input model. This is expected since the filtered model represents (unbiased) local averages (Zaroli et al., 2017). The filtered model is also rougher on short length scales compared to the input model. This can be explained by the local nature of SOLA inversions where each model parameter is inverted independently from the others. In this case, we notice this particularly because the input model itself is very smooth.

Some artefacts appear such as the fast velocity anomaly of SW Australia extending through the slow velocity of the Australian-Antarctic ridge. Some striations also appear in the fast velocity region in the NW Pacific, trending in the SW-NE direction. These artefacts are probably the result of anisotropic ray coverage, with many sources in East-Asia mostly recorded by stations in North-America. In addition to these artefacts, some local features disappear in the filtered model, such as the low velocity finger extending southward from the Aleutian trench, or the branch extending northwestward from Hawaii. Overall, the filtered input resembles the true input model well, as also reflected in the cross-sections underneath.

The resulting model solution based on SEM seismograms (Figure 8d) appears very similar to the filtered input (Figure 8b). Compared to the input and filtered input models described above, the model solution appears somewhat rougher due to the propagation of data errors into the model solution. The striations observed in the NW Pacific in the filtered model are also stronger in the model solution than in the filtered input. Finally, the cross-section indicates a good agreement between the filtered model and our model solution.

5.2 Quantitative proof of concept: uncertainty

Our model measurement uncertainty map (Figure 8c) is very similar to the 'uncertainty propagation factor' map in Figure 5. Uncertainty is typically higher where there are clusters of stations and at isolated stations with linear features following great circle paths. Uncertainty peaks at $\sim$ 87 km depth and decreases strongly deeper. In our SOLA inversions, the model uncertainty only arises from the propagation of data uncertainty (Equation 4). This means that robust data uncertainties need to be estimated in order for model uncertainties to be reliable. We estimate measurement data uncertainty at the measurement step. However, this estimate does not encompass the full uncertainty that should include effects due to theoretical errors. How much these contribute to the data uncertainty is generally difficult to determine, but the synthetic nature of this study allows us to estimate theoretical uncertainty and inform future studies.

We propose the following strategy to estimate the magnitude of the theoretical model uncertainty. Let $\boldsymbol{m}^{\mathbf{input}}$ and $\widetilde{\boldsymbol{m}}^{\mathbf{output}}$ be the input model and model solution respectively. Any discrepancy between the input model and model solution arises from the limited resolution and propagation of data uncertainty into model uncertainty. To rule out the effect of limited resolution, we apply the resolution to the input model to obtain the 'filtered' input model $\boldsymbol{R}\boldsymbol{m}^{\mathbf{input}}$. Therefore, in this synthetic setup, it is only the propagation of data errors into model errors that explains the discrepancy between the 'filtered' input model and the obtained model solution. This is confirmed by the fact that the model solution based on error-free analytical data reproduces the filtered input exactly. Let us define the model misfit normalised by the model uncertainty as:

$$\chi_{\widetilde{m}} = \sqrt{\frac{1}{\sum_{k\in\mathcal{P}} V_k} \sum_{k\in\mathcal{P}} V_k \frac{[(\widetilde{\boldsymbol{m}}^{\mathbf{output}})_k - (\boldsymbol{R}\boldsymbol{m}^{\mathbf{input}})_k]^2}{(\boldsymbol{\sigma}_{\widetilde{\boldsymbol{m}}})_k^2}}, \tag{8}$$

where k refers to the model parameter index, V_k is the volume of voxel k, $\mathcal{P}$ is the set of model parameters considered for the analysis, and $\boldsymbol{\sigma}_{\widetilde{\boldsymbol{m}}}$ refers to the model uncertainty estimate.

If the data uncertainty is well-estimated, then $\chi^2_{\widetilde{m}} = 1$. As an experiment, we add random noise with a known distribution to the analytical data (i.e. to those obtained using $\boldsymbol{d}^{\mathbf{analytical}} = \boldsymbol{G}\boldsymbol{m}^{\mathbf{input}}$). In this case, the simulated

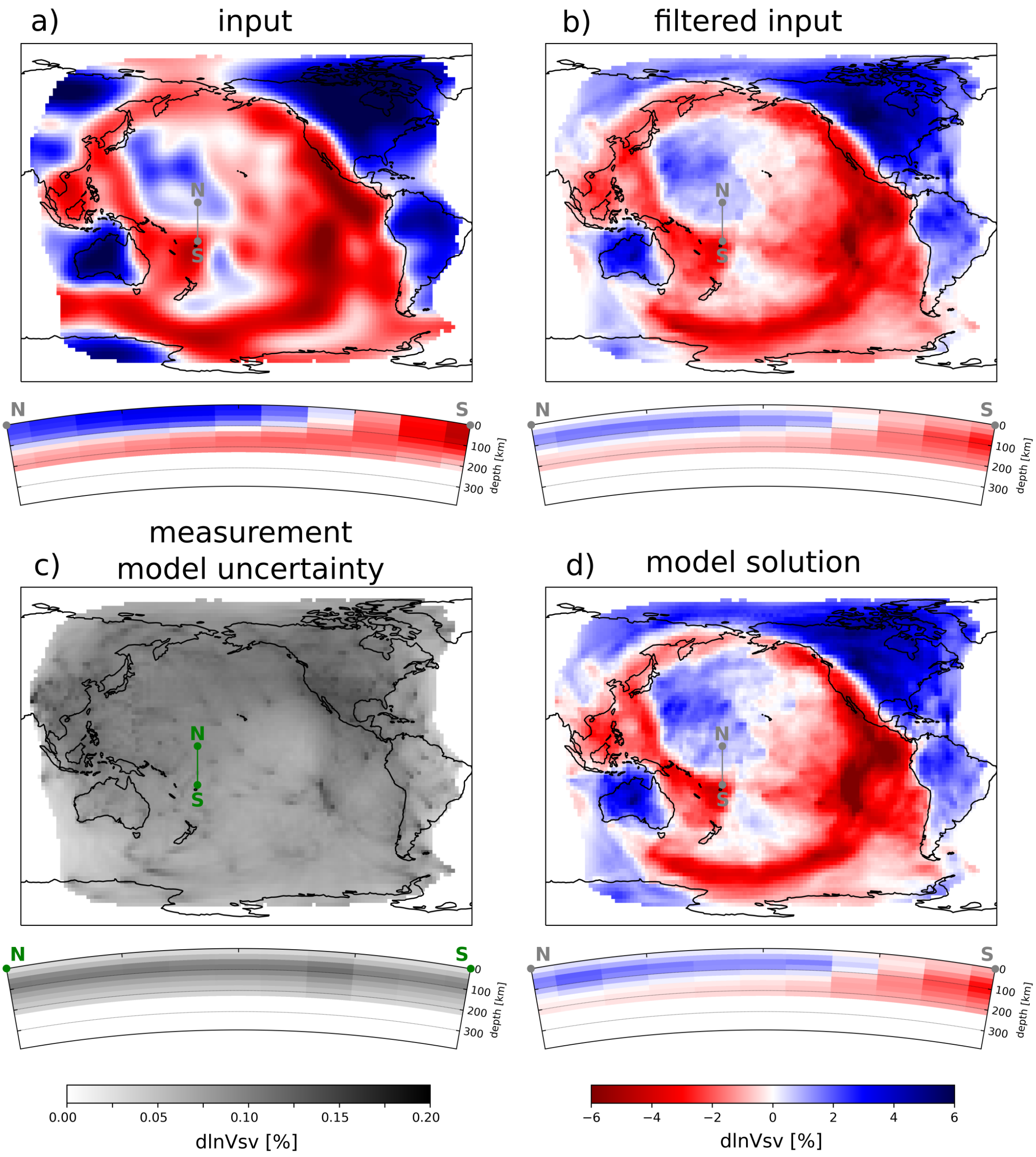


Figure 8 Summary of synthetic inversion results, comparing (a) input model S362ANI, (b) input model S362ANI filtered using our resolution matrix, (c) the model measurement uncertainty (propagated from data measurement uncertainty), and (d) the model solution retrieved using the measured data values (based on the SEM seismograms). All maps represent depth slices at 112 km depth, as in Figure 3. Below each map is a N-S vertical cross-section with the location indicated by the grey or green line on the maps.

data uncertainty is perfectly known and we obtain exactly $\chi^2_{\tilde{m}} = 1$. In the case of our synthetic tomography with phase delays measured on SEM waveforms, we obtain $\chi^2_{\tilde{m}} \approx 33 \gg 1$ when we only consider the propagation of data measurement uncertainty into model measurement uncertainty. This model uncertainty estimate is dramatically under-estimated as we may have underestimated the data measurement uncertainty and/or lack the theoretical uncertainty. We thus need to either upscale or add another component to the model uncertainty to account for this. We can write:

$$\sigma_{\tilde{m}_k}^{\text{total}\,2} = \alpha^2 \sigma_{\tilde{m}_k}^{\text{measurement}\,2} + \beta^2 \tag{9}$$

Here, α is the factor needed to upscale the model measurement uncertainty to account for the fact the measurement uncertainty itself might be underestimated. β is the theoretical uncertainty term that appears as an added component. We can now vary α and β independently and investigate for which combinations we obtain $\chi^2_{\tilde{m}} = 1$. Note that in this analysis the scaling factor α and the added uncertainty component β are both assumed to be constant over all model parameters involved (consisting here of all model parameters for V_{SV} at 112 km depth).

Figure 9 shows the evolution of $\chi^2_{\tilde{m}}$ for various combinations of α and β. We use this plot to illustrate three distinct cases. (i) The model measurement uncertainty serves as total model uncertainty, i.e. no upscaling nor added component, i.e. $\alpha = 1$ and $\beta = 0$. In this case, $\chi^2_{\tilde{m}} \approx 33$ falls in the under-estimated uncertainty region. (ii) We upscale the model measurement uncertainty without adding a component to obtain $\chi^2_{\tilde{m}} = 1$, i.e. $\beta = 0$, which requires $\alpha \approx 5.74$. (iii) We add an uncertainty component without upscaling the model measurement uncertainty to obtain $\chi^2_{\tilde{m}} = 1$, i.e. $\alpha = 1$, which requires $\beta \approx 0.49$. In this last case, β is very close to the total model uncertainty. This shows that the model measurement uncertainty explains only a small part of the discrepancy between the filtered input and the model solution. For comparison, the mean measurement model uncertainty is 0.09 (without upscaling). This means that the theoretical model uncertainty that needs to be added to the measurement uncertainty for a correct total model uncertainty is $0.49/0.09 \approx 5.5$ times the model measurement uncertainty (without any upscaling). Therefore, in this case, the total model uncertainty is thus dominated by what we refer to as theoretical uncertainty. In other words, the uncertainty provided by the measurement algorithm explains only a small fraction of the total magnitude of the uncertainty.

6 Discussion

The SOLA-finite-frequency framework for surface-wave tomography we present in this study makes it possible to obtain 3D resolution and uncertainty estimates in surface-wave tomography. Here, we discuss our findings regarding resolution and uncertainty in more detail and discuss possible future directions.

6.1 Full 3D resolution

Our setup offers many advantages for estimating seismic model resolution: we obtain the full resolution matrix in a computationally efficient way, the resolution is fully 3D, it is unbiased by construction (local averaging weights sum to 1), while at the same time we control the resolution we obtain by choosing the target kernels. This is in contrast with most other studies that typically have assessed the resolution through either inverting synthetic input models (e.g. French et al., 2013), checkerboard test (e.g. Zhou et al., 2006; Auer et al., 2014; Rawlinson and Spakman, 2016),

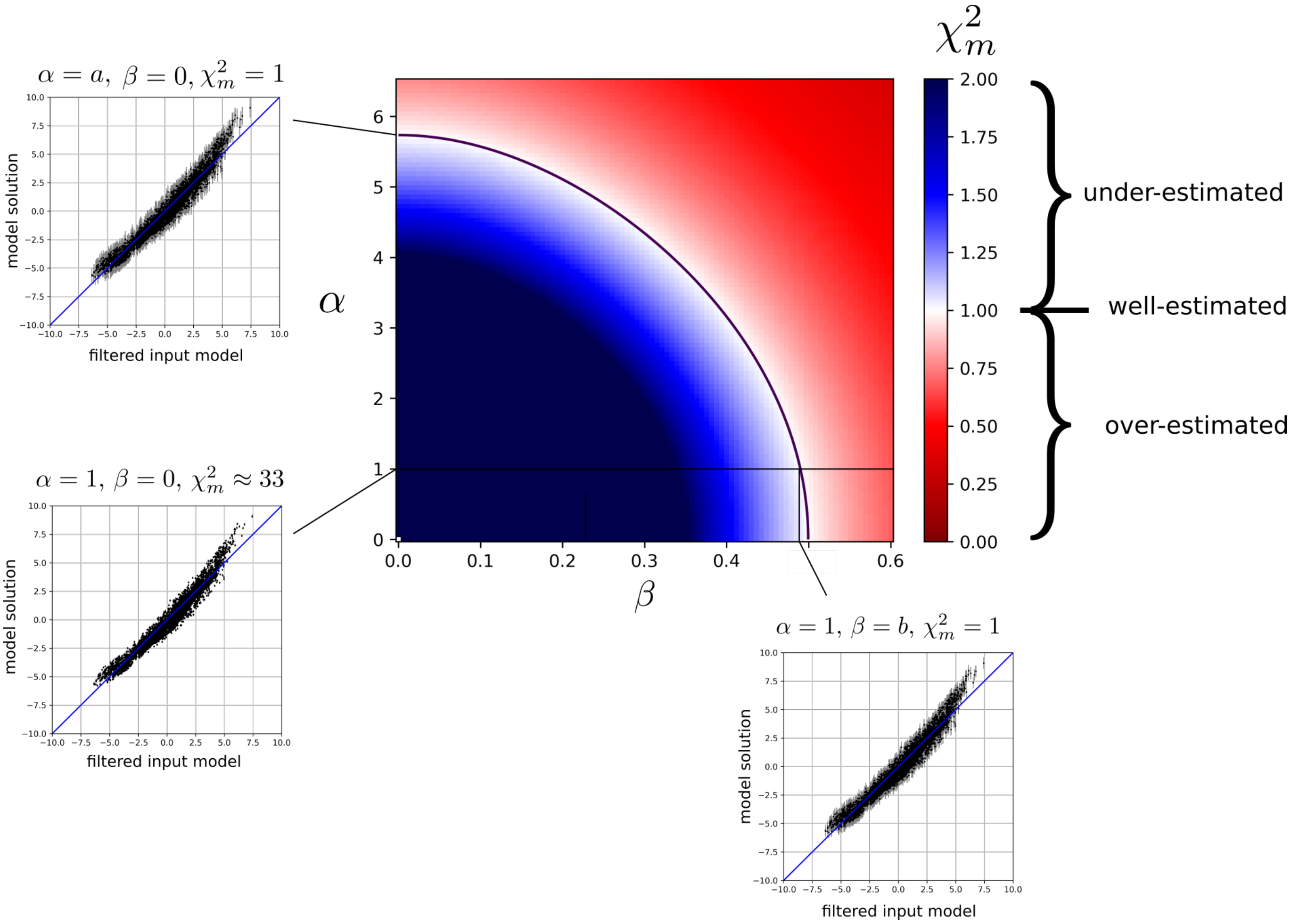


Figure 9 Model uncertainty analysis. The central plot shows the value of $\chi^2_{\widetilde{m}}$ (the misfit between the model solution and the filtered input model, normalised by the model uncertainty) for various combinations of the scaling factor α and added theoretical component β. In general, one should aim to find values of α and β that lead to $\chi^2_{\widetilde{m}} = 1$ (the black line in the white area). For small values of both α and β (blue region, or lower-left part of the plot), $\chi^2_{\widetilde{m}} > 1$, meaning that the model uncertainty is under-estimated, while the red regions indicate the model uncertainty is overestimated. The three cross-plots show the velocity variations in the model solution versus those in the filtered input model for three cases: (i) upscaled measurement uncertainty and no added component (upper-left), (ii) no upscaling nor added component (lower-left), and (iii) an added component, but no upscaling (lower-right). Note that only the error bars representing the total model uncertainty for various combinations of α and β change between these plots.

point spread functions (Ritsema et al., 2004; Bonadio et al., 2021), using the Hessian in the context of full-waveform inversion (e.g. Fichtner and Trampert, 2011), statistical methods using Monte Carlo approaches or transdimensional tomography (e.g. An, 2012; Bodin et al., 2012b; Sambridge et al., 2013), or other algebraic manipulations (e.g. Fichtner and Zunino, 2019; Shapiro et al., 2005; French and Romanowicz, 2014). While these approaches can handle non-linear inverse problems, they are typically computationally expensive, approximate, and only partially assess the resolution. In addition, since surface-wave tomography is often based on a two-step approach, estimates for the resolution are typically only 2D (lateral) or 1D (vertical). Moreover, data-fitting methods have great difficulties to provide direct control over the resolution, which can lead to biased local averages (e.g. Zaroli et al., 2017).

In this (synthetic) study, we find that the resolution is good enough laterally to qualitatively retrieve the main features of the input model (compare Figure 8a and b). This may be surprising given the small number of data in our inversion (47 700). We believe there are three main reasons for this: (i) we carefully select our input data, (ii) finite-frequency theory provides improved constraints compared to ray theory since one 3D sensitivity kernel constrains more model parameters than a thin ray, while it is also more accurate, and (iii) the SOLA inversion performs well in optimally utilising the data sensitivities. Point (ii) shares some similarities with adjoint methods used in full waveform inversion, given the volumetric nature of the adjoint sensitivity kernels (e.g. Monteiller et al., 2015).

The SOLA method consists of individual inversions for each model parameter without imposing any global constraint on all model parameters together (other than the target kernels). Therefore, the fact that we recover large-scale structure in the filtered model and model solution that are consistent with the input model is encouraging (Zaroli, 2016). However, compared to the input model, some short-scale variability arises in the filtered input, where adjacent cells show relatively strong differences. This is due to the pointwise nature of the SOLA inversion, combined with the absence of a smoothness criterion, and the smooth nature of the input model itself.

In the above, we typically assess the performance of the resolution by comparing the filtered model to the input model. In doing this, we must keep in mind that there is a dependency on the roughness of the input model itself. In particular, if the input model had shorter scale structure, we might not have been able to resolve it. While the resolution itself remains reliable, the comparison of input *versus* output models depends on the input itself; this bears some similarity with the inherent limitations of checkerboard tests (e.g. Lévêque et al., 1993; Rawlinson and Spakman, 2016). The full resolution itself remains necessary for robust model interpretations.

Since the data sensitivity and the resolution is fully 3D, we can confidently interpret the model resolution and uncertainty at all depths. This is a great advantage compared to our earlier 2D work (Latallerie et al., 2022), where the data sensitivity was imposed based on the lateral ray coverage (assuming ray theory). As a consequence, this study was likely too optimistic about the resolution at greater depth and therefore it was not possible to clearly state up to what depth the resolution and uncertainty estimates could be robustly interpreted. Moreover, since our resolution is fully 3D, we can investigate vertical resolution effects here. In addition to the well-known lateral smearing that arises in surface-wave tomography (discussed by Latallerie et al. (2022)), our averaging kernels indicate also significant vertical smearing (or depth leakage) in the cross-sections (Figures 3 and 4). Similar observations have been made in the context of full waveform inversion through assessment of the Hessian (e.g. Fichtner and Trampert, 2011). For some model parameters, the averages we recover relate primarily to structure above or below the true location as the averaging kernel is shifted upward or downward relative to the target kernel. In particular, the structure obtained at greater depth tends to be an average over shallower structure, with the effect becoming stronger with depth. Ignoring this full 3D resolution could thus lead to biased interpretations of surface-wave tomography, for example in studies of the age-depth trends of the oceanic lithosphere (e.g. Ritzwoller et al., 2004; Priestley and Mckenzie, 2006; Maggi et al., 2006b; Isse et al., 2019). This synthetic study thus emphasises the importance of taking vertical resolution into account when interpreting surface-wave tomography models and provides a quantitative way to estimate the depth to which a surface-wave tomography model should be interpreted.

Resolution and uncertainty are closely related; regions with high resolution tend to have high uncertainty, and *vice versa*. In this study, we find that the propagation of uncertainty decreases with depth (Fig. 5). This might be counter-intuitive as we expect the sensitivity of surface waves to decrease with depth. However, this observation has also been noted in other studies (e.g. Zhang et al., 2018; Earp et al., 2020; Latallerie et al., 2022). Our 3D resolution provides a robust explanation for the the decrease of uncertainty with depth. As depth increases, the resolution typically degrades (averages are estimated over larger volumes), leading to lower uncertainties. This illustrates that a combined analysis of uncertainty and 3D resolution is necessary to fully understand the limitations of surface-wave tomographic models.

6.2 Robust uncertainty estimates?

In this study, we estimate model uncertainty by propagating data uncertainty into model uncertainty using SOLA, which works for linear(ised) inverse problems. Other studies have used Bayesian approaches (e.g. Bodin et al., 2012b; Sambridge et al., 2013; Zhang et al., 2018), recently helped by machine learning approaches (e.g. Earp et al., 2020), where the posterior probability density function for the model can be interpreted as a measure of uncertainty. The Hessian has also been used in full waveform inversions (e.g. Fichtner and Trampert, 2011). However, in non-linear problems, the interpretation becomes more difficult. In both cases, we are left with the problem of estimating robust data uncertainty, which in the Bayesian philosophy entails finding the right prior probability distribution.

Since errors in the tomographic problem stem from both imperfect measurement and forward theory, we have separated the data uncertainty into two components: measurement and theoretical uncertainty. We have estimated the measurement uncertainty with repeated sampling, changing the time window using the multi-taper technique. This is similar to previous studies, which have used summary rays, bootstrapping or perturbation methods to estimate the data mean and measurement uncertainty (e.g. Maggi et al., 2006b; Koelemeijer et al., 2013; Amiri et al., 2023; Asplet et al., 2020). The general conclusion in such studies is that data uncertainty is typically underestimated. This is clear from the meta-analysis of published tomography models that show that the discrepancies are stronger than the typical error bars (e.g. Hosseini et al., 2018; Marignier et al., 2020; De Viron et al., 2021). This has led authors to use simple *ad hoc* criteria for upscaling the measurement uncertainty. For example, Latallerie et al. (2022) use a least-squares χ-test to upscale the uncertainty by a factor up to 3.4, while Lin et al. (2009) multiply their random error uncertainty estimates by 1.5 to obtain a more realistic model uncertainty estimate. While the measurement uncertainty might indeed be underestimated (which led us to define the factor α in section 5.2), the total uncertainty needs to account for additional theoretical uncertainty (the factor β in section 5.2).

Theoretical uncertainty has typically been estimated using synthetic tests during which input parameters are varied and the range of recovered data values is recorded as uncertainty. For example, for surface-wave dispersion measurements, Bozdağ and Trampert (2008) investigated the theoretical errors induced by imperfect crustal corrections, while Amiri et al. (2023) estimated the theoretical error induced by source mislocation. Similarly, Akbarashrafi et al. (2018) investigated the theoretical error produced by different coupling approximations on normal mode measurements, finding that reported data uncertainties need to be at least doubled to account for the errors due to theoretical omissions. In this work, we instead estimated the effect of the theoretical uncertainties on the model using a synthetic tomography setup that included many sources of theoretical uncertainty simultaneously. The effect of resolution was removed by filtering the input model so that discrepancies between our model estimate and the filtered input model represent the total uncertainty. Since we obtained the model measurement uncertainty resulting from the propagated data measurement uncertainty, we estimated the theoretical model uncertainty to be ~ 5.5 times larger than the model measurement uncertainty. The theoretical model uncertainty is thus larger than previously proposed factors of 1.5–3.4 (Lin et al., 2009; Latallerie et al., 2022), providing further evidence that the model uncertainty is indeed severely underestimated if we only propagate the data measurement uncertainty. Whether there is a need to upscale the measurement uncertainty naturally also depends on the specifics of the study and on the reliability of the measurement uncertainty estimate itself.

The main aim of this study is to provide a framework for surface-wave tomography with robust model statistics, including both the 3D resolution and total uncertainty. However, we still suffer from several drawbacks. For instance, although our measurement uncertainty should account for contamination by other phases or higher modes and cycle skipping, visual inspection indicates that this is not always the case (Figure 6). In the case of poor measurements (e.g. due to a missed cycle skip) with low uncertainty, we underestimate the measurement uncertainty and consequently overestimate the theoretical uncertainty. This is the rationale behind the factor α to upscale the measurement uncertainty in Section 5.2 and illustrates the difficulty of correctly estimating the measurement uncertainty. An interesting alternative approach was presented by Bodin et al. (2012a) who proposed to use a hierarchical transdimensional Bayesian approach where the data uncertainty is an output of the inverse process itself, rather than an input. However, this approach assumes a single uncertainty value for all data, which can be problematic since the relative magnitude of the data uncertainty is of interest in the inverse process itself as well as for obtaining the robust model uncertainty.

Another drawback of our approach is that our estimates of theoretical uncertainty depend on the input model used, i.e. S362ANI (Kustowski et al., 2008). The validity of the forward theory depends on several assumptions (e.g. forward scattering, paraxial approximation) whose validity depends on the properties of the medium in which waves propagate (e.g. Liu and Zhou, 2013; Parisi et al., 2015). It is therefore important to perform our analysis in an Earth-like model and further work could investigate the dependency on the input model.

Additionally, the scaling factor α (upscaling of the measurement uncertainty) and the added component β (representing the theoretical uncertainty) need to be determined for a sufficiently large number of model parameters for the results to be statistically significant (here we considered all model parameters at 112 km depth). In particular, we would recommend to determine these parameters for each depth in the model independently, as velocity structure and the magnitudes of measurement and theoretical uncertainties likely change with depth.

We further assume the data uncertainties to be uncorrelated, whereas in reality we expect them to be correlated to some extent – e.g an error in the source location or mechanism will impact several measurements. In theory, it is possible to account for correlations between data uncertainties, but estimating these correlations remains a challenge in surface-wave tomography. Furthermore, our assumption of a zero-mean Gaussian distribution for the data errors seems reasonable, but the use of more general probability distributions could be also investigated (e.g. Tarantola, 2005).

Lastly, we estimate the theoretical uncertainty from the discrepancy between the filtered input model and the model solution based on measurements on SEM seismograms. Since the crustal model we assume for the crustal corrections is exactly the same as in the input model, and the source parameters used for generating the reference seismograms are exactly the same as for the SEM seismograms, there is no theoretical error associated with errors in the crustal model or source solution in our synthetic framework. Nevertheless, these two components likely introduce non-negligible errors in reality (e.g. Marone and Romanowicz, 2007; Bozdağ and Trampert, 2008; Panning et al., 2010; Ferreira et al., 2010; Liu and Zhou, 2013; Latallerie, 2022; Amiri et al., 2023). Model uncertainty related to these components could be incorporated in the theoretical uncertainty estimate proposed in this study.

Despite the drawbacks outlined above, we believe that our study provides a valuable starting point to obtain 3D resolution and estimate theoretical model uncertainty in surface-wave tomography, upon which future work can build. This information is vital for robust model interpretations and to reconcile existing discrepancies between published tomography models (e.g. Hosseini et al., 2018; Marignier et al., 2020; De Viron et al., 2021).

6.3 Future directions

The depth sensitivity and thus resolution in this study is limited by the restriction to fundamental surface-wave data. This can be mitigated by adding measurements for surface-wave overtones. In theory, including these in the presented framework is trivial, but it will be important to carefully estimate the data uncertainty for these new measurements. The resolution and uncertainty produced in our setup can be used to inform other tomographic studies. Our 3D resolution maps indicate how well certain model parameters are constrained depending on their position and particularly with depth. Based on this, we may choose sets of source-receiver paths and frequencies that best suit a certain target. For example, to better homogenise the resolution with depth, we may want to increase the number and/or the relative weight of low frequency data.

The obvious next step is to apply the approach presented here to real data, using the lessons learned in this synthetic study. The information on 3D resolution and uncertainty obtained using SOLA would be particularly useful for testing geodynamic predictions (Freissler et al., 2022). In addition, this information would ensure that we only interpret the tomographic models to their limits, and not beyond, being aware of potential resolution artefacts, especially with depth.

There are many other directions for further development. For example, it is possible to extend the SOLA-finite-frequency framework for surface-wave tomography to other data and physical parameters, e.g. amplitude measurements to study anelasticity in the upper-mantle (e.g. Zhou, 2009b). These could be investigated independently, or through a joint approach, thus reducing theoretical uncertainty due to neglecting the effect of other physical parameters.

Conclusion

In this contribution, we have combined the Backus-Gilbert-based SOLA inverse method with finite-frequency theory in a synthetic study of the Pacific upper mantle. Our 3D modelling and inversion framework enables us to control and produce uncertainty and resolution information together with the surface-wave tomography model. We have used a synthetic framework to demonstrate the reliability of our approach and to investigate the effect of 3D resolution, laterally and vertically, in surface-wave tomography. We find that the limited resolution induces well-known artefacts, including lateral smearing effects where data coverage is poor or highly anisotropic. More importantly, we show that limited vertical resolution can induce strong artefacts with model parameters potentially representing averages of true Earth properties at much shallower depth. Knowledge of this full 3D resolution is crucial for robust interpretations of surface-wave tomography models. Our synthetic setup allows us to also explore the reliability of model uncertainty estimates. We find that the theoretical uncertainty, required to match the filtered input model, might be much larger than the measurement uncertainty in the data. This demonstrates the need to account for both measurement and theoretical uncertainty in surface-wave tomography. We believe that our study is a starting point towards better use and interpretation of surface-wave tomography models.

Acknowledgments

This research is supported by the UKRI NERC Large grant “Mantle Circulation Constrained (MC2): A multidisciplinary 4D Earth framework for understanding mantle upwellings” (NE/T01248X/1). PK acknowledges financial support from a Royal Society University Research Fellowship (URF \1 \180377). CZ acknowledges financial support from ITES (Institut Terre et Environnement de Strasbourg, UMR 7063) for a research visit to Oxford. This study used the ARCHER2 UK National Supercomputing Service (https://www.archer2.ac.uk). For this study we made extensive use of GNU/Linux and Python (including packages Scipy, Numpy, Matplotlib, Pandas and Multiprocessing). For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising.

Data and code availability

Seismic source solutions were downloaded from the Global Centroid Moment Tensor (GCMT) Catalog (Dziewonski et al., 1981; Ekström et al., 2012). The facilities of the EarthScope Consortium were used to access waveforms and related metadata and derived data products. These services are funded through the National Science Foundation's Seismological Facility for the Advancement of Geoscience (SAGE) Award under Cooperative Agreement EAR-1724509. All waveforms used in this study are SEM synthetics from the GlobalShakeMovie project (Tromp et al., 2010), and were obtained through IRIS DMC (Hutko et al., 2017; IRIS DMC, 2012). To compute the finite-frequency sensitivity kernels, we used software provided by Ying Zhou (Zhou, 2009b), available via their webpage. To compute the reference seismograms in a 1D radial Earth model using normal modes summation, we used MINEOS 1.0.2 (Masters et al., 2011) published under the GPL2 license. We thank the Computational Infrastructure for Geodynamics (http://geodynamics.org), which is funded by the National Science Foundation under awards EAR-0949446, EAR-1550901, and EAR-2149126 for making the code available.

The SOLA tomography code used in this study consists of running the LSQR algorithm of Paige and Saunders

(1982) with specific input matrices and vectors. These inputs can be constructed from the sensitivity matrix and target kernels as detailed in Appendix A1 of Zaroli (2016). The LSQR code is freely downloadable from the webpage of the Systems Optimisation Laboratory (Stanford University): https://web.stanford.edu/group/SOL/software/lsqr/. A pre-constructed software package for SOLA tomography is available from Christophe Zaroli (c.zaroli@unistra.fr) upon e-mail request.

References

Akbarashrafi, F., Al-Attar, D., Deuss, A., Trampert, J., and Valentine, A. P. Exact free oscillation spectra, splitting functions and the resolvability of Earth's density structure. *Geophysical Journal International*, 213(1):58–76, Apr. 2018. doi: 10.1093/gji/ggx539.

Amiri, S., Maggi, A., Tatar, M., Zigone, D., and Zaroli, C. Rayleigh wave group velocities in North-West Iran: SOLA Backus-Gilbert vs. Fast Marching tomographic methods. *Seismica*, 2(2), Dec. 2023. doi: 10.26443/seismica.v2i2.1011.

An, M. A simple method for determining the spatial resolution of a general inverse problem. *Geophys. J. Int.*, 191(2):849–864, Nov. 2012. doi: 10.1111/j.1365-246X.2012.05661.x.

Asplet, J., Wookey, J., and Kendall, M. A potential post-perovskite province in D" beneath the Eastern Pacific: evidence from new analysis of discrepant SKS–SKKS shear-wave splitting. *Geophys. J. Int.*, 221(3):2075–2090, June 2020. doi: 10.1093/gji/ggaa114.

Auer, L., Boschi, L., Becker, T. W., Nissen-Meyer, T., and Giardini, D. *Savani* : A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets. *J. Geophys. Res. Solid Earth*, 119(4):3006–3034, Apr. 2014. doi: 10.1002/2013JB010773.

Backus, G. and Gilbert, F. The Resolving Power of Gross Earth Data. *Geophys. J. Int.*, 16(2):169–205, Oct. 1968. doi: 10.1111/j.1365-246X.1968.tb00216.x.

Backus, G. E. and Gilbert, F. Uniqueness in the inversion of inaccurate gross Earth data. *Phil. Trans. R. Soc. A.*, 266(1173):74, 1970. doi: https://doi.org/10.1098/rsta.1970.0005.

Backus, G. E. and Gilbert, J. F. Numerical Applications of a Formalism for Geophysical Inverse Problems. *Geophys. J. Int.*, 13(1-3):247–276, July 1967. doi: 10.1111/j.1365-246X.1967.tb02159.x.

Barmin, M. P., Ritzwoller, M. H., and Levshin, A. L. A Fast and Reliable Method for Surface Wave Tomography. *Pure appl. geophys.*, 158:25, 2001.

Bassin, C., Laske, G., and Masters, G. The Current Limits of Resolution for Surface Wave Tomography in North America. *Eos Trans. AGU*, 81 (F897), 2000.

Boaga, J., Vignoli, G., and Cassiani, G. Shear wave profiles from surface wave inversion: the impact of uncertainty on seismic site response analysis. *J. Geophys. Eng.*, 8(2):162–174, June 2011. doi: 10.1088/1742-2132/8/2/004.

Boaga, J., Vignoli, G., and Cassiani, G. Reply to comment on 'Shear wave profile from surface wave inversion: the impact of uncertainty on seismic site response analysis'. *J. Geophys. Eng.*, 9(2):244–246, Apr. 2012. doi: 10.1088/1742-2132/9/2/244.

Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P. Transdimensional tomography with unknown data noise. *Geophys. J. Int.*, 189(3): 1536–1556, June 2012a. doi: 10.1111/j.1365-246X.2012.05414.x.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., and Rawlinson, N. Transdimensional inversion of receiver functions and surface wave dispersion: TRANSDIMENSIONAL INVERSION OF RF AND SWD. *J. Geophys. Res.*, 117(B2):n/a–n/a, Feb. 2012b. doi: 10.1029/2011JB008560.

Bonadio, R., Lebedev, S., Meier, T., Arroucau, P., Schaeffer, A. J., Licciardi, A., Agius, M. R., Horan, C., Collins, L., O'Reilly, B. M., Readman,

P. W., and Ireland Array Working Group. Optimal resolution tomography with error tracking and the structure of the crust and upper mantle beneath Ireland and Britain. *Geophys. J. Int.*, 226(3):2158–2188, June 2021. doi: 10.1093/gji/ggab169.

Bozdağ, E. and Trampert, J. On crustal corrections in surface wave tomography. *Geophys. J. Int.*, 172(3):1066–1082, Mar. 2008. doi: 10.1111/j.1365-246X.2007.03690.x.

Chen, L.-W. and Romanowicz, B. On accounting for the effects of crust and uppermost mantle structure in global scale full-waveform inversion. *Geophys. J. Int.*, 2024.

Dahlen, F. A. and Tromp, J. *Theoretical Global Seismology*. Princeton University Press, Dec. 1999. doi: 10.1515/9780691216157.

De Viron, O., Van Camp, M., Grabkowiak, A., and Ferreira, A. M. G. Comparing global seismic tomography models using varimax principal component analysis. *Solid Earth*, 12(7):1601–1634, July 2021. doi: 10.5194/se-12-1601-2021.

Debayle, E. SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion. *Geophys. J. Int.*, 137(3):747–754, June 1999. doi: 10.1046/j.1365-246x.1999.00832.x.

Debayle, E. and Kennett, B. L. N. The Australian continental upper mantle: Structure and deformation inferred from surface waves. *J. Geophys. Res.*, 105(B11):25423–25450, Nov. 2000. doi: 10.1029/2000JB900212.

Debayle, E. and Lévêque, J. J. Upper mantle heterogeneities in the Indian Ocean from waveform inversion. *Geophys. Res. Lett.*, 24(3): 245–248, Feb. 1997. doi: 10.1029/96GL03954.

Debayle, E. and Sambridge, M. Inversion of massive surface wave data sets: Model construction and resolution assessment. *J. Geophys. Res.*, 109(B2), Feb. 2004. doi: 10.1029/2003JB002652.

Dziewonski, A. M., Chou, T.-A., and Woodhouse, J. H. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J. Geophys. Res.*, 86(B4):2825–2852, Apr. 1981. doi: 10.1029/JB086iB04p02825.

Earp, S., Curtis, A., Zhang, X., and Hansteen, F. Probabilistic neural network tomography across Grane field (North Sea) from surface wave dispersion data. *Geophys. J. Int.*, 223(3):1741–1757, Oct. 2020. doi: 10.1093/gji/ggaa328.

Ekström, G. A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25-250 s: Global dispersion model GDM52. *Geophys. J. Int.*, 187(3):1668–1686, Dec. 2011. doi: 10.1111/j.1365-246X.2011.05225.x.

Ekström, G., Tromp, J., and Larson, E. W. F. Measurements and global models of surface wave propagation. *J. Geophys. Res.*, 102(B4): 8137–8157, Apr. 1997. doi: 10.1029/96JB03729.

Ekström, G., Nettles, M., and Dziewoński, A. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Phys. Earth planet. Inter.*, 200-201:1–9, June 2012. doi: 10.1016/j.pepi.2012.04.002.

Ferreira, A. M. G., Woodhouse, J. H., Visser, K., and Trampert, J. On the robustness of global radially anisotropic surface wave tomography. *J. Geophys. Res.*, 115(B4):2009JB006716, Apr. 2010. doi: 10.1029/2009JB006716.

Fichtner, A. and Trampert, J. Resolution analysis in full waveform inversion: Resolution in full waveform inversion. *Geophysical Journal International*, 187(3):1604–1624, Dec. 2011. doi: 10.1111/j.1365-246X.2011.05218.x.

Fichtner, A. and Zunino, A. Hamiltonian Nullspace Shuttles. *Geophys. Res. Lett.*, 46(2):644–651, Jan. 2019. doi: 10.1029/2018GL080931.

Foulger, G. R., Panza, G. F., Artemieva, I. M., Bastow, I. D., Cammarano, F., Evans, J. R., Hamilton, W. B., Julian, B. R., Lustrino, M., Thybo, H., and Yanovskaya, T. B. Caveats on tomographic images. *Terra Nova*, 25(4):259–281, Aug. 2013. doi: 10.1111/ter.12041.

Freissler, R., Zaroli, C., Lambotte, S., and Schuberth, B. S. A. Tomographic filtering via the generalized inverse: a way to account for seismic data uncertainty. *Geophys. J. Int.*, 223(1):254–269, Oct. 2020. doi: 10.1093/gji/ggaa231.

Freissler, R., Schuberth, B. S., and Zaroli, C. The relevance of full 3D-wavefield simulations for the tomographic filtering of geodynamic models, Mar. 2022. https://meetingorganizer.copernicus.org/EGU22/EGU22-11686.html.

Freissler, R., Schuberth, B. S. A., and Zaroli, C. A concept for the global assessment of tomographic resolution and uncertainty. *Geophys. J. Int.*, page ggae178, May 2024. doi: 10.1093/gji/ggae178.

French, S., Lekic, V., and Romanowicz, B. Waveform Tomography Reveals Channeled Flow at the Base of the Oceanic Asthenosphere. *Science*, 342(6155):227–230, Oct. 2013. doi: 10.1126/science.1241514.

French, S. W. and Romanowicz, B. A. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography. *Geophys. J. Int.*, 199(3):1303–1327, Dec. 2014. doi: 10.1093/gji/ggu334.

Greenfield, T., Gilligan, T., Pilia, S., Cornwell, D. G., Tongkul, F., Widiyantoro, S., and Rawlinson, N. Post-Subduction Tectonics of Sabah, Northern Borneo, Inferred From Surface Wave Tomography. *Geophys. Res. Lett.*, 49(3), Feb. 2022. doi: 10.1029/2021GL096117.

Hjörleifsdóttir, V. *Earthquake Source Characterization Using 3D Numerical Modeling*. PhD thesis, California Institute of Technology, Apr. 2007. https://resolver.caltech.edu/CaltechETD:etd-03212007-170259. Medium: PDF Version Number: Final.

Hosseini, K., Matthews, K. J., Sigloch, K., Shephard, G. E., Domeier, M., and Tsekhmistrenko, M. SubMachine: Web-Based Tools for Exploring Seismic Tomography and Other Models of Earth's Deep Interior. *Geochem. Geophys. Geosyst.*, 19(5):1464–1483, May 2018. doi: 10.1029/2018GC007431.

Hutko, A. R., Bahavar, M., Trabant, C., Weekly, R. T., Fossen, M. V., and Ahern, T. Data Products at the IRIS-DMC: Growth and Usage. *Seismol. Res. Lett.*, 88(3):892–903, May 2017. doi: 10.1785/0220160190.

IRIS DMC. Data Services Products: ShakeMovieSynthetics 1D & 3D synthetic seismograms from the Global ShakeMovie project, 2012. https://doi.org/10.17611/DP/SYN.1.

Isse, T., Kawakatsu, H., Yoshizawa, K., Takeo, A., Shiobara, H., Sugioka, H., Ito, A., Suetsugu, D., and Reymond, D. Surface wave tomography for the Pacific Ocean incorporating seafloor seismic observations and plate thermal evolution. *Earth and Planet. Sc. Lett.*, 510:116–130, Mar. 2019. doi: 10.1016/j.epsl.2018.12.033.

Koelemeijer, P., Deuss, A., and Ritsema, J. Observations of core-mantle boundary Stoneley modes. *Geophysical Research Letters*, 40(11): 2557–2561, June 2013. doi: 10.1002/grl.50514.

Komatitsch, D. and Tromp, J. Spectral-element simulations of global seismic wave propagation-II. Three-dimensional models, oceans, rotation and self-gravitation. *Geophys. J. Int.*, 150(1):303–318, July 2002. doi: 10.1046/j.1365-246X.2002.01716.x.

Komatitsch, D. and Vilotte, J.-P. The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures. *B. Seismol. Soc. Am.*, 88(2):368–392, Apr. 1998. doi: 10.1785/BSSA0880020368.

Kustowski, B., Ekström, G., and Dziewoński, A. M. Anisotropic shear-wave velocity structure of the Earth's mantle: A global model. *J. Geophys. Res.*, 113(B6):B06306, June 2008. doi: 10.1029/2007JB005169.

Laske, G. and Masters, G. Constraints on global phase velocity maps from long-period polarization data. *J. Geophys. Res.*, 101(B7):16059–16075, July 1996. doi: 10.1029/96JB00526.

Laske, G., Masters, G., and Zu¨rn, W. Frequency-dependent polarization measurements of long-period surface waves and their implications for global phase-velocity maps. *Phys. Earth planet. Inter.*, 84(1-4):111–137, July 1994. doi: 10.1016/0031-9201(94)90037-X.

Latallerie, F. *Seismic tomography of plume-like upwellings in the French Polynesia region using Backus-Gilbert inversion*. Theses, Université de Strasbourg, Dec. 2022. https://theses.hal.science/tel-04193813. Issue: 2022STRAH012.

Latallerie, F., Zaroli, C., Lambotte, S., and Maggi, A. Analysis of tomographic models using resolution and uncertainties: a surface wave example from the Pacific. *Geophys. J. Int.*, 230(2):893–907, Apr. 2022. doi: 10.1093/gji/ggac095.

Lebedev, S. and Nolet, G. Upper mantle beneath Southeast Asia from *S* velocity tomography: UPPER MANTLE BENEATH SE ASIA. *J. Geophys.*

Res., 108(B1), Jan. 2003. doi: 10.1029/2000JB000073.

Lin, F.-C., Ritzwoller, M. H., and Snieder, R. Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophys. J. Int.*, 177(3):1091–1110, June 2009. doi: 10.1111/j.1365-246X.2009.04105.x.

Liu, K. and Zhou, Y. Effects of crustal thickness variations on surface wave phase delays. *Geophys. J. Int.*, 192(2):773–792, Feb. 2013. doi: 10.1093/gji/ggs048.

Liu, K. and Zhou, Y. Global Rayleigh wave phase-velocity maps from finite-frequency tomography. *Geophys. J. Int.*, 205(1):51–66, Apr. 2016a. doi: 10.1093/gji/ggv555.

Liu, K. and Zhou, Y. Travelling-wave Green tensor and near-field Rayleigh-wave sensitivity. *Geophys. J. Int.*, 205(1):134–145, Apr. 2016b. doi: 10.1093/gji/ggv564.

Lévêque, J., Debayle, E., and Maupin, V. Anisotropy in the Indian Ocean upper mantle from Rayleigh- and Love-waveform inversion. *Geophys. J. Int.*, 133(3):529–540, June 1998. doi: 10.1046/j.1365-246X.1998.00504.x.

Lévêque, J. J., Rivera, L., and Wittlinger, G. On the use of the checker-board test to assess the resolution of tomographic inversions. *Geophys. J. Int.*, 115(1):313–318, Oct. 1993. doi: 10.1111/j.1365-246X.1993.tb05605.x.

Ma, Z., Masters, G., Laske, G., and Pasyanos, M. A comprehensive dispersion model of surface wave phase and group velocity for the globe. *Geophys. J. Int.*, 199(1):113–135, Oct. 2014. doi: 10.1093/gji/ggu246.

Maggi, A., Debayle, E., Priestley, K., and Barruol, G. Azimuthal anisotropy of the Pacific region. *Earth and Planet. Sc. Lett.*, 250(1-2):53–71, Oct. 2006a. doi: 10.1016/j.epsl.2006.07.010.

Maggi, A., Debayle, E., Priestley, K., and Barruol, G. Multimode surface waveform tomography of the Pacific Ocean: a closer look at the lithospheric cooling signature. *Geophys. J. Int.*, 166(3):1384–1397, Sept. 2006b. doi: 10.1111/j.1365-246X.2006.03037.x.

Magrini, F., Diaferia, G., El-Sharkawy, A., Cammarano, F., van der Meijde, M., Meier, T., and Boschi, L. Surface-Wave Tomography of the Central-Western Mediterranean: New Insights Into the Liguro-Provençal and Tyrrhenian Basins. *JGR Solid Earth*, 127(3), Mar. 2022. doi: 10.1029/2021JB023267.

Marignier, A., Ferreira, A. M. G., and Kitching, T. The Probability of Mantle Plumes in Global Tomographic Models. *Geochem. Geophys. Geosyst.*, 21(9), Sept. 2020. doi: 10.1029/2020GC009276.

Marone, F. and Romanowicz, B. Non-linear crustal corrections in high-resolution regional waveform seismic tomography. *Geophys. J. Int.*, 170:460–467, Feb. 2007. doi: 10.1111/j.1365-246X.2007.03399.x.

Marquering, H., Nolet, G., and Dahlen, F. Three-dimensional waveform sensitivity kernels. *Geophys. J. Int.*, 132(3):521–534, Mar. 1998. doi: 10.1046/j.1365-246X.1998.00426.x.

Masters, G., Woodhouse, J. H., and Freeman, G. Mineos v1.0.2, 2011. https://geodynamics.org/cig,2011.

Menke, W. *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, 1989.

Montagner, J.-P. Upper mantle low anisotropy channels below the Pacific Plate. *Earth and Planet. Sc. Lett.*, 202(2):263–274, Sept. 2002. doi: 10.1016/S0012-821X(02)00791-4.

Monteiller, V., Chevrot, S., Komatitsch, D., and Wang, Y. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method. *Geophys. J. Int.*, 202(2):811–827, Aug. 2015. doi: 10.1093/gji/ggv189.

Moulik, P., Lekic, V., Romanowicz, B., Ma, Z., Schaeffer, A., Ho, T., Beucler, E., Debayle, E., Deuss, A., Durand, S., Ekström, G., Lebedev, S., Masters, G., Priestley, K., Ritsema, J., Sigloch, K., Trampert, J., and Dziewonski, A. M. Global reference seismological data sets: multimode surface wave dispersion. *Geophys. J. Int.*, 228(3):1808–1849, Dec. 2021. doi: 10.1093/gji/ggab418.

Nolet, G. Solving or resolving inadequate and noisy tomographic systems. *J. Comput. Phys.*, 61(3):463–482, Dec. 1985. doi: 10.1016/0021-9991(85)90075-0.

Nolet, G. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*. Cambridge University Press, 1 edition, Sept. 2008. doi: 10.1017/CBO9780511984709.

Nolet, G., Montelli, R., and Virieux, J. Explicit, approximate expressions for the resolution and *a posteriori* covariance of massive tomographic systems. *Geophys. J. Int.*, 138(1):36–44, July 1999. doi: 10.1046/j.1365-246x.1999.00858.x.

Ouattara, Y., Zigone, D., and Maggi, A. Rayleigh wave group velocity dispersion tomography of West Africa using regional earthquakes and ambient seismic noise. *J Seismol*, 23(6):1201–1221, Nov. 2019. doi: 10.1007/s10950-019-09860-z.

Paige, C. C. and Saunders, M. A. LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. Math. Softw.*, 8(1):43–71, Mar. 1982. doi: 10.1145/355984.355989.

Panning, M. P., Lekić, V., and Romanowicz, B. A. Importance of crustal corrections in the development of a new global model of radial anisotropy. *J. Geophys. Res.*, 115(B12):B12325, Dec. 2010. doi: 10.1029/2010JB007520.

Parisi, L., Ferreira, A. M. G., and Capdeville, Y. Validity domain of the Born approximation for seismic waveform modelling in realistic 3-D Earth structure. *Geophysical Journal International*, 200(2):910–916, Feb. 2015. doi: 10.1093/gji/ggu446.

Park, J., Lindberg, C. R., and Vernon, F. L. Multitaper spectral analysis of high-frequency seismograms. *J. Geophys. Res.*, 92(B12):12675, 1987a. doi: 10.1029/JB092iB12p12675.

Park, J., Vernon, F. L., and Lindberg, C. R. Frequency dependent polarization analysis of high-frequency seismograms. *J. Geophys. Res.*, 92 (B12):12664, 1987b. doi: 10.1029/JB092iB12p12664.

Parker, R. L. Understanding Inverse Theory. *Annu. Rev. Earth Planet. Sci.*, 5(1):35–64, May 1977. doi: 10.1146/annurev.ea.05.050177.000343.

Percival, D. B. and Walden, A. T. *Spectral Analysis for Physical Applications*. Cambridge University Press, 1 edition, June 1993. doi: 10.1017/CBO9780511622762.

Pijpers, F. P. and Thompson, M. J. Faster formulations of the optimally localized averages method for helioseismic inversions. *Astron. Astrophys.*, 262:L33–L36, Sept. 1992.

Pijpers, F. P. and Thompson, M. J. The SOLA method for helioseismic inversion. *Astron. Astrophys.*, 281:231–240, Jan. 1994.

Priestley, K. Seismic evidence for a moderately thick lithosphere beneath the Siberian Platform. *Geophys. Res. Lett.*, 30(3):1118, 2003. doi: 10.1029/2002GL015931.

Priestley, K. and Mckenzie, D. The thermal structure of the lithosphere from shear wave velocities. *Earth and Planet. Sc. Lett.*, 244(1-2):285–301, Apr. 2006. doi: 10.1016/j.epsl.2006.01.008.

Rawlinson, N. and Spakman, W. On the use of sensitivity tests in seismic tomography. *Geophys. J. Int.*, 205(2):1221–1243, May 2016. doi: 10.1093/gji/ggw084.

Rawlinson, N., Fichtner, A., Sambridge, M., and Young, M. Seismic Tomography and the Assessment of Uncertainty. *Adv. Geophys.*, 55:1–76, 2014. doi: https://doi.org/10.1016/bs.agph.2014.08.001.

Restelli, F., Zaroli, C., and Koelemeijer, P. Robust estimates of the ratio between S- and P-wave velocity anomalies in the Earth's mantle using normal modes. *Phys. Earth planet. Inter.*, 347:107135, Feb. 2024. doi: 10.1016/j.pepi.2023.107135.

Ritsema, J., van Heijst, H. J., and Woodhouse, J. H. Global transition zone tomography: GLOBAL TRANSITION ZONE TOMOGRAPHY. *J. Geophys. Res.*, 109(B2), Feb. 2004. doi: 10.1029/2003JB002610.

Ritsema, J., McNamara, A. K., and Bull, A. L. Tomographic filtering of geodynamic models: Implications for model interpretation and large-scale mantle structure. *J. Geophys. Res.*, 112(B01303), Jan. 2007. doi: 10.1029/2006JB004566.

Ritzwoller, M. H., Shapiro, N. M., and Zhong, S.-J. Cooling history of the Pacific lithosphere. *Earth and Planet. Sc. Lett.*, 226(1-2):69–84, Sept. 2004. doi: 10.1016/j.epsl.2004.07.032.

Ruan, Y. and Zhou, Y. The effects of 3-D anelasticity (Q) structure on surface wave phase delays. *Geophys. J. Int.*, 181(1):479–492, Apr. 2010. doi: 10.1111/j.1365-246X.2010.04514.x.

Sambridge, M., Bodin, T., Gallagher, K., and Tkalčić, H. Transdimensional inference in the geosciences. *Phil. Trans. R. Soc. A.*, 371(1984): 20110547, Feb. 2013. doi: 10.1098/rsta.2011.0547.

Scales, J. A. and Snieder, R. To Bayes or not to Bayes? *GEOPHYSICS*, 62(4):1045–1046, July 1997. doi: 10.1190/1.6241045.1.

Schuberth, B. S. A., Bunge, H., and Ritsema, J. Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone? *Geochem Geophys Geosyst*, 10(5):2009GC002401, May 2009. doi: 10.1029/2009GC002401.

Seredkina, A. I. Surface Wave Tomography of the Arctic from Rayleigh and Love Wave Group Velocity Dispersion Data. *Izv., Phys. Solid Earth*, 55(3):439–450, May 2019. doi: 10.1134/S106935131903008X.

Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M. H. High-Resolution Surface-Wave Tomography from Ambient Seismic Noise. *Science*, 307(5715):1615–1618, Mar. 2005. doi: 10.1126/science.1108339.

Simmons, N. A., Schuberth, B. S. A., Myers, S. C., and Knapp, D. R. Resolution and Covariance of the LLNL-G3D-JPS Global Seismic Tomography Model: Applications to Travel time Uncertainty and Tomographic Filtering of Geodynamic Models. *Geophys. J. Int.*, 217(3): 1543–1557, June 2019. doi: 10.1093/gji/ggz102.

Simons, F. J., van der Hilst, R. D., Montagner, J.-P., and Zielhuis, A. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle. *Geophys. J. Int.*, 151(3):738–754, Dec. 2002. doi: 10.1046/j.1365-246X.2002.01787.x.

Slepian, D. Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-V: The Discrete Case. *Bell. Syst. Tech. J.*, 57(5):1371–1430, May 1978. doi: 10.1002/j.1538-7305.1978.tb02104.x.

Snieder, R. 3-D linearized scattering of surface waves and a formalism for surface wave holography. *Geophys. J. Int.*, 84(3):581–605, Mar. 1986. doi: 10.1111/j.1365-246X.1986.tb04372.x.

Snieder, R. and Nolet, G. Linearized scattering of surface waves on a spherical Earth. *Journal of Geophysics*, 61(1):55–63, 1987.

Socco, L. V., Foti, S., Comina, C., and Boiero, D. Comment on 'Shear wave profiles from surface wave inversion: the impact of uncertainty on seismic site response analysis'. *J. Geophys. Eng.*, 9(2):241–243, Apr. 2012. doi: 10.1088/1742-2132/9/2/241.

Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Jan. 2005. doi: 10.1137/1.9780898717921.

Tarantola, A. and Valette, B. Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys.*, 20(2):219, 1982. doi: 10.1029/RG020i002p00219.

Thomson, D. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70(9):1055–1096, 1982. doi: 10.1109/PROC.1982.12433.

Tian, Y., Zhou, Y., Sigloch, K., Nolet, G., and Laske, G. Structure of North American mantle constrained by simultaneous inversion of multiple-frequency *SH* , *SS* , and Love waves. *J. Geophys. Res.*, 116(B2):B02307, Feb. 2011. doi: 10.1029/2010JB007704.

Trampert, J. Global seismic tomography: the inverse problem and beyond. *Inverse Probl.*, 14(3):371–385, June 1998. doi: 10.1088/0266-5611/14/3/002.

Tromp, J., Komatitsch, D., Hjörleifsdóttir, V., Liu, Q., Zhu, H., Peter, D., Bozdag, E., McRitchie, D., Friberg, P., Trabant, C., and Hutko, A. Near real-time simulations of global CMT earthquakes: Near real-time simulations of CMT earthquakes. *Geophysical Journal International*, 183(1):381–389, Oct. 2010. doi: 10.1111/j.1365-246X.2010.04734.x.

Wiggins, R. A. The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure. *Rev. Geophys.*, 10 (1):251, 1972. doi: 10.1029/RG010i001p00251.

Yoshizawa, K. and Kennett, B. L. N. Multimode surface wave tomography for the Australian region using a three-stage approach incorporating finite frequency effects: THREE-STAGE SURFACE WAVE TOMOGRAPHY. *J. Geophys. Res.*, 109(B2), Feb. 2004. doi: 10.1029/2002JB002254.

Yoshizawa, K. and Kennett, B. L. N. Sensitivity kernels for finite-frequency surface waves. *Geophys. J. Int.*, 162(3):910–926, Sept. 2005. doi: 10.1111/j.1365-246X.2005.02707.x.

Zaroli, C. Global seismic tomography using Backus-Gilbert inversion. *Geophys. J. Int.*, 207(2):876–888, Nov. 2016. doi: 10.1093/gji/ggw315.

Zaroli, C. Seismic tomography using parameter-free Backus-Gilbert inversion. *Geophys. J. Int.*, 218(1):619–630, July 2019. doi: 10.1093/gji/ggz175.

Zaroli, C., Koelemeijer, P., and Lambotte, S. Toward Seeing the Earth's Interior Through Unbiased Tomographic Lenses. *Geophys. Res. Lett.*, 44(22):11,399–11,408, Nov. 2017. doi: 10.1002/2017GL074996.

Zhang, X., Curtis, A., Galetti, E., and de Ridder, S. 3-D Monte Carlo surface wave tomography. *Geophys. J. Int.*, 215(3):1644–1658, Dec. 2018. doi: 10.1093/gji/ggy362.

Zhou, Y. Multimode surface wave sensitivity kernels in radially anisotropic earth media. *Geophys. J. Int.*, 176(3):865–888, Mar. 2009a. doi: 10.1111/j.1365-246X.2008.04010.x.

Zhou, Y. Surface-wave sensitivity to 3-D anelasticity. *Geophys. J. Int.*, 178(3):1403–1410, Sept. 2009b. doi: 10.1111/j.1365-246X.2009.04230.x.

Zhou, Y., Dahlen, F. A., and Nolet, G. Three-dimensional sensitivity kernels for surface wave observables. *Geophys. J. Int.*, 158(1):142–168, July 2004. doi: 10.1111/j.1365-246X.2004.02324.x.

Zhou, Y., Dahlen, F. A., Nolet, G., and Laske, G. Finite-frequency effects in global surface-wave tomography. *Geophys. J. Int.*, 163(3):1087–1111, Dec. 2005. doi: 10.1111/j.1365-246X.2005.02780.x.

Zhou, Y., Nolet, G., Dahlen, F. A., and Laske, G. Global upper-mantle structure from finite-frequency surface-wave tomography. *J. Geophys. Res.*, 111(B04304), 2006. doi: 10.1029/2005JB003677.

Appendix A: Phase delay measurements using multi-taper technique

Let $s(\omega) = A(\omega)e^{\phi(\omega)}$ be the mathematical expression of the reference seismogram computed for the 1D reference model for a given source-receiver pair at some frequency ω, with amplitude A and phase ϕ. Let $o(\omega) = A^o(\omega)e^{\phi^o(\omega)}$ be defined equivalently for the observed seismogram, or the SEM seismogram in the case of this synthetic study. The accumulated phase results from source and receiver effects, caustics and the propagation itself (e.g. Ekström, 2011; Ma et al., 2014; Moulik et al., 2021). We typically assume the first three terms are the same for both the reference and observed seismograms. In that case, the phase delay can be directly related to the propagation and thus perturbations in the Earth model. These phase delays are what we are interested in measuring here.

Waveforms are first pre-processed (e.g. resampled at 1 Hz, instrumental response removed if necessary). As suggested by Zhou et al. (2005) and Zhou (2009a), we then use a multi-taper technique to measure the phase-delays and to obtain an estimate of the measurement uncertainty (e.g. Thomson, 1982; Park et al., 1987a,b; Laske et al., 1994; Laske and Masters, 1996; Hjörleifsdóttir, 2007). The technique uses the first few Slepians (after Slepian, 1978) defined

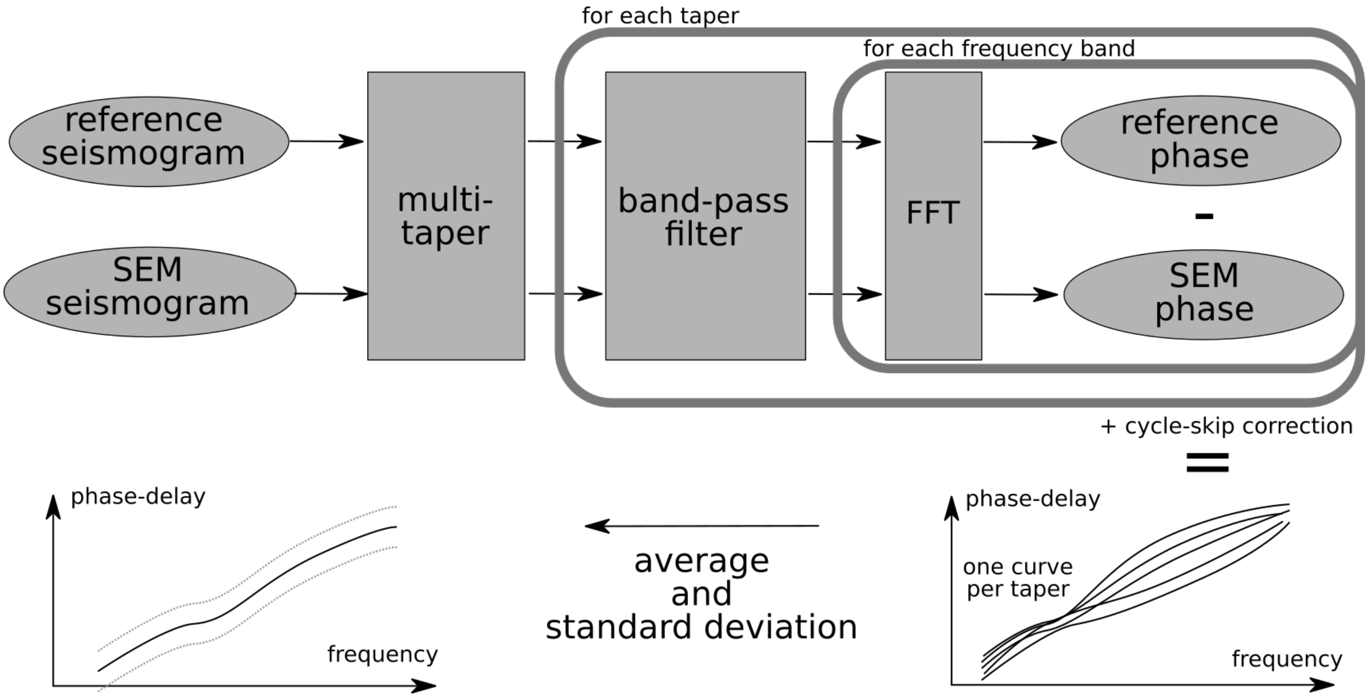


Figure 10 Overview of the measurement workflow. We compute a reference seismogram for the reference radial Earth model, which we use to measure the phase-delay of a SEM-computed seismogram (acting in this synthetic setup as observed seismogram). We apply a set of tapers (the five first Slepians), thus leading to 5 tapered traces. We filter each in a set of frequency bands, before we take the FFT. In the frequency domain, we then compute the phase difference for all frequencies for all tapers, producing a set of 5 dispersion curves. We apply a cycle-skip correction and then take the mean of all 5 tapers as the final measurement, with the measurement uncertainty given by the standard deviation of the five tapers.

over a 801 s window. Slepians are an infinite series of functions with optimal frequency spectrum (therefore reducing frequency leakage) that weigh different parts of the waveform (thus reducing bias in the time-domain). With a 801 s-long time-window and 1 Hz sampling rate, we should use only the first 5 Slepians (see Percival and Walden, 1993, pp. 331). To position the Slepians, we compute the predicted group arrival time at the frequency of interest, starting the Slepian time window 150 s before the expected arrival. We then apply a 4 mHz-wide bandpass filter around the frequency of interest before we compute the Fast Fourier Transform. Finally, we subtract the phase component of the tapered and filtered observed (or SEM here) waveform from the reference waveform in the frequency domain. Usually, we obtain a smooth dispersion curve, except for when the phase delay reaches $\pm\pi$, where the dispersion curve makes jumps of $\pm 2\pi$. Low frequencies are less likely to suffer from cycle-skips. Therefore, we make our measurements at increasingly higher frequency, starting at 6 mHz. When we detect these so-called cycle-skips (we use a threshold of ± 4 radians for the detection), we add or remove 2π to obtain a smooth dispersion curve and apply this correction accordingly to all higher frequencies.

For each source-receiver pair, we end up with 5 dispersion curves for the 5 Slepians, corrected for cycle-skipping. We use the average of these 5 curves as our final measurements and the standard deviation as the data measurement uncertainty. In some cases, we note an inaccurate detection of cycles-skipping (either as false-positive or false-negative). These false detections typically do not occur on all five tapers, leading to a sharp increase in measurement uncertainty. In addition, some fundamental mode measurements are contaminated by the interference of other phases or higher modes. This usually does not affect all five tapers, thus also leading to an increase in the measurement uncertainty.

Appendix B: Computational considerations

In this study, we use $N = 47\,700$ fundamental mode phase delays as data and we parameterise the spatial domain into $M = 259\,200$ voxels (cells of size $2° \times 2°$ laterally and 25 km depth for the first 400 km depth of the whole mantle). Therefore, the sensitivity matrix $\boldsymbol{G}$ of size $N \times M$ is reasonably large. To optimise the sparsity of the sensitivity matrix, we only consider the sensitivity kernels in the two first Fresnel zones laterally, since their amplitude is negligible further away. The sensitivity is also negligible at depths greater than 400 km depth. Our resulting matrix thus contains $645\,282\,622$ non-zero elements, i.e. the sparsity is approximately 5.2%. The SOLA optimisation problem (Equation 5) leads to a set of normal equations taking the form of another $(M+1) \times (N-1)$ matrix $\boldsymbol{Q}$ that is less sparse than $\boldsymbol{G}$ (see Zaroli, 2016, Appendix A). Reordering the lines of $\boldsymbol{G}$ with the sparsest row first helps to improve the sparsity of $\boldsymbol{Q}$. In this study, $\boldsymbol{Q}$ contains $657\,124\,288$ non-zero elements, i.e. sparsity is approximately 5.3%. On disk, we use a 'coordinate list' (COO) storage strategy, and $\boldsymbol{Q}$ takes up ∼17 GB. On RAM, we use a reversed linked-chain storage strategy to improve compute time. In this case, the $\boldsymbol{Q}$ matrix takes up ∼35 GB. This large memory requirement is the primary limiting factor for increasing the number of data and model parameters.

The computation time of the LSQR inversion for a single model parameter depends on the target resolution and trade-off parameter. With the choices made in this study, it takes ∼100 s per model parameter. As we invert for $69\,200$ model parameters, a full model estimate thus requires $\sim 692\,000$ s CPU time (or 192 CPUh). In practice, we invert for model parameters in parallel on several nodes with 128 CPU each using a multi-threading approach with OpenMP. The scaling is not fully linear due to input/output operations, but this strategy reduces the wall time to ~ 20 h.

Response to the reviewers

We thank the Editor for handling the review process and both reviewers for their critical assessment of our work. Below, we provide a summary of their primary concerns and the main changes we have made to address these, before responding to each comment in detail. We have included the sticky notes provided by Reviewer 1 in an annotated PDF with line numbers to facilitate the review process. Corresponding changes in the manuscript are indicated in blue (new) or red (removed) text. Line numbers correspond to line numbers in the PDF with tracked changes.

Summary

The concerns of Reviewer 1 and 2 relate to a few main points. A primary point of concern raised by Reviewer 1 (Point 1.1 and 1.2) focused on our definitions of uncertainties. Particularly, they did not like our treatment of theoretical uncertainty, and how we defined it as a component of the data uncertainty. We agree with their comment, and we have amended the text accordingly throughout the paper, while we specify early on what the theoretical uncertainty entails. In addition, some concepts related to the SOLA inversion were not sufficiently clear in the submitted manuscript. As suggested by Reviewer 2 (Point 2.1) and the Editor, we have added an appendix (Appendix A in the revised manuscript) to provide more details on the SOLA inversion, making this paper more self-contained, and hopefully clarifying some fundamentals on the SOLA inversion. Several comments of Reviewer 2 concern the choice of η, the trade-off parameter and how this was chosen. While this is a choice in SOLA, all models based on different values of η are valid solutions, as long as they are interpreted together with their respective uncertainty and resolution. We have clarified this in responses to Point 2.3 and Point 2.6. Other comments focus on the choice of input model, the weighting of data (Point 2.2), the data and measurement procedure (Points 2.4 and 2.5 and non-linearity Point 2.7). We hope our responses, together with the changes made in the manuscript, address the comments satisfactorily.

Reviewer 1

This manuscript describes a combination of the SOLA method (a variant of Backus-Gilbert inversion) with finite-frequency theory for surface waves. Following a detailed introduction into surface wave tomography and the need to quantify uncertainties, the authors provide a derivation of their method and a synthetic inversion at global scale. The manuscript is well written and a pleasure to read. I have no doubt that the method, maybe after some improvements to increase computational efficiency, will be an attractive alternative to existing approaches, especially when uncertainty quantification is critical for the scientific question that one actually aims to answer. Also, the paper will surely become a valuable addition to the existing seismic tomography literature. Most of my comments are minor and contained in the annotated manuscript. They primarily concern inaccuracies in some lines of arguments. Two more significant concerns, detailed below, are related to (1) sources of model uncertainty that may be more significant than the ones addressed by the authors, and (2) the concept and treatment of forward modelling (theory) errors. In summary, all of my comments can be addressed with improvements of the text and maybe some additions to the discussion section. Since no additional simulations are needed, I would classify this as a minor revision.

Reply: We thank Reviewer 1 for their critical review and we hope our answers to the following comments and subsequent changes in the manuscript will have addressed their concerns.

Major

Reviewer Point P 1.1 — Sources of model uncertainties

1. The authors mainly consider two sources of model uncertainties: errors in the data and forward modelling errors, i.e., inaccuracies of the underlying theory. While these two are certainly important, they are only part of the story.
2. Nonlinearity can have a major impact, and not taking it into account is a limitation of the method that deserves a more honest discussion. There are non-linear versions of the Backus-Gilbert method, e.g., by Roel Snieder, that should probably be mentioned. Using a fixed, and unavoidably inaccurate, crustal model is related to nonlinearity, too. Iteratively improving tomographic models can have an enormous impact, as evidenced by recent models that use hundreds instead of just a few tens of iterations (e.g., REVEAL by Thrastarson et al., or WUS324 by Rodgers et al.).
3. Probably the most important source of model uncertainties are all the little choices that we make along the way. This includes the shape and size of the target resolution kernels, the code used to compute artificial data, approximations made to compute sensitivity kernels, and, again, the choice of the crustal model. Without being able to proof this, my gut feeling is that this is the zeroth-order contribution to model uncertainties. Of course, I would not request the authors to solve this issue, but it deserves to be mentioned in the discussion.

Reply:

1. The forward modelling uncertainty estimate accounts for a number of errors mentioned below by Reviewer 1. Non-linearity, approximations in the sensitivity kernels and reference seismograms, should contribute to the discrepancy between the filtered input and output models, and therefore contribute to what we regard as theoretical error, though non-linearities are undoubtedly stronger in reality than in the input model we have chosen. It is true however that some contributions to the theoretical uncertainty are not accounted for, particularly uncertainties in the crustal model or the earthquake source parameters.

   We have clarified what should be accounted for what we regard as the theoretical uncertainty estimate when introducing it in the introduction at L115-123. Additionally, the paragraph in the discussion (section 6.2, L710-716) makes clear what errors are lacking in the theoretical uncertainty estimate (where we also added the the fact we use a optimal reference model as factor, and that non-linearities in the Earth are expected to be stronger than in S362ANI).

2. We completely agree that iterative non-linear methods are important and that is certainly a direction that seismology is heading. However, that does not mean there is no merit in linearised methods. They allow us to test and investigate assumptions in data and modelling, while the results can often be a valuable starting point for non-linear approaches. In addition, the resolution and uncertainty of models developed using non-linear methods are more difficult to assess.

   We have added this limitation in the discussion (L724-730) and mentioned here also several of the suggested studies suggested to contrast with SOLA.

3. Most of 'the little choices that we make along the way' should contribute to our definition of the theoretical uncertainty. In our linearised setup, we can separate resolution and uncertainties, in contrast with non-linear approaches. Specifically, the model uncertainty is defined as the variance attached to local average estimates. Therefore, the resolution itself does not contribute to the uncertainty.

   We have emphasised this in the text at L192-193, and we hope that the added appendix with further details on the SOLA method (as suggested by Reviewer 2 and the Editor) will make this point clearer.

Reviewer Point P 1.2 — Theory errors
I can understand the authors' idea behind the estimation of forward modelling errors. Loosely speaking, everything above an rms misfit of 1 is interpreted as forward modelling errors, which are added to the observational errors.

1. Apart from the fact that this line of arguments breaks down for non-Gaussian errors, there are a few other issues
2. An rms significantly above 1 can also result from a parameterisation that is too coarse, or from a target resolution that is too low. This has nothing to do with forward modelling errors. Hence, by changing these subjective choices, the forward modelling error changes, thereby turning it into an arbitrary quantity.
3. Forward modelling errors are not random but deterministic. Hence, it does not make sense to add them to the data errors and to treat them as a random variable.
4. Going a bit further, I would even argue that there is no such thing as a forward modelling error. In fact, the notion of "error" is meaningful only in the context of a reference. Error with respect to what? For a forward model such a reference does not exist because all models are wrong from the outset, and there is no universal measure for the quality of a model. (They are mere abstractions of nature.) It follows that it does not actually make sense to introduce a forward modelling error, because it is an arbitrary quantity anyway.

Reply:

1. The reviewer is correct in pointing out that our argumentation breaks down in the case of non-Gaussian errors. This is a simplification that we cannot avoid in the current work, but others have worked on this with SOLA. The only thing we can do at present is to be honest about this limitation.

   We mention this limitation in the theory section at L161-162 and we also remind the reader of it in the discussion at L701-707.

2. We also agree that the parameterisation contributes to the theoretical errors.

   We mention this at L121-122 (when addressing point P 1.1).

   It is true that the model uncertainty due to theoretical errors depends on the resolution. However, this is natural since the model uncertainty is for local averages in the SOLA approach. In other words, the theoretical uncertainty is constant, but its propagation into model uncertainty is resolution-dependent.

We added this point in the discussion at L694-700.

3. Forward modelling errors are indeed deterministic. However, for mathematical convenience, we treat them statistically as random variables.

   We mention this now in the discussion at L651-652.

4. We agree that it is difficult, or maybe impossible, to define what the general impact of the forward modelling is and how it relates to our theoretical uncertainty. However, we believe we can clearly identify assumptions in our forward modelling that will lead to errors in the model solution compared to the true Earth structure. In this study, this is relatively clear as we work in a synthetic setup: the 'true' forward modelling is the Spectral Element Method used to compute the SEM seismograms, which function as observations. Every assumption we make in the linear finite-frequency approach used to compute the sensitivity kernels that deviates from the SEM leads to an increase in the theoretical errors. This is nevertheless dependent on the SEM seismograms representing thee ground truth.

   We added text on this in the discussion section at L721-723.

Minor

Reviewer Point P 1.3 — L43: I do not quite understand why you want to limit the importance of your work to the oceans. There are many more regions where you do not have good coverage: the Antarctic, Russia, South America, most of Africa, ...

Reply: Amended.

Reviewer Point P 1.4 — L49: This statement is only party correct. Even if you had perfect coverage, you would still not have a unique inverse.

Reply: We have reformulated the text.

Reviewer Point P 1.5 — L60: This statement is obviously wrong. The vast majority of tomographic models come with some sort of resolution estimate, although this may sometimes not be very sophisticated.

Reply: We meant to say that these approaches do not account *directly* for these, with examples of different strategies that have been used to provide estimates given in the following sentence.
We have adapted the text to clarify this, which also addresses point P 2.8 of Reviewer 2.

Reviewer Point P 1.6 — L97: I think you are missing a classic here: Yomogida et al., 1992

Reply: Added.

Reviewer Point P 1.7 — L113-120: I think the logic here is not correct. Data errors and forward modelling errors are two completely different creatures. Yet, when solving an inverse problem, you can treat these two in a similar way.

Reply: We agree with Reviewer 1 that even if the maths is similar, the concepts are distinct. In the revised paper, we now clearly distinguish between measurement (or data) uncertainty and theoretical uncertainty, which propagate into model measurement uncertainty and model theoretical uncertainty, respectively.
We have reformulated the whole paragraph (penultimate paragraph of the introduction) where we introduce the different uncertainties to make this clear. We have also amended this throughout the manuscript wherever needed.

Reviewer Point P 1.8 — L113: This statement is quite obviously wrong. Data have errors due to noise, even when you do not even have a forward theory. My 4-year-old son can pick travel times, but he does not know about ray theory!

Reply: It is useful to see that our statement can be interpreted in different ways.
We hope the rephrasing made to address point P 1.7 also addresses this comment of the reviewer.

Reviewer Point P 1.9 — L115-123: Sorry, but this is not very clearly expressed.

Reply: We agree that this was not phrased clearly (perhaps related to the fact that defining theoretical uncertainty is difficult).
We have removed this sentence at L115-123, and instead now explicitly list all the contributions that we identify as theoretical errors.

Reviewer Point P 1.10 — L133: I am curious to see how you do this later in the paper. At this point, this sounds like a self-contradictory statement because theoretical errors are not random but deterministic.

Reply: As mentioned in our response to point P 1.2, we primarily make this choice for mathematical convenience. This is indeed discussed later in the paper.

Reviewer Point P 1.11 — L144: Control of what?

Reply: Rephrased.

Reviewer Point P 1.12 — L155-157: Note some inconsistency in your logic. In the introduction you mention that accounting for forward modelling errors is a major contribution of your work. Here, however, you write that you make simplifications that are probably much more important: Gaussian uncorrelated errors!

Reply: The reviewer is correct that this is a significant drawback. However, it is a commonly made assumption and therefore, we feel that even under this assumption, we can learn more about the contribution of theoretical errors to the model uncertainty.
We added a sentence at L160-162 to make clear that this is a significant drawback that should motivate future work. We have also rephrased the main text in the abstract and introduction to be more honest on what we can achieve given this limitation.

Reviewer Point P 1.13 — L164-170: I am struggling a bit with these sentences because they do not seem to be connected. It may help to say more explicitly that ill-posedness motivates regularised least-squares solutions, and then you provide a way of writing these down.

Reply: We agree that these sentences were poorly phrased.
We now clarify the main aim of data-fitting approaches, discussing that this is fundamentally different from the SOLA philosophy, before we start describing the SOLA method in detail.

Reviewer Point P 1.14 — L178: Note that this is a contradiction in itself. There is no such thing as a true model! Absolutely all models are wrong because they are simplified abstractions of nature.

Reply: We refer here to the velocity structure and in our synthetic case, the 'truth' is very clear and well defined, as it is the Vs structure of model S362ANI that we aim to retrieve. In general, we agree that there might be some subjectivity in how we define the 'truth', we also believe that it is the point of statistical methods to find an estimator that estimates some 'true' physical property.
To make clear that there may be different definitions of 'true', we therefore have added simple quotes around this word wherever it appears in the manuscript.

Reviewer Point P 1.15 — L185-187: A reference would be useful here.

Reply: This is the affine transformation of a multivariate normal distribution.
We have added this description in the text to make clear where this equation stems from, but there is no straightforward reference for this.

Reviewer Point P 1.16 — L189-192: Maybe add that you define uncertainty to be the standard deviation. Strictly speaking, however, this is not correct. Uncertainty is a much broader term, which include inter-parameter covariances, for example. It may be better to just call it the standard deviation.

Reply: Thanks for this comment. That we assume uncorrelated data noise is a limitation, but such correlations would be difficult to estimate. However, in SOLA, we do not aim to consider off-diagonal terms of the model covariance matrix for the uncertainty. While we could compute these, this information is already embedded in the resolution kernels, and we therefore do not think there is a need to account for them in the uncertainty.
We have now clarified that we use the standard deviation as the uncertainty, both for the data and model estimate, at L158 and L189-193.

Reviewer Point P 1.17 — L195-199: Try to be fair to all the people who developed methods that go far beyond that. With statements like that you implicitly say that the vast body of literature on Monte Carlo methods, for instance, can be disregarded.

Reply: We did not mean to imply with this statement that the literature on these methods can be disregarded. However, as far as we understand, Monte Carlo methods still stem from a data-misfit point of view, in contrast to Backus-Gilbert methods. Both these methods and data-fitting approaches have their pros and cons and we were primarily justifying the use of SOLA to achieve the particular objective of accounting robustly for resolution and uncertainty.

Nevertheless, we understand from the reviewer's comment that this statement might have come across as too strong, and we have therefore rephrased it.

Reviewer Point P 1.18 — L205: The left-hand side of this equation is missing. What is this quantity that you are optimising?

Reply: Thanks for pointing this out, the equation has been corrected.

Reviewer Point P 1.19 — L209-212: Maybe add a sentence to explain this. Why exactly shall we bother about this?

Reply: We have added a sentence to explain the importance of this constraint.

Reviewer Point P 1.20 — L219: This statement is incorrect. The generalised inverse never depends on the data! This is not a particularity of SOLA.

Reply: We intended to state that the optimisation problem is not driven by the data, while the data are directly involved when minimising the data misfit in other approaches.
We have reformulated this statement and believe that the new appendix may also help to clarify this.

Reviewer Point P 1.21 — L226-227: Is this one of the simplifications you consider in the forward modelling errors? Maybe mention this here.

Reply: Indeed. We now mention this at L230-231, and we also include this in the list of theoretical error contributions (see our response to point P 1.9).

Reviewer Point P 1.22 — L235-239: Not sure the logic of this statement is correct. Strictly speaking, there is not approximation here. What you do is to compute a first derivative operator. Such an operator is by definition linear and for wave propagation problems it corresponds to single scattering because higher-order scattering corresponds to later terms in the Taylor expansion, i.e., not the first derivative anyway.

Reply: Yes, this is what we mean here. Stating that we use only the first order of the Taylor expansion (in the mathematical sense) or assume single scattering (in the physical sense) is equivalent. This is an approximation in that we ignore higher order terms or multiple scattering.
We have added a sentence to mention the equivalence with the terms of the Taylor expansion for clarity.

Reviewer Point P 1.23 — L244: Style: Sentences should not start with math symbols.

Reply: Amended.

Reviewer Point P 1.24 — L241: I would suggest to simplify the equation above if you do not use this anyway.

Reply: Amended.

Reviewer Point P 1.25 — L270-273: Maybe explain which kind of grid you use. It probably is not latitude and longitude as this would make a very small spacing near the poles.

Reply: We do use a local parameterisation in latitude and longitude for the whole sphere, but only the first 400 km depth, as stated in the manuscript. Note that we account for the norm of the basis functions wherever it applies through the $\mathcal{V}_j$ terms.

Reviewer Point P 1.26 — L287-289: I think it would be good to explain why you actually need to store that matrix explicitly. Most tomographic methods compute matrix-vector products on the fly, avoiding matrix storage. Why exactly can you not do this?

Reply: We made this choice to improve data I/O, which we expect to be slower if the matrix is not stored in RAM. We would however welcome suggestions on how to improve this. We have added a sentence here to explain this.

Reviewer Point P 1.27 — Fig 2: Please add lat/lon labels to the figures and enlarge the labels in panel d) bottom.

Reply: This is something we have thought about ourselves, but we believe that latitude/longitude labels are not necessary here. The scale of the maps and the coastlines make clear what region we are looking at, and we do not refer to lat/lon coordinates anywhere in the text. Therefore, to maximise the aesthetic of these maps, we prefer not to add these labels.
In response to the reviewer's comment, we have removed the depth labels in the cross-section and we now refer to the meaning of the depth lines in the figure caption (for this figure and all other similar figures in the paper).

Reviewer Point P 1.28 — L305: kilometres (I think you mostly write in BE.)

Reply: Amended everywhere.

Reviewer Point P 1.29 — L312-315: This exactly the point I was trying to make above. It is this freedom of choice that controls the difference between tomographic models, and not the forward modelling uncertainty.

Reply: We believe it is both, the resolution as well as the uncertainty due to choices in the forward modelling. This is also what we write at L46-48.

Reviewer Point P 1.30 — L326-327: Sorry, but this statement contradicts itself! You ignore effect X because you want to study effect X.

Reply: This is not what we intend to say or do. We ignore X (the uncertainty) in driving the inversion (to obtain the generalised inverse), but then we do use X to study the effect of X, i.e. we do propagate the actual uncertainty we measure into model measurement uncertainty. We do this intentionally to purely analyse the propagation of errors, but we then return to this in the discussion.
We have reformulated the text to make this clearer.

Reviewer Point P 1.31 — L329: Note that "unitary" has a meaning that is different from what you have in mind here!

Reply: We have rephrased this sentence to remove this expression from the text.

Reviewer Point P 1.32 — Fig 3: The labels in this figure are a bit small.

Reply: As discussed in our response to point P 1.27 we have now removed the depth labels from the cross-sections and instead added the meaning of the depth lines in the caption.

Reviewer Point P 1.33 — Fig 4: Please enlarge the labels a bit.

Reply: See the response above.

Reviewer Point P 1.34 — L372-373: I am wondering if this is a reasonable choice. S362ANI is such a simple model that it can be reconstructed with any tomographic method. Would it not have been more interesting to use an input model that is a bit more challenging? (Basically, with such a simple model you can only find out that your method works, but you cannot explore its limitations, which is more interesting.)

Reply: We thank the reviewer for this suggestion. It would certainly be interesting to test an independent (not tomographically derived model of the mantle) and this is certainly something that could be done in future. However, we also would not want to discard the value of the lessons learnt with model S362ANI for which synthetic seismograms were already available (thus also providing a computational advantage). With a more challenging model, we would likely note that the resolution does not allow to resolve smaller-scale features. Additionally the uncertainties (measurement and theoretical) would probably be larger for a model that is more complex. We mention this in the Discussion: in the 4th paragraph of section 6.1 (L588-594) on the resolution, and in the 6th paragraph of section 6.2 (L684-688) on the uncertainty. More importantly, since SOLA is mostly concerned with obtaining the statistics underlying the model estimate, the method would perform better for a more challenging model as it would inform us about specific features of that model that cannot be retrieved. As an aside, the SOLA minimisation problem to find the generalised inverse does not depend on the input model. Therefore, the approach should work for any model.

Reviewer Point P 1.35 — L377: Same comment as above. Is this not too much of an inverse crime?

Reply: We commit three types of inverse crimes in this synthetic study: using the perfect 1D reference model, the perfect crustal model, and perfect earthquake source solutions. The first favours linearity and therefore limits the contribution of non-linear effects to the theoretical errors, the last two fully remove the contribution of errors in the crustal model and source solutions to the theoretical errors. It would certainly be interesting to investigate the impact of these errors on the model uncertainty. Unfortunately, it is difficult to have reliable uncertainty estimates on the source, crustal model, and 1D Earth model, and to investigate their effect would represent a significant amount of work. This truly deserves a separate study and there is ongoing work in our group focused on the uncertainties due to the source parameters.
We have extended the third to last paragraph of the discussion in section 6.2 'Robust model uncertainty?' (L710-723) to mention the inverse crimes related to the earthquake source and crustal model and we have added the contribution of the reference model. We also now mention these factors early in the

manuscript in the introduction (L123 of the revised manuscript) when listing the contributions to the theoretical errors.

Reviewer Point P 1.36 — L387: Since this is a synthetic study, this seems to eliminate the data that are actually the most interesting.

Reply: Although this is a synthetic study, we aim to mimic a real study, where similar criteria would be considered reasonable.

Reviewer Point P 1.37 — L397: Why not?

Reply: In our linear setup, we require these corrections as the inversion for crustal structure is highly non-linear. We have added this reasoning at L397-398.

Reviewer Point P 1.38 — L400: I see your logic here, but still note that this amplifies the extent to which this is an inverse crime.

Reply: Please see our response to point P 1.35.

Reviewer Point P 1.39 — Fig 6: The labels are too small.

Reply: Amended.

Reviewer Point P 1.40 — L462-463: This statement is not correct. No matter how perfect your data are, there is still a nullspace! Hence, there are infinitely many models that are equally good. That your output model looks the same as the input model simply means that your inversion methodology has a bias towards producing a model that you like.

Reply: Reviewer 1 may have misinterpreted this sentence. We primarily state that in the case of error-free data the output model is the same as the *filtered* input, not the input model itself. We can show this mathematically: perfect data are computed with $\boldsymbol{d} = \boldsymbol{G}\boldsymbol{m}_{\text{input}}$, and the output model is computed with $\boldsymbol{m}_{\text{output}} = \boldsymbol{G}^{\dagger}\boldsymbol{d}$, thus replacing by the perfect data we have, $\boldsymbol{m}_{\text{output}} = \boldsymbol{G}^{\dagger}\boldsymbol{G}\boldsymbol{m}_{\text{input}} = \boldsymbol{R}\boldsymbol{m}_{\text{input}} = \boldsymbol{m}_{\text{filtered}}$. As a note, with SOLA, we break the non-uniqueness by looking at local averages. This philosophy allows us to retrieve the filtered input exactly in the case of error-free data.

Reviewer Point P 1.41 — L471: For my taste this paragraph is a bit too long. After all, this is a synthetic study where it has limited meaning to discuss how different features do or do not come out.

Reply: We agree that there is little value in analysing specific geological features in this synthetic setup. However, we believe that there is value in checking whether the method is able or not to retrieve particular features in the input model, as this will be a primary objective when applying the method to real data. It is also informative to note where the method performs well, where it fails, and to infer why this is the case (see also our response to point P 2.5.

Reviewer Point P 1.42 — L472-473: I do not think this logic is correct. Input = output can also mean that your input model is so simple (smooth) that you can recover it well no matter what you do. (See my earlier comment on choosing an input model that may be too simple.)

Reply: Please see our response to point P 1.34 on this topic. This is now also discussed thoroughly in the discussion section, but we have added a sentence here at L477-478 as well to specify that we would likely miss smaller scale features when using a rougher input model.

Reviewer Point P 1.43 — L500: Should this be the other way around?

Reply: No, this is the correct way as can be seen on the maps, as well as in the cross-sections. This is because poorly constrained regions are overcompensated by very poor resolution, which can lead to low uncertainty. For example, it can be counter-intuitive that uncertainty decreases at large depth, but this is simply explained by the fact that the resolution does not represent what we think we are looking at (but represents shallower structure). This shows the importance of having both resolution and uncertainty when interpreting a model. Reviewer 1's comment also shows the importance of actually quantifying resolution and uncertainty as these may vary in counter-intuitive ways.
We discuss this point in the last paragraph of section 6.1, where we link the variation of resolution with that of model uncertainty: the model uncertainty can only be interpreted together with the resolution.

Reviewer Point P 1.44 — L525-526: I can follow the idea but I still think it is incomplete. Please see main text of my review.

Reply: We hope our responses to previous comments and changes in the manuscript will have addressed this concern concerning the definition of the measurement and theoretical uncertainties that is indeed central in this review.

Reviewer 2

The authors have provided an interesting exploratory paper, trying to bring together several well-known problematic aspects in seismic imaging and possible directions to address them. I think this study could be important for people interested in mantle imaging with methods that deviate from the common assumptions adopted (damped least squares or probabilistic sampling). The manuscript is well written, even if I think some additional details could be beneficial to it (see comments below), the title and the abstract fits well the topics discussed. All the pictures are relevant in the context they are presented in.

Reply: We thank Reviewer 2 for their views and their comments that have helped to improve the manuscript.

Reviewer Point P 2.1 — I had to familiarize a little with the methodology used, reading the work of Latallerie et al. 2022 and the precedent works of Zaroli et al. 2016 and Zaroli et al. 2017. Considering this, my first comment is that this study works well following up the precedent papers, but it could be not enough self-explanatory to be fully understandable by itself. For this reason, I suggest expanding the methodology section to provide some additional information about the SOLA algorithm (eventually, this could be even presented in the Appendix). At the end, this study revolves mostly around methods, not results, so I suppose the methodology should be outlined with some additional details.

Reply: We appreciate that Reviewer 2 took the time to read through previous papers that have used the SOLA inversion to familiarise themselves with the methodology. As suggested by Reviewer 2 and the Editor and to make this paper self-consistent, we have added an appendix on SOLA (Appendix A in the revised manuscript), containing an illustration, that we hope will help future readers to grasp the main ideas of the SOLA method without having to read through other literature.

Major

Reviewer Point P 2.2 — L324-327: I suppose that the idea of building the generalized inverse matrix assuming uncorrelated equal errors in all the observations (data covariance matrix = identity matrix) rises from the difficulty in assessing the true uncertainty (measurement + theoretical) and the influence this can have in the inversion process. However, since you are working with a wide range of surface waves' periods, I would expect the noise in the measurements to manifest some kind of correlation with the frequency. So, either - by using the estimated uncertainties you consider more "robust" - you justify assuming the same error (i.e., weight) at all periods in the first step, or I would like to see discussed the impact different weights attributed to the data can have on the generalized inverse and on the final solution. I suppose fixing the data covariance matrix with reasonable uncertainty estimates and building multiple solutions can illuminate this. Another reason why I consider this relatively important is because of the strong – and unexplained (not undescribed) - deviations your solution model manifests from the input filtered model. I wonder if one possible cause could be in the way the generalized matrix is built. The deviations between the solution model and the filtered input model seems to be largely due to fine scale structures, I would not be surprised if the average deviation could be reduced by a more appropriate weighting scheme.

Reply: The reviewer makes an interesting suggestion. In the current setup, our preference was to use equal data weights, in order to limit the number of subjective choices for designing the generalised inverse. Note that we do use the measurement uncertainty estimate and propagate this into model uncertainty. We slightly rephrased this passage to make this clear.

Given the different sensitivities, we agree that setting all weights equal to one is not optimal. Nevertheless, given this limitation, it is promising to see that the model estimate does not deviate much from the input model. While it would be informative to build a range of generalised inverses with different choices of weights, this also represents a significant computational burden that we prefer to avoid here. However, this will be investigated further in future, and we will consider a better weighting for constructing a more optimal generalised inverse in real applications.

We added a sentence on this at the end of the paragraph at L330-331

Reviewer Point P 2.3 — L355-357: the part about resolved deep kernels representing an average of the shallower portions of the model is particularly interesting. I suppose this is dependent on the choice of the trade-off parameter "eta" that prioritizes low covariances for the model increasing the deviation of the resolved kernels from the target ones. However, this makes me wonder if the trade-off parameter could be selected to partially prevent these cases (?). If I understood correctly, error propagation factor decreases around 200 km depth (fig. 5) because resolved kernels represent the average of large shallower regions of space (fig. 4), enclosing many finite frequency kernels, resulting in low resolution and low model's uncertainty. Is this correct? I would like to see a comparison between target and resolved kernels toward 400 km depth, where the uncertainty propagation factor seems to drop to 0.

Reply: This is totally correct. We have tried to retrieve the deeper structure with different values of 'eta', but without much success. In fact, even when using 'eta' values to favour a low resolution-misfit, we observe that averaging kernels tend to over-fit the targets laterally, which dramatically increases the uncertainties without much improvement vertically. Instead, it is likely that the data themselves cannot constrain the depth better, which make sense as we only use fundamental modes here.

A comparison of the target and averaging kernels is given in figures 3 and 4 (blue ellipses versus coloured grid cells) for the depth 112 km and 212 km. While it would be interesting to investigate the structure at 400 km depth, we did not run the inversion below 212 km depth where the depth leakage is already strong and the structure is unconstrained. We are currently searching for an approach to balance this depth leakage better, for example by changing the trade-off parameter with depth, but this represents additional computational costs. *We added a sentence on this at L611-614*.

Reviewer Point P 2.4 — L384-396: since your method uses a subset of the dataset to reduce the storage required, it is my opinion that an approach more robust than randomly selecting rays excluding similar paths would be i) using a binning technique to merge similar-paths, extracting more robust estimates for the observables (and also uncertainties…), or ii) using a procedure similar to bootstrapping, where the dataset is randomly subsampled multiple times, and the solution models resulting from each subset are brought together in an ensemble of solutions. This allows, using metrics like average and standard deviation, to explore central features and poorly constrained anomalies highly dependent on the specific subset. Considering that many of your measurements have the uncertainties potentially underestimated, method ii) could be interesting to compare the uncertainties in your model coming from bootstrapping with the error propagation you use to compute model's covariance matrix. I will not explicitly ask to add these to the study, but if you agree that these approaches could work, maybe this discussion should be added to the text, or considered for future experiments.

Reply: We thank Reviewer 2 for these suggestions. We have considered approach i), but this would be difficult to implement in our setup given that finite-frequency sensitivity kernels depend on the specific source mechanism, so that they would not be straightforward to compute for the summary rays. Approach ii) however is very interesting for exploring features in the dataset. It would be more difficult to investigate the effect on the model estimate, as this would require to build and invert several sensitivity matrices, which requires considerable computational resources.
We now mention these approaches when presenting our approach for obtaining the measurement uncertainty at L393-395, as well as in the discussion about uncertainty at L640-643, and we may consider them in future work.

Reviewer Point P 2.5 — L409-429:

1. I find a little concerning that your forward modelling strategy can describe only 30% of the selected dataset. Considering that the - relatively accurate - approximations adopted in this study should account for the primary elements controlling surface wave phase observables (especially if simulated), I cannot fully understand why almost 70% of the measurements deviate so much from the predictions of the true model.
2. I suppose phenomena like cycle-skipping could be identified by comparing the same event-station phase measurement at different periods and checking if one breaks an expected trend (since it should be wavelength dependent…).

3. If a wide range of frequencies gives rise to cycle-skips, like in fig. 6, I wonder if alternative measurement methods based on nearby stations cross-correlations (Jin & Gaherty, 2015), could provide more reliable observations, at least for some problematic cases.

4. In the pictures you show in Fig. 6 (top), the case representing a class III measurement is shown as a surface wave traveling across North America, while in the other two – non problematic – cases the propagation is across the Pacific. Could you plot a map with the "rays" connecting events to stations (like Latallerie et al. 2022, fig.1a) colored by the deviation (measurement – input prediction)? Just to check if the deviations manifest some sort of correlation with the waves' paths.

5. Does using finite frequency sensitivity kernels improve data-misfit with respect to ray-theory predictions (considering the true model)? I would like to better understand if this is mostly a problem with the measurement algorithm, or with the forward modeling you adopt in the code. Either way, this could affect real data applications of the inference code, where no true model predictions exist to compare with the measurements.

Reply:

1. The volume of data falling in each class depends on the (arbitrary) definition of these classes, that is the number of radians and the number of standard deviations between the analytical predictions and measurements. The results presented in figure 8 make clear that the overall dataset is not as bad as the volumes in each class portray it to be. We note that 56% of the dataset has a difference of less than 3 radians between the predictions and the measurements.
   We have adjusted the definition of the classes slightly to reflect the overall behaviour of the data set, and we have clarified this situation in the text at L430-442.

2. Regarding the cycle-skip, we do correct for this by detecting 'jumps' as frequency increases in the dispersion curve for each event-station pair, as described in Appendix A. However, the algorithm misses some cycle-skips. This is a recognised issue in surface-wave tomography, as shown by Moulik et al. [2021] who compare a range of published phase-delay datasets. In our synthetic setup, we are able to spot these missed cycle-skips, but we keep them as we aim to mimic a realistic situation where we may be unaware of them. An advantage of our measurement approach is that most cycle-skips that are missed lead to an increased data measurement uncertainty. The real problematic case is when this does not lead to an increased uncertainty, as discussed in the manuscript.
   We have clarified the situation regarding the cycle-skip at L418-425. We also refer the reader to the appendix about the cycle-skip correction, and now cite Moulik et al. [2021] on this recurring issue in surface-wave tomography.

3. We thank Reviewer 2 for the suggestion to use another measurement algorithm. However, it is important to use a measurement algorithm that is tailored to the forward theory we use. For example, in the current setup the sensitivity kernels are derived with the measurement algorithm in mind, and the tapers that we apply to the waveforms are also applied to the sensitivity kernels. Cross-correlation measurements at nearby stations as in Jin and Gaherty [2015] is difficult to apply in our context, where most stations are isolated.
   We may consider comparing our measurements to those obtained with other algorithms in future, but for now we refer the reader to Moulik et al. [2021] who compared several global data sets.

4. We have followed the reviewer's suggestion and plotted the ray-paths coloured by the deviation between the analytical predictions and measurements. There is indeed a spatial pattern that arises, with higher deviations mostly found at ocean-continent boundaries or along ridges, as well as in the central Pacific at lower frequencies. These are probably locations where non-linear effects are stronger.
   These plots are indeed informative, and we have therefore added them to Figure 7, as described in the text at L453-456. This also relates to Point P 2.22 as the locations of the most discrepant measurements correlate well with the locations where the model solution deviates the most from the filtered input.

5. A general comparison between ray-theory and finite-frequency has been conducted by Zhou et al. [2005]. Repeating this would be a study in itself and is therefore outside the scope of our study. In addition, it would only provide limited insights for understanding the discrepancy due to the assumptions that are made in ray-theory. It might be more useful to conduct more synthetic experiments, varying some aspects that we include in our theoretical uncertainties. Again, this is outside the scope of the current study.

Reviewer Point P 2.6 — L524:

1. since many deviations from the input filtered model are small scale features, I wonder if varying the trade-off parameter "eta" could have any effect.
2. In this study I don't see any trade-off curve, or a similar criterion, for the selection of the trade-off value. How did you choose this parameter?
3. Is there a trade-off value that could decrease significantly chi^2? Would that be far from your choice (50)?

Reply:

1. Indeed, small-scale structures are the most difficult to resolve. The better the resolution, or the smaller η, the more likely it is we can resolve them. However, better resolution also means higher uncertainty. Therefore, while we may better recover small-scale structures, these are more likely to be uncertain, i.e. we are still not able to interpret them. We have added a note on this at L585-587.

2. This is legitimate question from the reviewer, teaching us that a fundamental aspect of SOLA may remain unclear: the trade-off parameter η is fundamentally different between data-fitting optimisation problems and in SOLA. In data-fitting optimisations, it controls the smoothness of the model versus the data-fit and it is subjective in that the tomographer decides to what degree they trust their data and want a 'high-resolution' model. In contrast, in the SOLA approach it relates to the design of a local average: for a certain point in space, do we prefer a small, but highly uncertain local average, or a large, but more certain average? This means that compared to data-fitting inversion all choices of η are valid. We therefore do not need to rely on some arbitrary criterion to select one value of η and there is no particular need for an L-curve (see also the supporting information of Zaroli et al. [2017]). In practice, we do experiment with different values of the trade-off parameter and target resolution, but only for a small selection of model

parameters. This is primarily to make sure that the results are robust, rather than to make subjective choice.
We hope that the added appendix on the SOLA approach will help to clarify this fundamental aspect of the method. We have also added a note in the text about this at L315-320.

3. In this case, χ^2 is a model misfit (normalised by model uncertainty), not a data misfit. To avoid any confusion, we have swapped χ for ξ in the text. The trade-off parameter is related to the resolution misfit, not the model misfit. The relationship with that (between resolution and the ξ^2 value) is not trivial because a higher resolution does not always correspond to a lower model misfit. We have clarified this point in paragraph 8 of section 6.2 (L694-700).

Reviewer Point P 2.7 — This is a general comment about the discussion section.

1. I agree that SOLA is an interesting option to consider when compared with damped least squares methods, even if lack of smoothing constraints in SOLA seems to give rise to relatively coarse solution models.
2. However, it remains non-trivial to interpret the robustness of anomalies bringing together resolution and uncertainty propagation (especially in an unambiguous quantitative way).
3. Moreover, the method remains limited to linear(ised) problems, because any sort of iterative linearization would make error propagation non-trivial (due to the iterative perturbative approach).
4. In Bayesian Monte Carlo inference, I agree that prior probability functions have their influence on the posterior density, but i) non-informative priors (like uniforms) can be chosen for the fields directly investigated in the inversion, ii) even if informative priors were chosen, Bayesian inference at the end is all about comparing priors with posteriors, to see where information deviates from prior knowledge. These solvers provide uncertainty estimation directly sampling from the posterior models that fit the data within the uncertainty.
5. Moreover, in the case of the trans-dimensional (t-dim) methods you mention (Bodin and Sambridge, 2009), many parameters that typically need to be selected a priori for the model's parametrization are removed, resulting in a self-parametrized multi-resolution approach. I have not fully understood your claim that t-dim solvers are approximate (??) and provide partial information about the resolution (L563-564), because this is the entire point behind trans-dimensional imaging. If data strongly requires a feature, small or large scale, that will be manifested in many models in the ensemble, resulting in the average and uncertainty maps. I have also some thoughts about your claim that hierarchical methods, where noise is a random variable in the posterior, use a single noise parameter for all data (L681-683). This is a partial consideration because of the following reasons: i) data can be divided in "classes", or different observables, and independent noise parameters can be introduced for each one (Zhang et al., 2018; surface-waves phase and group travel times; Del Piccolo et al. 2024; body-waves P&S delays and splitting intensity). The "classes" of observables can be chosen arbitrarily; hypothetically, I could also separate different ranges of periods for surface waves and invert for independent noise parameters, or different classes based on a "measurement quality metric"... Another option is ii) to use more complicated parametrizations to represent

noise depending on physical quantities, for example introducing polynomial functions (or other basis functions) depending on the wave-paths' ranges (Bodin et al., 2012; Zhang et al., 2020), on the periods, etc... defining the coefficients of the polynomials as posterior's random variables to be inferred during the inversion. In this process, noise enters directly the posterior definition with the possibility to be inferred, but also directly controls model's total uncertainty acting on the width of the likelihood functions.

Reply: We agree that there are advantages and disadvantages to the SOLA method. In the revised manuscript, we have adapted the text to be fair in our comparisons with other methods, as described below.

1. Regarding the lack of smoothness, this is due to the fundamental difference of SOLA with damped data fitting approaches. However, we believe it is not a drawback of the method. One must remember that we are not looking for one single model that satisfies the tomographer (given that the smoothing that other methods apply is arbitrary). We instead aim to obtain individual local average estimates that we can quantitatively interpret with the resolution and uncertainty. It may be misleading that we plot these local averages side by side, which gives the impression that they can be interpreted together as 'model', but we have not yet found a more adequate visualisation. See also our response to Point P 2.6.
2. It is true that interpreting the model solution, resolution, and uncertainty is not trivial, and this is exactly the point made by Latallerie et al. [2022]. Studies that have used SOLA so far have proposed various ways to make use of the model resolution and uncertainty for robust interpretations, but given we work here in a synthetic setup, we prefer to leave this to a future application of the method.
3. We acknowledge that the method remains limited to linear problems, and accounting for non-linearity would be difficult. We do not aim to do otherwise here, but there are possible ways of improvement, following the work of Snieder [1991]. This has also been discussed in response to Point P 1.1.
   We have added a paragraph in the discussion on this at L724-730 where we cite Snieder [1991].
4. We thank the reviewer for pointing this out. We have now clarified the role of priors in Bayesian methods by mentioning that non-informative priors could be used and that prior distributions could also be compared directly with the posterior distributions at L632-633.
5. We thank the reviewer for correcting these statements. We agree that our claim that t-dim was 'approximate' was not justified and we had missed the point that t-dim methods could indeed treat data with different uncertainties.
   We have removed the statements the reviewer refers to and now cite the relevant references [Bodin and Sambridge, 2009, Zhang et al., 2020, Del Piccolo et al., 2024] at L678-683.

Minor

Reviewer Point P 2.8 — L58-61: I would say that the reason why data-misfit based methods do not directly account for resolution is because this metric is influenced by the grid choice (model's parametrization) and smoothing constraints in a non-trivial combination. It is not clear what you mean by "computational reasons". Are you referring to full-waveform inversions? Because otherwise these solvers are typically extremely efficient when working with secondary observables (like travel-times).

Reply: In linear problems, one could approximate the resolution using point spread functions for example, which would be relatively efficient, while bootstrapping could be used to obtain the uncertainties. For large-scale problems, however, it gets expensive (computationally), to obtain the generalised inverse. In general, it is true that these statistics would also be biased by the parameterisation and regularisation, which is not the case with SOLA.
We have changed the text to clarify the above points, see L58-66 in the revised manuscript.

Reviewer Point P 2.9 — L63: what do you mean exactly by "crude approximation" of the resolution?

Reply: We meant to say that the resolution obtained is not always exact, but only an approximation of the resolution.
We agree that this was not very clear, and we have removed this expression, also in response to Point P 2.8.

Reviewer Point P 2.10 — L189-191: does the model's covariance matrix also show non-zero elements outside the diagonal representing covariances between parameters? Would accounting for covariances change significantly the estimated model's uncertainty?

Reply: Yes, the off-diagonal terms of the model covariance matrix can be non-zero. Actually, one row of the model covariance matrix looks very similar to an averaging kernel when plotted, as we are working in a linear setup. In SOLA we do not consider them explicitly because the information they carry is already embedded in the resolution, and this is the point of view we take here.
We have added a note on this in the discussion, see the section about the uncertainty at L707-709.

Reviewer Point P 2.11 — L220-221: it is my understanding that, if we refer to linear problems, matrix G (and the generalized inverse) does not depend on the values of data d (also for "data-fitting" methods). A typical example is a body wave imaging experiment with fixed rays and a grid parametrization for the velocity model; the element Gij would represent the length of the segment of the i-th ray in the j-th voxel, so also the generalized inverse – as a manipulation of G, would not depend on data d (travel times). Probably you were referring to iterative linearized solvers, where G represents the Fréchet derivatives, and it is in general a function on the data d, but this would make the comparison with SOLA (only linear) unfair I suppose. Could you elaborate this point further?

Reply: In this instance, we are discussing linear problems. Indeed, the sensitivity matrix G does not depend on the data, but the generalised inverse does. In fact, we meant that the minimisation problem that we solve to produce a model solution is driven by the data in data-misfit approaches, while this is not the case in SOLA. We agree that this was not very clear. This also relates to Points P 2.21 and P 1.20.
We have changed the text to clarify this at L219, and also made this clear in the figure in the added appendix.

Reviewer Point P 2.12 — L227: please, specify what kind of data δ_{ϕ_l} in eq. (6) represents.

Reply: Now specified at L224-225.

Reviewer Point P 2.13 — L270-271: do you expect that an alternative 3D parametrization in voxels different from 2°x2°x25 km would significantly impact your conclusions about resolution? Or is this strategy robust for even smaller voxels?

Reply: The parameterisation will have no impact on our conclusions about the resolution, but it should be chosen in such a way that the voxels are smaller than the target kernels if these are to be honoured. However, it would have an impact on the theoretical uncertainty as discretisation onto the tomographic grid degrades the accuracy of the sensitivity kernels.
We have added a note on this at L276-279.

Reviewer Point P 2.14 — 2(b): correct title "data sensitivity".

Reply: Amended.

Reviewer Point P 2.15 — L312-315: do you think that a different choice for the target resolution, for example depending on the coverage, could lead to an improved data-misfit with respect to the $chi_2 = 33$ you achieved in this study?

Reply: We should clarify that χ^2 is a model-misfit (normalised by the model uncertainty), not a data-misfit. To avoid this confusion in future, we have replaced the letter χ by the letter ξ in the revised manuscript. While a better resolved model may lead to a lower model misfit, the model uncertainty in the denominator would also likely increase. Therefore, the relationship between ξ^2 and the resolution is not trivial.
We have added a paragraph at L694-700 on this in the discussion on uncertainty.

Reviewer Point P 2.16 — L324: I would think that data uncertainty always (not potentially) influences solution models in multi-observables (like multi-period) inversions.

Reply: Here, we meant to say that the data uncertainty could potentially influence the design of the generalised inverse. As can be seen in the minimisation equation for SOLA (equation 5), the uncertainty can play a role through the data covariance matrix. In our case, since we set all data uncertainty equal to one when obtaining the generalised inverse, the data uncertainty does not influence the minimisation.
We have slightly rephrased the text at L324-328 to clarify this.

Reviewer Point P 2.17 — L338-340: was the L-curve used to select a trade-off parameter value equal to 50 in this study?

Reply: As discussed in our response to Point P 2.6, there is no need for an L-curve for choosing the value of η in the SOLA method. All target resolutions are possible, so long as the model solution is interpreted in light of its resolution and uncertainty. In general, one would use a trade-off that is useful for the object that is to be characterised, depending whether one wants to see a very detailed structure with high uncertainty, or only broad features with low uncertainty. Here, since it is a synthetic study, we have no reason to choose a particular value of η, but adopted one so that the model shows all main tectonic features. While we have not computed an L-curve to select η, we have experimented with various values of η, as well as target resolution sizes and found that they do not alter our conclusions.

Reviewer Point P 2.18 — L389: this is where I suppose something like bootstrapping could produce more informative models in terms of uncertainty.

Reply: See our response to Point P 2.4

Reviewer Point P 2.19 — L419-422: is the undetected cycle-skip at 8mHz responsible for all the discrepancies between analytical predictions and measurements in that 70% fraction of the dataset? It is not completely clear from the manuscript how you justify the deviations of such a significant fraction of the dataset.

Reply: No, this is only the case for this specific example. We agree that this was unclear in the original manuscript.
We have split this paragraph into two paragraphs and rewritten a significant portion of it. These are now the 3rd and 4th to last paragraphs of section 4 (L409-442). The fact that the measurements appear to deviate from the predictions for a large fraction of the dataset has been addressed in the response to Point P 2.5. We discuss there that this is not really the case, but instead due to the arbitrary definition of the classes we used.
As discussed, we have changed the definition of the classes in the revised manuscript (paragraphs 3 and 4 of section 4, L409-442).

Reviewer Point P 2.20 — Looking at figure 7(a), it looks like there is weak tilt in the relationship between the measured and the analytical phase delays. Would applying a rotation (meaning that either measured delays are systematically overestimated or predictions are systematically underestimated) lead to the analytical delays describing a wider range of observations (more than the current 30%)?

Reply: A rotation may improve the agreement between predictions and measurements, but we would be hesitant to apply this without knowing whether this tilt arises from a systematic error in the predictions or from a bias in the measurements. In addition, the tilt is not entirely symmetric and seems to affect positive phase delays more, which may hint at non-linear effects. Importantly, we can spot this apparent tilt since this is a synthetic test, but in order to mimic a real case scenario, we believe it is better to keep the dataset as it is. As discussed in the responses to Points P 2.5 and P 2.19, the agreement is not as bad as it seems (see also the comparison work by Moulik et al. [2021]).

Reviewer Point P 2.21 — L479-482: I wonder if there could be a correlation between the rays corresponding to class III measurements (poor measurement unrealistically low uncertainty) and regions where you recover stronger anomalies with respect to the filtered input. This would also follow up the comment above about the possibility of predictions being systematically underestimated due to the local average of resolved kernels. Would this mean that resolved kernels are too wide? Or the trade-off parameter not appropriate?.

Reply: Investigating whether cells with a strong difference between filtered input and output are mostly constrained by particular, e.g. class III measurements, is not trivial. With SOLA, it is possible to investigate this using the generalised inverse (one row of the generalised inverse contains the contribution of each data to a specific cell). However, it remains not trivial to visualise and analyse the information that this provides.

Nevertheless, as noted in our responses to Points P 2.5 and Point P 2.22, the locations of the most discrepant ray-paths correlate well with the regions where the model solution deviates the most from the filtered input model. We thank Reviewer 2 for suggesting to plot the ray paths and model differences as they have helped to observe this correlation.

We do not understand the second part of this comment. There is no relationship between the resolution and the predictions, which are computed using the forward theory in the input model. In addition, the averaging kernels are never too wide or too small, nor the trade-off parameter not appropriate. Any combination of these is valid, so long as the model solution is interpreted together with its resolution and uncertainty, as discussed further in the response to Point P 2.6.

Reviewer Point P 2.22 — Fig. 8: I would add a map showing the deviation (filtered input – model solution). It can be difficult to directly compare the two solution maps. Less important and not necessary to add, but I would really like to see what the input model looks like when applying a filter with the target kernels.

Reply: We agree that a map showing the deviations of the model solution to the filtered input model would be useful.
We have added this to Figure 8 of the revised manuscript, together with a map where these deviations are normalised by the model uncertainty. This suggestion has also helped to observe that the most discrepant regions correlate with the location of the most discrepant measurements, which relates to Point P 2.5 and P 2.21.

The role of the target resolution is to design the expected actual resolution, but once the actual resolution is obtained, one should forget about the target. However, we understand that Reviewer 2 may be curious about this information, which we have therefore computed. As Figure 1 shows, the input model filtered by the target resolution looks very similar to the input model itself. This is because the target kernels are very small, almost Dirac deltas. We have decided not to include it in the revised paper as we think it would add confusion.

Reviewer Point P 2.23 — L563-568: as mentioned in the main comments, some words here like “approximate” and “partial” probably needs some additional context. Moreover, some recent studies using Monte Carlo inference adopt a single-step scheme for the inversion of surface waves (Zhang et al., 2018; Zhang et al., 2020), providing more reasonable estimates for uncertainty compared to the 2-steps inversions.

Reply: The first part of this comment has been addressed in responses to earlier comments, e.g. Point P 2.7 and point P 2.9 and these phrases have been removed from the revised manuscript.

We appear to have missed these applications of 3D transdimensional tomography. We thank reviewer 2 for pointing us to these studies, and we have added a note on this in the revised manuscript, citing them at L568.

Reviewer Point P 2.24 — L578-579: I agree that no explicit global constraints are applied to the model’s parameters. However, it is my opinion that limiting the model’s covariance acts similarly to an implicit “smoothing” constraint on the parameters, since a smooth model is typically expected to be more “stable” than a model with very rough (i.e., small-scale) features. I would expect that

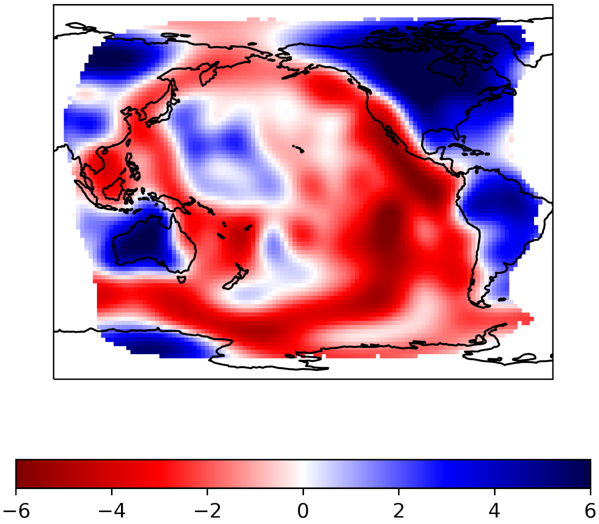


Figure 1: Input model filtered by the target resolution. The colorscale gives dlnVs in %.

by changing the trade-off parameter, and promoting lower covariance models, they would also look smoother than the one in this study (because of the increasing size of the resolving kernels).

Reply: There are several factors that affect the smoothness of the model. Regarding the covariance, this would be true if we were limiting the off-diagonal terms of the model covariance matrix. However, this is not the case, we only limit the covariance of model parameters with themselves, and therefore no implicit smoothing is implied here. However, it is true that the overlap of averaging kernels for nearby model parameters guarantees some spatial coherency in the model. In addition, some spatial coherency in the model arises due to the data sensitivity itself. To take this comment into account, we have added a sentence at L581-582 on the role of the overlap of the averaging kernels in providing some spatial coherency.

Note that we are not aiming to obtain a smooth model with SOLA. Instead, we tend to search for models where the resolution is as small as possible, and therefore the objective is almost opposite to that of looking for a smooth model.

Reviewer Point P 2.25 — L569-570: I suppose the large-scale features are also the ones on which almost all the surface waves tomography models agree on. . . (?)

Reply: Yes. We have clarified this at L570-571.

Reviewer Point P 2.26 — L599-606: this point is extremely interesting. I suppose that this would manifest as well in teleseismic imaging with body waves using SOLA algorithm. Anomalies tend to be smeared up to the surface, meaning the shallow parameters' kernels are likely to represent the average of deeper sections. Again, interesting how the trade-off parameter could control these biases, but probably there is not an intuitive way to tune it for this purpose.

Reply: Indeed, this effect has also been seen in body-wave tomography with SOLA (e.g. Freissler 2024). With SOLA, we have some control over the resolution using the target kernels and thus we

would hope to remove this effect. However, we remain limited by our choice of data, so in practice we we can only reduce the effect. In our case, we have tried various trade-off parameters and target kernels to avoid this depth leakage, but the results presented in the paper are the best we could achieve with the data set. This is primarily due to the fact that we only use fundamental mode measurements for now, and so only have good sensitivity at shallow depth. There may be ways of improvement, for example by changing the trade-off parameter with depth or by including data weighs in the inversion (for example by using values other than one in the data covariance matrix or including a weighing matrix, to give more weight to low frequency data). These are all ways of improvement to investigate in future.
To take this comment into account, we mention some of the ideas above in the text at L611-614.

We thank the editor and reviewers again for their comments which have improved and clarified many aspects of our work. We hope our responses and the changes made to the revised manuscript address their concerns.

References

T. Bodin and M. Sambridge. Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, 178(3):1411–1436, Sept. 2009. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2009.04226.x. URL `https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.2009.04226.x`.

G. Del Piccolo, B. P. VanderBeek, M. Faccenda, A. Morelli, and J. S. Byrnes. Imaging Upper-Mantle Anisotropy with Transdimensional Bayesian Monte Carlo Sampling. *Bulletin of the Seismological Society of America*, 114(3):1214–1226, June 2024. ISSN 0037-1106, 1943-3573. doi: 10.1785/0120230233. URL `https://pubs.geoscienceworld.org/bssa/article/114/3/1214/633693/Imaging-Upper-Mantle-Anisotro`

G. Jin and J. B. Gaherty. Surface wave phase-velocity tomography based on multichannel cross-correlation. *Geophysical Journal International*, 201(3):1383–1398, June 2015. ISSN 1365-246X, 0956-540X. doi: 10.1093/gji/ggv079. URL `http://academic.oup.com/gji/article/201/3/1383/758678/Surface-wave-phasevelocity-tomography-b`

F. Latallerie, C. Zaroli, S. Lambotte, and A. Maggi. Analysis of tomographic models using resolution and uncertainties: a surface wave example from the Pacific. *Geophysical Journal International*, 230(2):893–907, Apr. 2022. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggac095. URL `https://academic.oup.com/gji/article/230/2/893/6544670`.

P. Moulik, V. Lekic, B. Romanowicz, Z. Ma, A. Schaeffer, T. Ho, E. Beucler, E. Debayle, A. Deuss, S. Durand, G. Ekström, S. Lebedev, G. Masters, K. Priestley, J. Ritsema, K. Sigloch, J. Trampert, and A. M. Dziewonski. Global reference seismological data sets: multimode surface wave dispersion. *Geophysical Journal International*, 228(3): 1808–1849, Dec. 2021. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggab418. URL `https://academic.oup.com/gji/article/228/3/1808/6408466`.

R. Snieder. An extension of Backus-Gilbert theory to nonlinear inverse problems. *Inverse Problems*, 7(3):409–433, June 1991. ISSN 0266-5611, 1361-6420. doi: 10.1088/0266-5611/7/3/008. URL `https://iopscience.iop.org/article/10.1088/0266-5611/7/3/008`.

C. Zaroli, P. Koelemeijer, and S. Lambotte. Toward Seeing the Earth's Interior Through Unbiased Tomographic Lenses. *Geophysical Research Letters*, 44(22):11,399–11,408, Nov. 2017. ISSN 0094-8276, 1944-8007. doi: 10.1002/2017GL074996. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074996`.

X. Zhang, C. Roy, A. Curtis, A. Nowacki, and B. Baptie. Imaging the subsurface using induced seismicity and ambient noise: 3-D tomographic Monte Carlo joint inversion of earthquake body wave traveltimes and surface wave dispersion. *Geophysical Journal International*, 222(3):1639–1655, Sept. 2020. ISSN 0956-540X, 1365-246X. doi: 10.1093/gji/ggaa230. URL `https://academic.oup.com/gji/article/222/3/1639/5835227`.

Y. Zhou, F. A. Dahlen, G. Nolet, and G. Laske. Finite-frequency effects in global surface-wave tomography. *Geophysical Journal International*, 163(3):1087–1111, Dec. 2005. ISSN 0956540X, 1365246X. doi: 10.1111/j.1365-246X.2005.02780.x. URL `https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.2005.02780.x`.

Non-peer reviewed manuscript submitted to SEISMICA

Towards surface-wave tomography with 3D resolution and uncertainty

Franck Latallerie * [1], **Christophe Zaroli** [2], **Sophie Lambotte** [2], **Alessia Maggi** [2], **Andrew Walker** [1], **Paula Koelemeijer** [1]

[1]Department of Earth Sciences, University of Oxford, Oxford, United Kingdom, [2]Institut Terre et Environement de Strasbourg, UMR7063, Université de Strasbourg, EOST/CNRS, 67084, Strasbourg CEDEX, France

Author contributions: *Conceptualization*: FL, CZ, SL, AM, AW, PK. *Methodology*: FL, CZ, SL. *Software*: FL, CZ, SL. *Data curation*: FL. *Formal Analysis*: FL. *Investigation*: FL, CZ, PK. *Writing - Original draft*: FL. *Writing - Review & Editing*: FL, CZ, SL, AM, AW, PK. *Visualization*: FL. *Supervision*: CZ, SL, AM, AW, PK. *Project administration*: AW, PK. *Funding acquisition*: AM, AW, PK.

Abstract Surface-wave tomography is crucial for mapping upper-mantle structure in poorly instrumented regions such as the oceans. However, data sparsity and errors lead to tomographic models with complex resolution and uncertainty, which can impede meaningful physical interpretations. Accounting for the full 3D resolution and robustly estimating model uncertainty remains challenging in surface-wave tomography. Here, we propose an approach to provide direct control over the model resolution and uncertainty and to produce these in a fully three-dimensional framework by combining the Backus-Gilbert-based SOLA method with finite-frequency theory. Using a synthetic setup, we demonstrate the reliability of our approach and illustrate the artefacts arising in surface-wave tomography due to limited resolution. We also indicate how our synthetic setup enables us to discuss the theoretical model uncertainty (arising due to assumptions in the forward theory), which is often overlooked due to the difficulty in assessing it. We show that the theoretical uncertainty components may be much larger than the measurement uncertainty, thus dominating the overall uncertainty. Our study paves the way for more robust and quantitative interpretations in surface-wave tomography.

Non-technical summary In the oceans, several surface features such as isolated volcanic islands or variations in the depth of the seafloor result from dynamic processes in the underlying mantle. To understand these processes, we need to image the three-dimensional structures present in the subsurface. While long-period surface waves can be used for this, the data are typically noisy and provide poor coverage of the oceans. This limits the quality of our images and therefore the interpretations that can be drawn from them. In addition, limitations of our images are difficult to quantify with current methods, which makes interpretations even more difficult.

---

*Corresponding author: franck.latallerie@earth.ox.ac.uk

In this study, we propose an approach to produce high-quality maps of 3D structures in the upper mantle, which also gives information on the quality of the images. We present the method in a synthetic framework, which serves to demonstrate our ability to retrieve an input Earth model and enables us to estimate theoretical model uncertainties. Our approach will enable more robust interpretations of surface-wave tomography models in the future.

1 Introduction

Many important geological processes (e.g. melting at mid-ocean ridges, spreading, subduction and hotspot volcanism) result from dynamic processes in the upper mantle. To improve our understanding of these processes, we need to robustly image the structure of the upper mantle. In poorly instrumented regions, such as the oceans, this imaging relies heavily on surface-wave tomography. However, surface-wave data have poor spatial coverage, both laterally due to the uneven distribution of earthquakes (sources) and seismic stations (receivers), and vertically due to how surface-wave sensitivity varies with depth. Surface-wave data also contain errors due to imperfect measurement and physical theory. Poor data coverage and data errors render the inverse problem ill-posed and lead to complex model resolution and model uncertainty (e.g. Parker, 1977; Menke, 1989; Tarantola, 2005). These explain the strong discrepancies between published tomography models (e.g. Hosseini et al., 2018; Marignier et al., 2020; De Viron et al., 2021). Over time, seismic images have become more detailed and are being used to inform research in other fields. To guarantee the usefulness of surface-wave tomographic images however, we need to account for their full 3D resolution and uncertainty (e.g. Ritsema et al., 2004; Foulger et al., 2013; Rawlinson et al., 2014). Equipped with these, we will be able to avoid interpreting non-significant anomalies (e.g. Latallerie et al., 2022), set up meaningful comparisons with theoretical predictions (e.g. Freissler et al., 2020), or include tomography models in further studies such as earthquake hazard assessments (e.g. Boaga et al., 2011; Socco et al., 2012; Boaga et al., 2012).

Many approaches have been proposed to solve ill-posed inverse problems in seismology (e.g. Wiggins, 1972; Parker, 1977; Tarantola and Valette, 1982; Nolet, 1985; Scales and Snieder, 1997; Trampert, 1998; Nolet, 2008). Most take a data-misfit point of view and search for a model whose predictions are 'close enough' to observations. However, such approaches have difficulties in accounting directly for model resolution and uncertainty, either for computational reasons or because, in these approaches, resolution and uncertainty depend in complex ways on the parameterisation and regularisations used (Nolet et al., 1999; Barmin et al., 2001; Ritsema et al., 2004; Shapiro et al., 2005; Ritsema et al., 2007; Fichtner and Trampert, 2011; An, 2012; Fichtner and Zunino, 2019; Simmons et al., 2019; Bonadio et al., 2021). Synthetic tests, sometimes in the form of checkerboard tests, can be useful to assess resolution, but these have been shown to be potentially misleading (e.g. Lévêque et al., 1993; Rawlinson and Spakman, 2016).

Other approaches for solving ill-posed inverse problems move away from the data-misfit point of view and instead concentrate on directly optimising model resolution and uncertainty. These approaches are typically based on Backus–Gilbert theory (Backus and Gilbert, 1967, 1968, 1970). One such approach, the SOLA (Subtractive Optimally Localized Averages) formulation, was derived for helioseismology by Pijpers and Thompson (1992, 1994) before being introduced and adapted to linear body-wave tomographic inversions by Zaroli (2016) and Zaroli (2019). Besides body waves, the method has been successfully applied to normal-mode splitting data to constrain ratios between seismic velocities (Restelli et al., 2024) and to surface-wave dispersion data to build group-velocity maps (Ouattara et al., 2019; Amiri et al., 2023) or 2D maps of the vertically polarised shear-wave velocity V_{SV} (Latallerie et al., 2022). Although SOLA can be applied only to linear problems, it requires no prior on the model solution, provides direct control on model resolution and uncertainty, and produces solutions free of averaging bias (Zaroli et al., 2017).

Traditionally, surface-wave tomography studies are based on ray-theory. This infinite-frequency approximation requires a two-step procedure that can be performed in either order. One order is first to solve the inverse problem laterally (to produce 2D phase or group-velocity maps) and subsequently to solve for velocity structure with depth (to produce 1D velocity profiles) (e.g. Ekström et al., 1997; Montagner, 2002; Yoshizawa and Kennett, 2004; Ekström, 2011; Ouattara et al., 2019; Seredkina, 2019; Isse et al., 2019; Magrini et al., 2022; Greenfield et al., 2022). The other approach is to solve first for velocity structure with depth for independent source-receiver pairs (to produce 1D path-averaged velocity profiles) and subsequently for lateral variations (to produce 2D velocity maps) (e.g. Debayle and Lévêque, 1997; Lévêque et al., 1998; Debayle, 1999; Debayle and Kennett, 2000; Simons et al., 2002; Lebedev and Nolet, 2003; Priestley, 2003; Debayle and Sambridge, 2004; Maggi et al., 2006b,a; Priestley and Mckenzie, 2006). This second approach was adopted by Latallerie et al. (2022) who applied the SOLA method to the second step (lateral inversion) to produce 2D lateral resolution and uncertainty information, in addition to their tomography model. Because the first step is a non-linear depth inversion, it could not be performed using SOLA – a purely linear method. Therefore, this study was not able to provide high-quality information about vertical resolution, a significant drawback given the complex depth sensitivity of surface-waves.

In this study, we extend the approach of Latallerie et al. (2022) to 3D using the framework of finite-frequency theory (Snieder, 1986; Snieder and Nolet, 1987; Yomogida, 1992; Marquering et al., 1998; Dahlen and Tromp, 1999; Yoshizawa and Kennett, 2004; Zhou et al., 2004, 2005; Yoshizawa and Kennett, 2005; Zhou, 2009a,b; Ruan and Zhou, 2010; Tian et al., 2011; Zhou et al., 2006; Liu and Zhou, 2016b,a). In this framework, surface-wave dispersion data are linearly related to perturbations in the 3D upper-mantle velocity structure. This makes it possible to perform a one-step inversion and thus to obtain 3D resolution information using SOLA. Finite-frequency inversions come with higher memory costs because the sensitivity kernels are volumetric (with both a lateral and depth extent) and the whole 3D model must be stored all at once (large number of model parameters). However, with smart data selection and ever increasing computational power, this memory cost is becoming less of an issue.

Model uncertainty arises from data uncertainty (or measurement uncertainty) as well as theoretical uncertainty. Data uncertainty is often estimated by comparing the dispersion of measurements for nearby rays (e.g. Maggi et al., 2006b). However, this approach dramatically underestimates the data uncertainty and accounts poorly for systematic biases (e.g. Latallerie et al., 2022). This is less of an issue if we are only interested in the relative uncertainty between individual data (e.g. when we weigh data contributions in a data-driven inversion). Underestimated data uncertainty and bias become problematic, however, if we want to interpret the ‘true’ magnitude of the model uncertainty. It therefore becomes important to estimate data uncertainties carefully. Additionally, we need to account for imperfections in the forward theory, which give rise to ‘theoretical uncertainty’. This theoretical uncertainty arises from a range of approximations commonly made: single-scattering, which relates to non-linearity; the forward-scattering approximation; the paraxial approximation; neglected sensitivity to other parameters; discretisation onto the tomographic grid; linear crustal correction strategy; errors in the crustal model; and errors in the earthquake source parameters. These last two contributions are not accounted for in this study. The theoretical component is often missing in model uncertainty estimates, which may partly explain why these estimates appear to be small. Importantly, both measurement uncertainty and theoretical uncertainty contribute to model uncertainty. Here, we distinguish the two contributions to the model uncertainty by using the terms 'measurement model uncertainty' and 'theoretical model uncertainty'. We take advantage of the synthetic nature of this study to discuss the contribution of both contributions.

In this study, we show that it is possible to obtain detailed 3D resolution and robust uncertainty information using surface waves with SOLA within a finite-frequency framework, thus extending the approach of Latallerie et al. (2022) to 3D. By working in a synthetic setup, we demonstrate the feasibility of our approach, and discuss the contribution of theoretical errors. To achieve these aims, we develop a complete workflow from dispersion measurements on the waveforms to analyses of the resulting 3D model, its resolution and uncertainty. In Section 2, we introduce the SOLA method and the forward modelling approach. In Section 3, we describe the tomography setup, including the data geometry, target resolution and generalised inverse. Subsequently, in Section 4, we discuss the data and their uncertainty in detail. In Section 5, we present our tomographic results, both qualitatively and quantitatively. Finally, in Section 6, we discuss the 3D resolution and uncertainty estimates of our model and indicate possible future directions.

2 Theory

We present here the main building blocks of our approach. Firstly, we briefly introduce the general forward problem. We then discuss the inverse problem, introducing the discrete linear SOLA inverse method (Zaroli, 2016) that provides control on the resolution and the propagation of uncertainty, and produces the tomographic model with full resolution and uncertainty information. Finally, we present the finite-frequency theory that allows the surface-wave inverse problem to be expressed in a linear and fully three-dimensional framework.

2.1 General forward theory

Let $\boldsymbol{d} \in \mathcal{R}^N$ be a data vector and let $\boldsymbol{m} \in \mathcal{R}^M$ be a model vector containing model parameters given a pre-defined parameterisation. Let $\boldsymbol{G} \in \mathcal{M}(N \times M)$ be the sensitivity matrix (in the set of matrices of size $N \times M$), describing a linear relationship between model parameters and data. We can then write the forward problem as:

$$\boldsymbol{d} = \boldsymbol{G}\boldsymbol{m}. \tag{1}$$

Rows of $\boldsymbol{G}$ are the sensitivity kernels and $\boldsymbol{G}$ thus contains all the information regarding the sensitivity of the entire dataset to all model parameters; this is what we refer to as the data geometry.

To account for data errors, we treat $\boldsymbol{d}$ as a normally distributed multi-variate random variable with data covariance matrix $\boldsymbol{C_d} \in \mathcal{M}(N \times N)$. We assume uncorrelated noise, thus the data covariance matrix is diagonal and we can write $\boldsymbol{C_d} = \mathrm{diag}(\sigma^2_{d_i}), i \in [|1, N|]$, where σ_{d_i} is the standard deviation of the error on the i^{th} datum. Throughout this study, we refer to the standard deviation as the data uncertainty. Note that under the Gaussian hypothesis both theoretical errors (due to imperfect forward theory) and measurement errors (due to imperfect measurements) can be included in $\sigma_{d_i}^2$ (see e.g. Tarantola, 2005). As it is challenging to estimate error correlations, we assume uncorrelated errors, which we further assume to be Gaussian for mathematical simplicity. The assumption of Gaussian uncorrelated errors remains an important limitation that should motivate future work.

2.2 SOLA inverse method

Poor data geometry in seismic tomography makes the inverse problem ill-constrained as the sensitivity matrix $\boldsymbol{G}$ is not invertible. This justifies the use of various existing methods for obtaining model solutions (see e.g. Parker, 1977; Trampert, 1998; Scales and Snieder, 1997; Nolet, 1985; Tarantola and Valette, 1982; Wiggins, 1972; Nolet, 2008). Most of these methods use a data-misfit approach, where a model solution is found by minimising the discrepancy between predicted data and the actual data. With SOLA, we do not use a data-misfit to drive towards a model solution, but instead focus on designing a 'generalised inverse' of the sensitivity matrix $\boldsymbol{G}$. We describe the SOLA method briefly below, with more details in Appendix A.

Let $\boldsymbol{G}^\dagger$ be the 'generalised inverse' such that the model solution is expressed as linear combinations of the data:

$$\widetilde{\boldsymbol{m}} = \boldsymbol{G}^\dagger \boldsymbol{d}. \tag{2}$$

Using Equation 1, we obtain a relation between the model solution and the 'true' model:

$$\widetilde{\boldsymbol{m}} = \boldsymbol{G}^\dagger \boldsymbol{G} \boldsymbol{m}. \tag{3}$$

Each parameter in the model solution is a linear combination of the 'true' model parameters linked by the resolution matrix $\boldsymbol{R} = \boldsymbol{G}^\dagger \boldsymbol{G}$. In other words, the value of a model parameter in the model solution represents a spatial weighted average of the whole 'true' model (plus some errors propagated from data noise). The resolution for a model parameter is determined by this averaging and is referred to as 'resolving' or 'averaging kernel'. In general, we prefer the averaging for a model parameter to be focused around that parameter location. The full resolution matrix thus acts as a 'tomographic filter' (e.g. Ritsema et al., 2007; Schuberth et al., 2009; Zaroli et al., 2017). Note that in the hypothetical case where the data geometry constrains all model parameters perfectly, the sensitivity matrix $\boldsymbol{G}$ is invertible, the generalised inverse $\boldsymbol{G}^\dagger$ is the exact inverse, the resolution matrix is the identity matrix, and, in the case of error-free data, the model solution is exactly the 'true' model.

Since $\widetilde{\boldsymbol{m}} = \boldsymbol{G}^\dagger \boldsymbol{d}$ is a linear mapping of a multivariate normal distribution, we obtain the model covariance matrix from the data covariance matrix using:

$$\boldsymbol{C}_{\widetilde{\boldsymbol{m}}} = (\boldsymbol{G}^\dagger) \boldsymbol{C}_{\boldsymbol{d}} {\boldsymbol{G}^\dagger}^T, \tag{4}$$

where ^{T} denotes the matrix transpose. The diagonal elements of the model covariance matrix are the standard deviations of the model parameters, i.e. $\sigma_{\widetilde{m}^{(k)}} = \sqrt{C_{\widetilde{m}_{kk}}}$. Analogue to the data uncertainty, we refer to the model standard deviations as the model uncertainty. Note that model uncertainties are thus given for local average estimates, not estimates at absolute points in space. In summary, the generalised inverse $\boldsymbol{G}^\dagger$ determines the model solution, model resolution, and model uncertainty.

While data-misfit approaches have many advantages (e.g. treatment of non-linearity, computational efficiency), they do not directly control the resolution and uncertainty of the solution; estimating this information can be challenging depending on the inverse method used. With the SOLA method, which is based on Backus-Gilbert theory (Backus and Gilbert, 1967, 1968, 1970; Pijpers and Thompson, 1992, 1994; Zaroli, 2016), we explicitly design $\boldsymbol{G}^{\dagger}$ to achieve certain objectives regarding the resolution and model uncertainty. In particular, we design a target resolution $\boldsymbol{T}$ and seek a generalised inverse that leads to a resolution as close as possible to the target, while minimising model uncertainty. These are two contradictory objectives that are balanced in an optimisation problem:

$$\boldsymbol{G}^{\dagger(\boldsymbol{k})} = \arg \min_{\boldsymbol{G}^{\dagger(\boldsymbol{k})}} \sum_j [A_j^{(k)} - T_j^{(k)}]^2 \mathcal{V}_j + {\eta^{(k)}}^2 \sigma^2_{\widetilde{m}^{(k)}}, \quad \text{s.t.} \sum_j R_j^{(k)} = 1, \tag{5}$$

where k is the index of the model parameter for which we are solving (the target), j is a dummy index that iterates over model parameters, $\mathcal{V}_j$ is the volume of cell j, $A_j^{(k)} = R_j^{(k)}/\mathcal{V}_j$ is the averaging (or resolving) kernel (normalised by the cell volume), and $\eta^{(k)}$ is a trade-off parameter that balances the fit to the target resolution with the minimisation of model uncertainty. The constraint $\sum_j R_j^{(k)} = 1$ guarantees that local averages are unbiased. This is important because an uneven data distribution can artificially increase or decrease the value of the estimated parameters, as demonstrated by Zaroli et al. (2017). The optimisation problem leads to a set of equations (see Appendix A1 from Zaroli, 2016) that we solve for each model parameter using the LSQR algorithm of Paige and Saunders (1982), as suggested by Nolet (1985).

The SOLA inversion is point-wise, i.e. the minimisation problem is solved for each parameter independently from the others. This makes SOLA inversions straightforward to solve in parallel. Note that we do not need to solve for all model parameters nor do we need to solve for the whole region to which the data are sensitive (a necessity in data-fitting inversions): we have the possibility to solve only for model parameters of particular interest (the targets). Furthermore, note that the data $\boldsymbol{d}$ do not appear in the optimisation equation 5. We provide information on the computational costs of this study in Appendix C.

2.3 Finite-frequency forward theory

In order to make the implementation of SOLA for surface-wave tomography fully three-dimensional, we need a linear relation between surface-wave data and 3D physical properties of the Earth mantle. Here, we consider as data vertical-component Rayleigh-wave phase delays $\delta\phi_l(\omega)$ measured at frequencies ω for particular source-receiver pairs l. If we assume these delays are primarily sensitive to perturbations in the vertically polarized S-wave velocity δV_{SV} in the 3D mantle $\bigoplus$, we have the following relationship between data $\delta\phi_l(\omega)$ and model $\delta \ln V_{SV}(\boldsymbol{x})$:

$$\delta\phi_l(\omega) = \iiint_{\bigoplus} K_l(\omega; \boldsymbol{x}) \delta \ln V_{SV}(\boldsymbol{x}) d^3\boldsymbol{x}, \tag{6}$$

where $\boldsymbol{x}$ indicates the location, and $K_l(\omega; \boldsymbol{x})$ is the sensitivity kernel. We neglect the sensitivity to other physical parameters (e.g. V_{SH}, V_{PV}, density), but this contributes to the theoretical errors.

Analytical expressions of surface-wave sensitivity kernels have been derived based on the scattering principle in

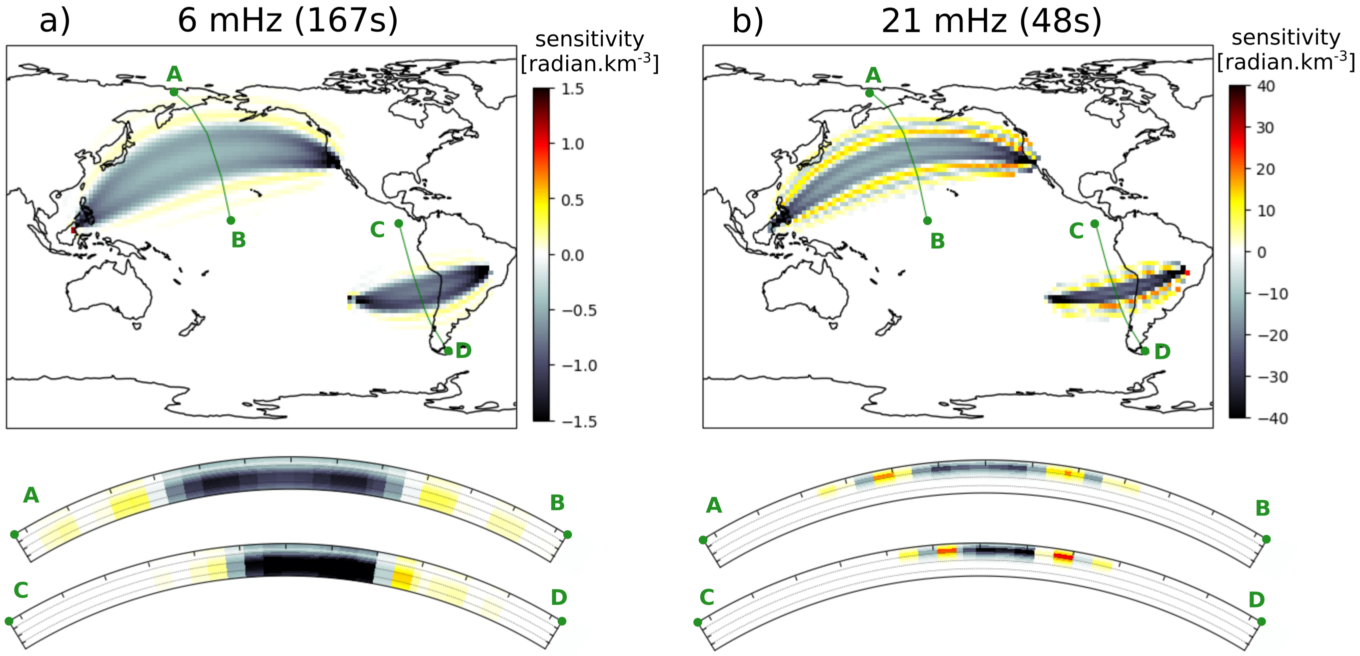


Figure 1 Examples of sensitivity kernels at a) 6 mHz and b) 21 mHz for two source-receiver pairs. The maps are plotted at depths of 87 km and 237 km depth respectively, which are the depths where the kernels reach their respective maximum amplitudes. Below each map, we also show a vertical cross-section through each kernel, as indicated on the maps, and the dotted lines indicate depths of 100, 200 and 300 km. The northern kernel is for a M_w 6.1 earthquake in Borneo (2015) recorded by station DSN5. The southern kernel is for a M_w 6.1 earthquake in the Easter Island region (2011) recorded by station BDFB. Note the difference in amplitude between the two frequencies shown in a) and b).

the framework of normal mode theory. Here, we use formulations from Zhou et al. (2004), later extended to multi-mode surface waves and anisotropy by Zhou (2009b). These assume far-field propagation, single forward scattering, and use a paraxial approximation. Thanks to the single-scattering assumption, also known as Born approximation, the resulting relationship between data and model is linear, which makes it tractable with SOLA. Single-scattering is equivalent to neglecting terms of order higher than 1 in the Taylor expansion of the Green tensor perturbations with respect to structural parameters (e.g. Dahlen, 2000). This single-scattering approximation also contributes to the theoretical errors. In this study, we restrict ourselves to fundamental modes, but extension of the theory to overtones is straightforward. The sensitivity kernels for the fundamental modes can be expressed as:

$$K(\omega;\boldsymbol{x}) = \mathrm{Im}\left(\frac{S'\,\Omega\,R'' e^{-i[k'\Delta'+k''\Delta''-k\Delta+(s'+s''-s)\frac{\pi}{2}+\frac{\pi}{4}]}}{S\,R\sqrt{8\pi(\frac{k'k''}{k})(\frac{sin|\Delta'||sin|\Delta''|}{|sin\Delta|}}}\right). \tag{7}$$

Symbols with prime $'$ refer to the source-scatterer path, ones with double prime $''$ to the scatterer-station path, and those without prime to the great-circle source-station path; k is the wave-number and s the Maslov index (here $s = 0$ or $s = 1$, i.e. single orbit); Δ is the path length, S the source radiation in the direction of the path, and R the projection of the polarisation onto the receiver orientation. The exponent term indicates the phase delay due to the detour by the scatterer, while the other terms express the relative amplitude of the scattered wave relative to the initial unperturbed wavefield. This relative strength depends on the source and receiver terms (the scattered wave leaves the source and arrives at the receiver with some angle compared to the unperturbed wave), on the geometrical spreading (the scattered wave makes a detour compared to the unperturbed wave), and on the scattering coefficient

Ω. The scattering coefficient depends linearly on physical model properties, for which detailed expressions can be found in Zhou (2009a). In practice, we use a slightly different form of Equation 7 to include the effect of waveform tapering in the measurement algorithm (see Zhou et al., 2004, for more details).

We use routines from Zhou (2009b) to compute the sensitivity kernels for the fundamental mode, assuming self-coupling. We only compute these in the top 400 km of the mantle as their amplitude decreases sharply with depth. We consider the first two Fresnel zones laterally as their side-lobes become negligible further away. Examples of sensitivity kernels are given in Figure 1, where they are projected onto the tomographic grid. The kernels have particularly strong amplitude at the source and station. This is caused by a combination of natural high sensitivity at end-points of a path and the far-field approximation (e.g. Liu and Zhou, 2016b). Low-frequency kernels peak at deeper depths, have a broader lateral and vertical extent, and have weaker amplitudes than high-frequency kernels. Although the projection onto the tomographic grid degrades the shape and amplitude of the sensitivity kernels, their main properties are retained on a tomographic grid that is sufficiently fine.

3 Tomography setup

In this section, we present the construction of the forward problem (the sensitivity matrix) and the inverse solution (the generalised inverse) that determines the resolution, the propagation of data uncertainty into model uncertainty, and the propagation of data values into model estimates. We will describe the data and data uncertainty in the next section. These will feed into the inverse solution to produce the tomography model and the measurement model uncertainty.

3.1 Parameterisation

We use a local model parameterisation and split the 3D spatial domain into voxels of size $2^\circ \times 2^\circ$ laterally (latitude and longitude) and 25 km depth vertically. We parameterise the whole sphere laterally, but only the top 400 km depth, since the sensitivity of fundamental mode surface waves to V_{SV} becomes negligible at greater depths. This leads to $M = 259\,200$ voxels. It is worth recalling that with SOLA we do not need to solve for all M model parameters nor for the whole region to which the data are sensitive. For example, we could solve only for cells where the data sensitivity is sufficiently high or only for a particular region of interest. Note that the parameterisation does not impact the SOLA inversion in the same way as in data-fitting approaches. Primarily, the parameterisation should be chosen finer than the target kernels if these are to be honoured. However, the parameterisation is expected to have an impact on the theoretical uncertainty, as the discretisation of the sensitivity kernels degrades the accuracy of the forward theory.

3.2 Data geometry

We select 312 earthquakes with M_w between ~6.0 and 7.7 and depth between ~12 and 87 km, all located in the Pacific region, occurring between July 2004 and December 2020. We consider 1228 stations, also located in the Pacific region (see Fig. 2). Sources and stations are both selected in a way to avoid strong spatial redundancy. For all paths, we consider 16 frequencies ranging from 6 to 21 mHz (48-167s), in steps of 1 mHz.

Compared to ray-theory, finite-frequency theory is fully three-dimensional. This makes the sensitivity matrix larger because we need to consider the whole 3D spatial extent of the model domain all at once, and less sparse

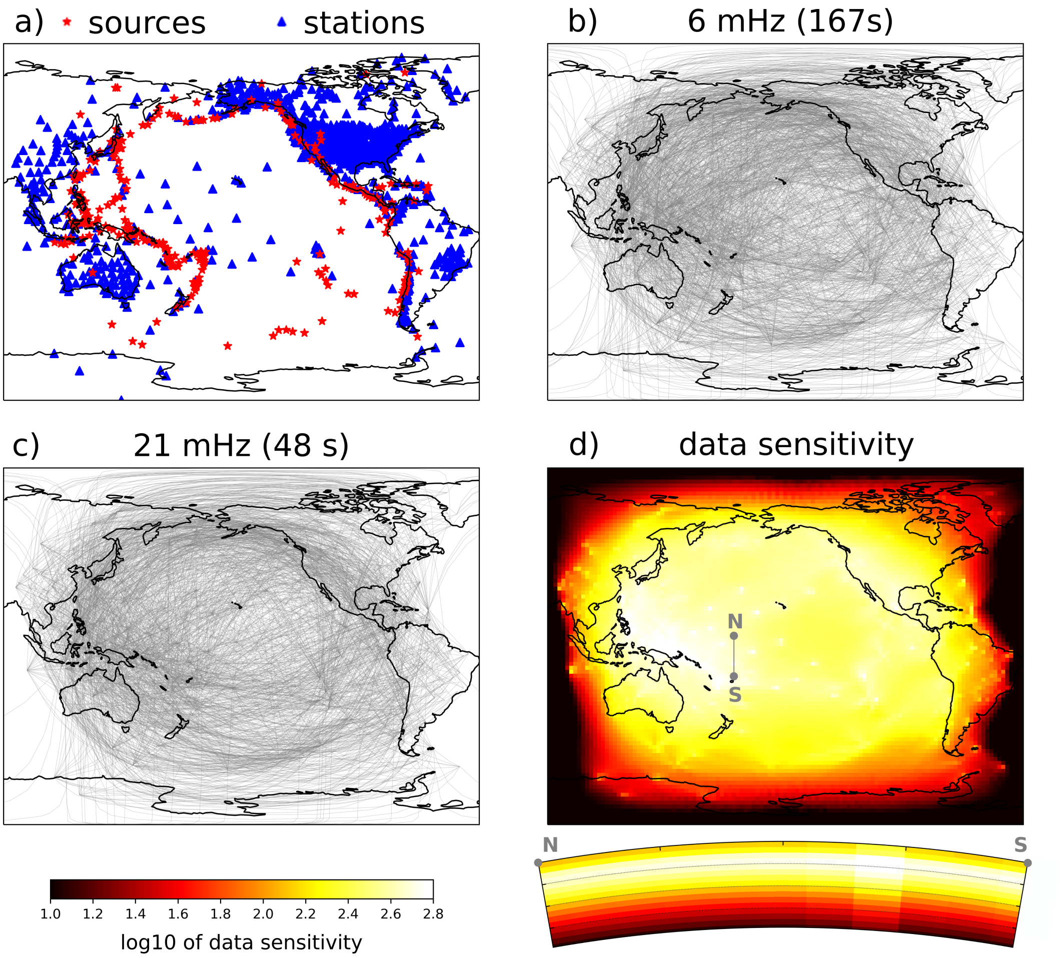


Figure 2 Data geometry of our tomography, showing a) the distribution of sources and receivers, b) the selected ray paths at 6 mHz and c) at 21 mHz, and d) the decimal logarithm of the data sensitivity, $\log_{10} \sum_i |G_{ij}|$. The data sensitivity is plotted at 112 km depth, with a N-S oriented vertical cross-section below it, indicated by the grey line on the map view, and the dotted lines indicate depths of 100, 200 and 300 km.

because finite-frequency sensitivity kernels have a volumetric extent. Since we store the whole sensitivity matrix in RAM to favour fast computation, this is a challenging issue that limits the number of data we can take into account in the inversion. For a computational node with 254 GB of RAM, and our current strategy for storing matrices in RAM, we estimate that we can incorporate at most $N = 300\,000$ measurements (more information on the computational costs of this study is given in Appendix C). Here, we restrict ourselves to $N \approx 50\,000$ measurements, making it possible to expand our work to overtones in the future. To achieve $N \approx 50\,000$ data, we carefully select our data with the aim to homogenise the lateral distribution of rays (see Section 4). We end up with 47,700 data in total, with approximately 3,000 data per frequency (figure 2).

For each selected measurement, we compute the corresponding 3D finite-frequency sensitivity kernel to build the sensitivity matrix $\boldsymbol{G}$, with examples shown in Figure 1. As a measure of the constraint offered by the data on the structure of the 3D upper mantle, we compute the decimal logarithm of the data sensitivity, $\log_{10} \sum_i |G_{ij}|$, where i and j designate a particular datum and model parameter respectively (see figure 2, lower right).

3.3 Target resolution, uncertainty propagation, and their trade-off

The shape of the target kernels used in the SOLA inversion is arbitrary. Ideally, it is chosen such as to produce results oriented towards addressing a specific key question. In this study, we wish for the resolution to represent simple, easy-to-interpret 3D local averages. For a given model parameter, we therefore choose the target kernel to be a 3D ellipsoid. The lateral resolution we can achieve with surface-wave data is controlled by the distribution of sources and receivers (and, to some extent, frequency). Our experience shows that it is rarely better than a few hundreds of kilometres for the frequency range used here. The vertical resolution is mostly controlled by the frequency content of the signal and it is typically on the order of tens to hundreds of kilometres. Therefore, a reasonable target kernel at a given point in the 3D grid would resemble a thick pancake centred at the query point. More formally, we design the target kernel of a model parameter as an ellipsoid whose major and semi-major axes are equal and aligned with the north-south and east-west directions at the location of the model parameter, and whose minor axis is vertical. The resulting target kernels are thick versions of the 2D kernels of Latallerie et al. (2022) and Amiri et al. (2023) and they represent a horizontally isotropic target resolution.

With SOLA, it is possible to adapt the size of the target kernels for each model parameter (i.e. for each location). For example, we could choose to achieve the best resolution possible at each location in the model given the data coverage, or we may prefer a homogeneous resolution or constant uncertainty across the spatial domain (see Freissler et al., 2024). This freedom illustrates the typical non-uniqueness of tomographic inversions. We could compute an L-curve for the resolution size versus model uncertainty to choose an optimal trade-off parameter. However, this L-curve would have a very different meaning than that computed for data-fitting approaches that typically consider data-fit versus model smoothness. With SOLA, we do not need to compute an L-curve as any choice of the trade-off parameter that fits the purpose of the study can be considered 'good', so long as the tomographic model is analysed together with its resolution and uncertainty (see also supporting information of Zaroli et al. (2017)). In this study, for simplicity, we make all target kernels the same, with 200 km long horizontal major and semi-major axes and 25 km long vertical minor axis. Figures 3 and 4 illustrate the extent of our target kernels for 10 different locations (blue ellipses).

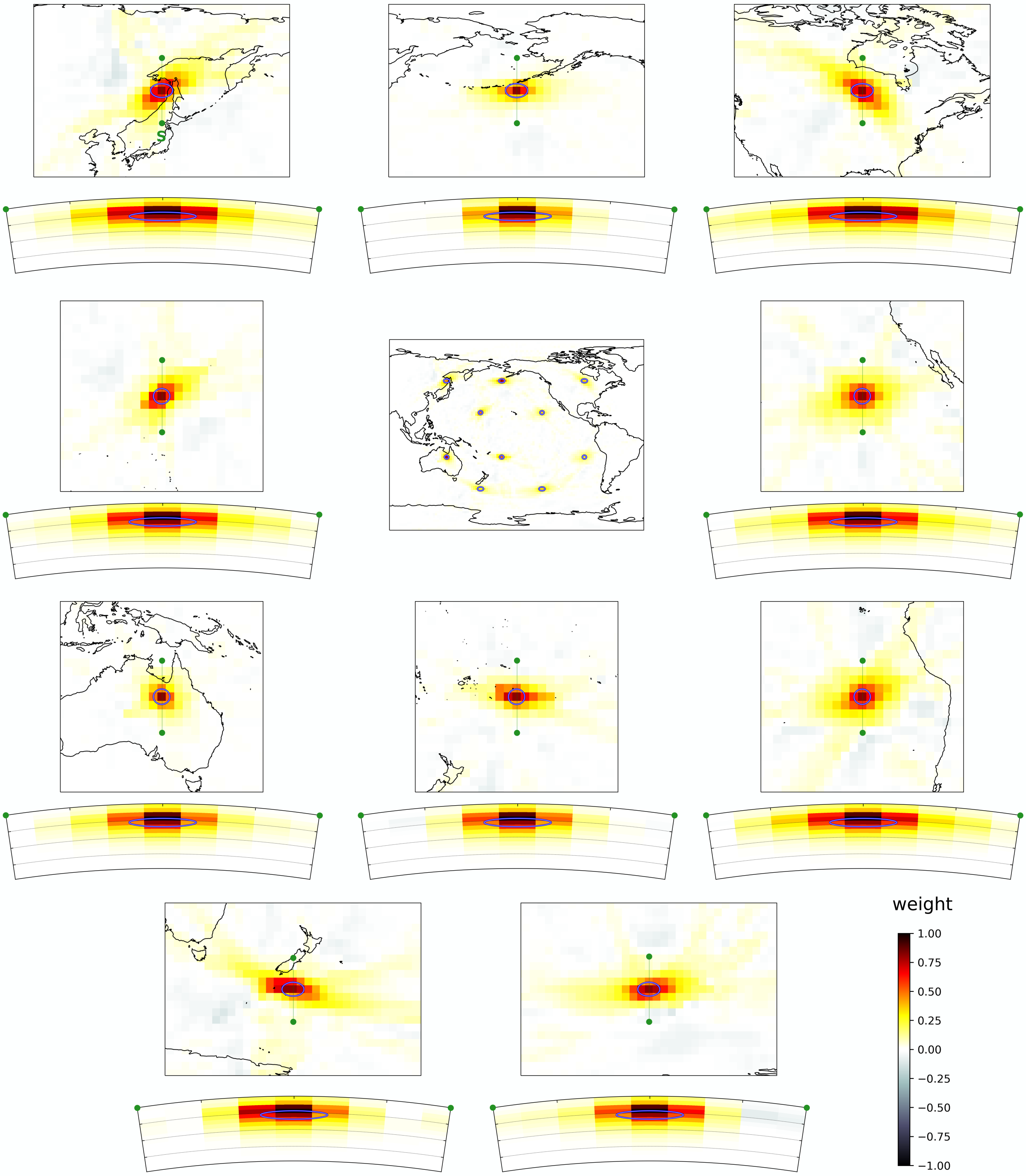


Figure 3 Resolution at 112 km depth illustrated for a selection of 10 model parameters. The centre map shows the locations of the 10 target and resolving kernels. This is shown as a sum, which may exaggerate the apparent strength of the tails. The surrounding panels are close-ups on individual kernels, both in map-view and as cross-section. All maps represent depth slices at 112 km depth and below each map is a ~ 3100 km long, N-S oriented (left to right) cross-section as indicated in green in the maps, with the dotted lines indicating depths of 100, 200 and 300 km. Blue ellipses show the lateral extent of the target kernels. All averaging kernels are normalised by their maximum, and the color scale indicated in the lower right applies to all panels.

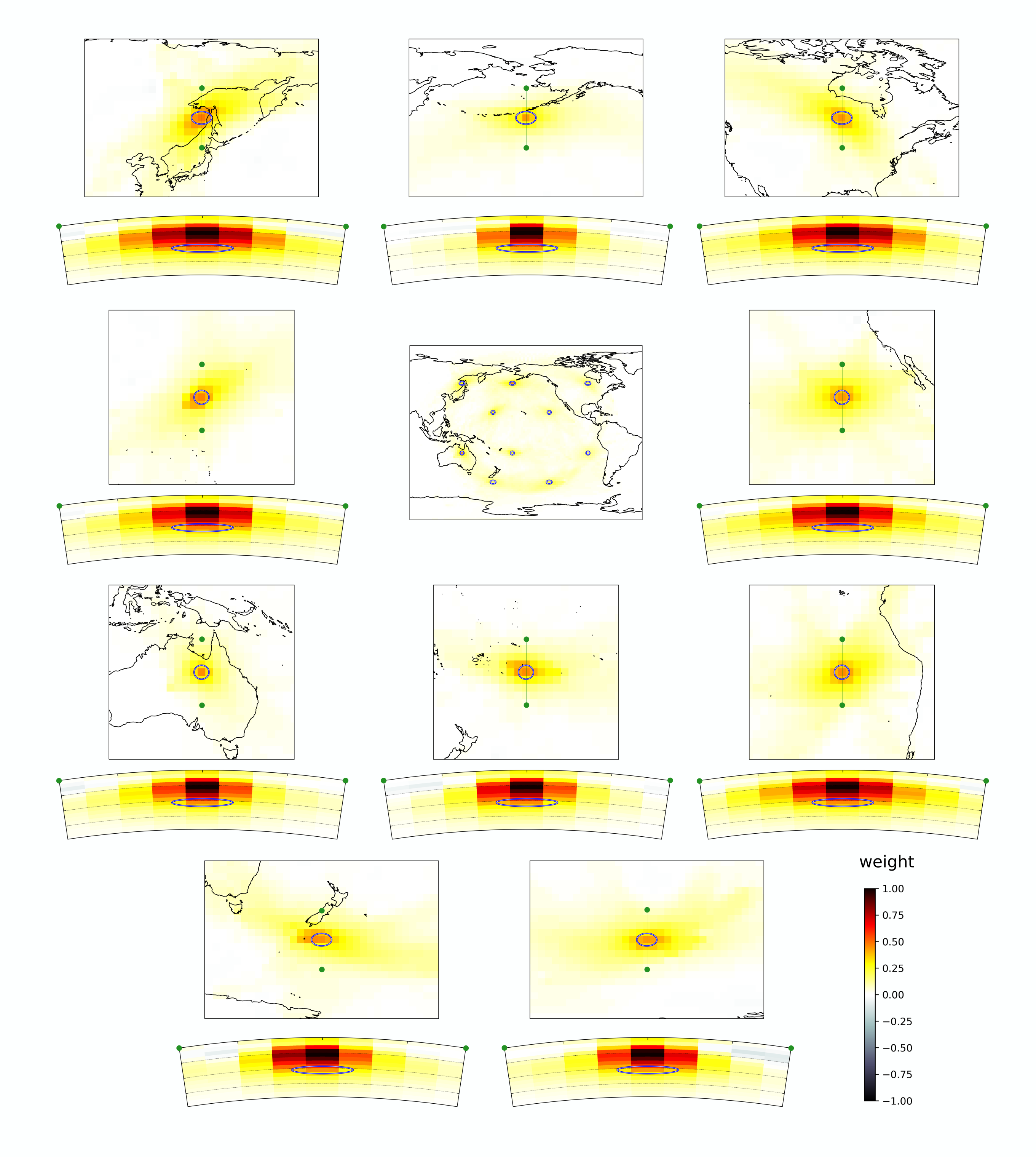


Figure 4 Same as figure 3, but for target locations at 212 km depth.

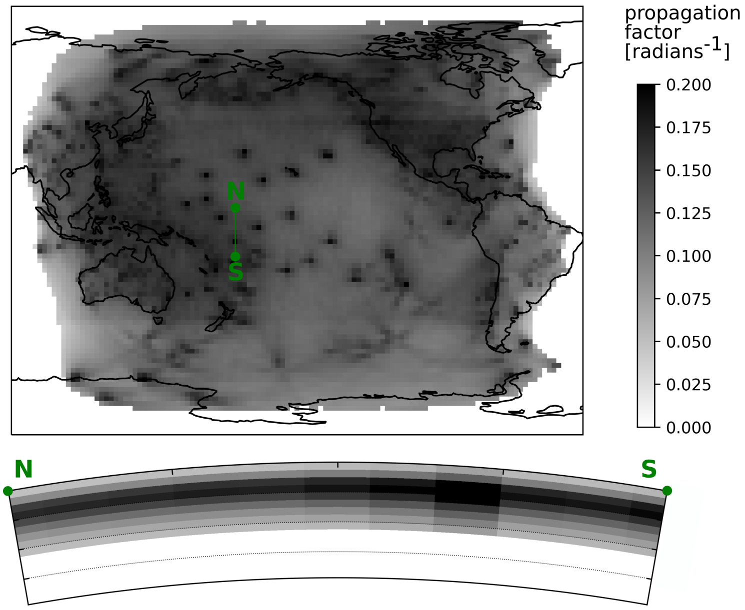


Figure 5 Illustration of the propagation of data uncertainty into model uncertainty. The map shows the ‘propagation factor’ at 112 km depth, defined as the model uncertainty given unit data uncertainty. The cross-section below the map indicates the depth dependence of the propagation factor along a vertical 2500-km long N-S oriented profile as indicated by the green line on the map, with the dotted lines indicating depths of 100, 200 and 300 km.

The data uncertainty could influence the generalised inverse we obtain with SOLA through the second term in the optimisation problem in Equation 5. However, as we aim to study the robustness of the data uncertainty itself in this study, we decide not to take it into account in designing $\boldsymbol{G}^{\dagger}$. Thus, we initially set $\boldsymbol{C_d} = \boldsymbol{I}$ and therefore $\boldsymbol{C}_{\widetilde{\boldsymbol{m}}} = (\boldsymbol{G}^{\dagger})^T\boldsymbol{G}$. This choice is only for designing $\boldsymbol{G}^{\dagger}$: once the generalised inverse has been computed, we propagate the actual measurement uncertainty into model uncertainty through $\boldsymbol{C}_{\widetilde{\boldsymbol{m}}} = (\boldsymbol{G}^{\dagger})^T\boldsymbol{C_d}\boldsymbol{G}$. Depending on the application, different data weighting (including data uncertainty), could be considered to produce an optimal generalised inverse.

The optimisation problem involves the minimisation of the difference between target and actual resolution on the one hand, and the magnitude of model uncertainty on the other hand. These two terms are balanced by the trade-off parameter η, which we set equal to 50 for all parameters. Again, it is possible to choose different values of η for different model parameters, but in practice it is computationally easier to keep η constant (see Appendix A1 of Zaroli, 2016). If, for example, one wants to give more weight to the resolution of a particular model parameter, this can also be obtained by designing a smaller size target kernel. If we vary the trade-off parameter, we obtain a typical L-shaped trade-off curve for resolution versus model uncertainty for each target (Latallerie et al., 2022; Restelli et al., 2024).

3.4 Generalised inverse: Resolution and uncertainty propagation

The seismic tomography inversion is fully characterised by the generalised inverse $\boldsymbol{G}^{\dagger}$: it determines the resolution (from $\boldsymbol{R} = \boldsymbol{G}^{\dagger}\boldsymbol{G}$) as well as the propagation of data uncertainty into model uncertainty (from $\boldsymbol{C}_{\widetilde{\boldsymbol{m}}} = (\boldsymbol{G}^{\dagger})^T\boldsymbol{C_d}\boldsymbol{G}^{\dagger}$). Lastly, it determines the propagation of data into model solution (from $\widetilde{\boldsymbol{m}} = \boldsymbol{G}^{\dagger}\boldsymbol{d}$).

It is difficult to represent the full 3D resolution as it is most easily understood in terms of an extended 3D resolving kernel associated with each model parameter. A detailed analysis thus requires 3D rendering software or the production of simple proxies, for example those proposed by Freissler et al. (2024). Here, we instead illustrate the resolution by selecting example resolving kernels. At 112 km depth (Figure 3), the resolving kernels match the target location well laterally. Their lateral size is roughly 250-450 km (if we take the radii of a circle containing 68% of the kernel). This can be compared to the length of the major and semi-major axes of the target kernels of 200 km. Some averaging kernels are significantly anisotropic, indicating lateral smearing due to the heterogeneous ray path distribution. Vertically, the resolving kernels appear also to be focused with a half-thickness of roughly 50 km. This can be compared to the length of the minor axis of the target kernels of 25 km. However, they appear slightly shifted upward from the target. Deeper down, at 212 km depth (Figure 4), the resolving kernels still match the target locations laterally, but they appear broader (300-700 km). They now also poorly match the target kernel depth-wise. Instead of peaking at 212 km depth, the resolving kernels peak at 112 km depth and tail off deeper down. This implies that what we observe in the tomographic model at 212 km depth is actually an average of the 'true model' at shallower depth.

We show the 'error propagation factor' in Figure 5. This can be interpreted as the model uncertainty for unit data uncertainty ($\boldsymbol{C_d} = \boldsymbol{I}$), obtained from $(\boldsymbol{G}^{\dagger})^T\boldsymbol{G}^{\dagger}$. We observe a positive correlation between data coverage and error propagation factor: the error propagation tends to be high where data coverage is high (e.g. North America, South-East Asia). We also clearly see patches of high error propagation in the Pacific Ocean at locations of isolated stations. This is due to the high data sensitivity at stations where many oscillatory sensitivity kernels add together. Furthermore, we note linear features with high error propagation that follow great-circle paths radiating away from some isolated stations. These probably outline sensitivity kernels that repeatedly sample similar regions. With depth, we find that the propagation factor increases down to 87 km depth and then decreases again deeper down. While this decrease may seem surprising, it is balanced by poor resolution at greater depth. In general, SOLA tends to produce models with better resolution where data sensitivity is high, at the cost of a larger error propagation factor. By choosing different sizes for the target kernels, this can be balanced (Freissler et al., 2024).

4 Input data and measurement uncertainty

We measure phase delays between 'observed' and 'reference' seismograms for 16 different frequencies ranging from 6 to 21 mHz (48-167s), in steps of 1 mHz. In this synthetic study, we use as 'observed seismograms' waveforms computed using SPECFEM3D_GLOBE (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 2002) for the 3D input model S362ANI (Kustowski et al., 2008) combined with CRUST2.0 on top (Bassin et al., 2000). Hereafter, we refer to these as SEM seismograms or SEM measurements. They were obtained from the GlobalShakeMovie project data base (Tromp et al., 2010) and downloaded from Earthscope, formerly IRIS (IRIS DMC, 2012; Hutko et al., 2017). Reference seismograms were computed using normal-mode summation with the Mineos software (Masters et al., 2011) for the 1D radial model stw105 (Kustowski et al., 2008), consistent with S362ANI. For both sets of seismograms, we use source solutions obtained from the Global-CMT project (Ekström et al., 2012) and station metadata from Earthscope. To measure the phase delay between the two sets of seismograms, we use a multi-taper measurement algorithm as suggested by Zhou et al. (2004) and detailed in appendix B. The multi-taper technique has the advantage of providing an estimate for the measurement data uncertainty as the standard deviation of the measurements across all tapers. This uncertainty estimate is particularly sensitive to cycle-skipping and contamination by higher modes and other phases.

Considering only source-receiver combinations for which the measurement time window (150 s before to 650 s after the predicted group arrival time) does not include the event origin time, we obtain 2,414,515 measurements of Rayleigh wave phase delays. We select a subset of these measurements based on the following criteria: similarity between the seismograms (cross-correlation > 0.8), source radiation in the direction of the station (> 80% of maximum radiation), measurement uncertainty (< 1.9 radians), outlier removal (1% of the dataset). This leads to 564,940 potential measurements. Due to memory limitations (as explained in section 3.2), we select a subset of $N = 47,700$ data to reduce the size of $\boldsymbol{G}$. This is achieved by randomly selecting one ray, then removing all rays whose endpoints are within 800 km radius of the endpoints of the selected ray, and repeating this process until we reach the desired number of measurements, at the frequency of interest. This gives the vector of measured data that we denote $\boldsymbol{d}^{\mathbf{measured}}$. Other approaches, such as 'bootstrapping' or 'summary ray' techniques could be experimented with to further investigate the uncertainty in the dataset or to compare to the uncertainty that we obtain with the multitaper technique. As a check, we also compute the corresponding analytical data $\boldsymbol{d}^{\mathbf{analytical}}$ by applying our forward theory $\boldsymbol{G}$ to the 3D input model S362ANI ($\boldsymbol{m}^{\mathbf{input}}$), i.e. $\boldsymbol{d}^{\mathbf{analytical}} = \boldsymbol{G}\boldsymbol{m}^{\mathbf{input}}$.

The inversion for crustal structure is highly non-linear and often avoided in surface-wave tomography. SOLA cannot handle this non-linearity and we therefore apply a crustal correction to our measurements (e.g. Marone and Romanowicz, 2007; Bozdağ and Trampert, 2008; Panning et al., 2010; Liu and Zhou, 2013; Chen and Romanowicz, 2024). For consistency with the synthetic 'observed' waveforms, we also use CRUST2.0 to compute the crustal correction (Bassin et al., 2000). We first construct 1D radial models for a combination of stw105 and CRUST2.0 at every location in a $2° \times 2°$ grid. For each grid point, we then solve a normal-mode eigenvalue problem using Mineos (Masters et al., 2011) to obtain the local phase velocity, thus building phase velocity maps for the reference model with the added crustal structure. For each source-receiver path and all frequencies in our dataset, we subsequently compute the phase accumulated in this model $\boldsymbol{\phi}^{\mathbf{ref+crust}}$ as well as in the reference model $\boldsymbol{\phi}^{\mathbf{ref}}$, assuming ray-theory (i.e. great-circle approximation). The difference in phase due to the crustal structure $\boldsymbol{\delta\phi}^{\mathbf{crust}} = \boldsymbol{\delta\phi}^{\mathbf{ref}} - \boldsymbol{\delta\phi}^{\mathbf{ref+crust}}$ is then used to correct the measured data: $\boldsymbol{d}^{\mathbf{corrected}} = \boldsymbol{d}^{\mathbf{measured}} - \boldsymbol{\delta\phi}^{\mathbf{crust}}$.

Examples of our dispersion measurement procedure and results are given in Figure 6 and used to illustrate three typical cases. In Case I (left column), measurements agree well with the analytical predictions and have low uncertainty. In Case II (middle column), measurements do not agree well with the analytical predictions, but this is compensated by high data uncertainty. In Case III (right column), which is more problematic, the measurement has low uncertainty, but it does not match the analytical prediction. In this example, it appears that the cycle-skip correction (see Appendix B) has failed to detect a cycle-skip at 8 mHz. Since the measurements are consistent for all tapers, the uncertainty estimation fails to pick-up the cycle-skip and the uncertainty remains low. Therefore, the final measurement includes a cycle-skip difference with the analytical data above 8 mHz that is not reflected in the uncertainty. This is relatively common in surface-wave tomography (e.g. Moulik et al., 2021). Even if we could spot measurements with cycle-skips in a synthetic tomography setup, we do not remove them from the dataset to mimic a real case application. Note that discrepancies between analytical predictions and measurements are due both to errors in the measurement (poorly measured data), as well as to errors in the forward theory (poor analytical data).

At this stage, we ignore uncertainty arising from theoretical errors.

To get a feeling of the volume of data falling in each of these three cases, we define three classes based on the difference between analytical prediction and measurement: (i) below 3 radians and within 3 standard deviations for Class I; (ii) above 3 radians and within 3 standard deviations for Class II; and (iii) above 3 radians and outside 3 standard deviations for Class III. For completeness, we also define Class IV as below 3 radians and outside 3 standard deviations. Classes I, II, III and IV contain respectively 27%, 1%, 43%, and 29% of the dataset. In other words, 27% of the dataset show a good agreement between the predictions and measurements and this difference is also within 3 times the measurement uncertainty. 1% of the data does not show a good agreement (i.e. above 3 radians), but is still within 3 times the measurement uncertainty. 43% shows poor agreement and is also outside 3 times the measurement uncertainty, and 29% is in good agreement, but outside 3 times the measurement uncertainty (indicating a small uncertainty). In summary, 56% of the dataset shows good agreement (class I and IV), and 28% has a difference smaller than the measurement uncertainty (class I and II). Note that the boundaries of these classes, namely the threshold of 3 radians and 3 standard deviations, are somewhat arbitrary and primarily given to provide a sense of the data volume falling within each case illustrated in figure 6.

Figure 7 presents statistics summarising our measurements and associated uncertainty. Our measured phase delays are typically larger than the analytical predictions ($\boldsymbol{d}^{\mathbf{analytical}} = \boldsymbol{G}\boldsymbol{m}^{\mathbf{input}}$) for both positive and negative delays, possibly due to non-linear effects. We may therefore expect increased positive and negative anomalies in our resulting tomographic model. We also observe a parallel branch of negative measured phase-delays with respect to the analytical predictions, likely due to non-detected cycle-skips. Our measurement uncertainty peaks around 0.3-0.5 radians, with the peak uncertainty shifting to higher values (to the right) for higher frequencies (darker colours). The effect of this shift on the resulting model uncertainty is not easy to predict as different frequencies impact the model solution in different ways (e.g. low frequency data have overall lower sensitivity). We also observe two additional peaks for higher uncertainty values, probably due to cycle-skipping and contamination with higher modes. However, measurements with these uncertainty values are not included as we apply a cut-off of 1.9 radians in our data selection. We also observe a spatial pattern in the deviation between analytical and measured data in panels c) and d). Higher differences tend to be found for rays along ridges or along the ocean-continent boundaries. High deviations are also found in the central Pacific at lower frequencies. These may be due to limitations in the forward theory as non-linearities are to be expected for these regions.

We now have a dispersion data set with an estimate of the measurement uncertainty. While this measurement uncertainty provided by the measurement algorithm accounts for cycle-skips and contamination by other phases or higher modes, to some extent, it does not capture the theoretical errors. We estimate these in the following section.

5 Results

In the perfect case of error-free analytical data $\boldsymbol{d}^{\mathbf{analytical}}$, an inversion should produce a model solution that is exactly the same as the filtered input. We confirm that by comparing the analytical model solution $\widetilde{\boldsymbol{m}}^{\mathbf{analytical}} = \boldsymbol{G}^{\dagger}\boldsymbol{d}^{\mathbf{analytical}}$ to the filtered input $\boldsymbol{R}\boldsymbol{m}^{\mathbf{input}}$. When we instead use the measurements on SEM waveforms $\boldsymbol{d}^{\mathbf{corrected}}$, differences between the filtered input model $\boldsymbol{R}\boldsymbol{m}^{\mathbf{input}}$ (Figure 8b) and the obtained model solution $\widetilde{\boldsymbol{m}}^{\mathbf{output}}$ (Fig-

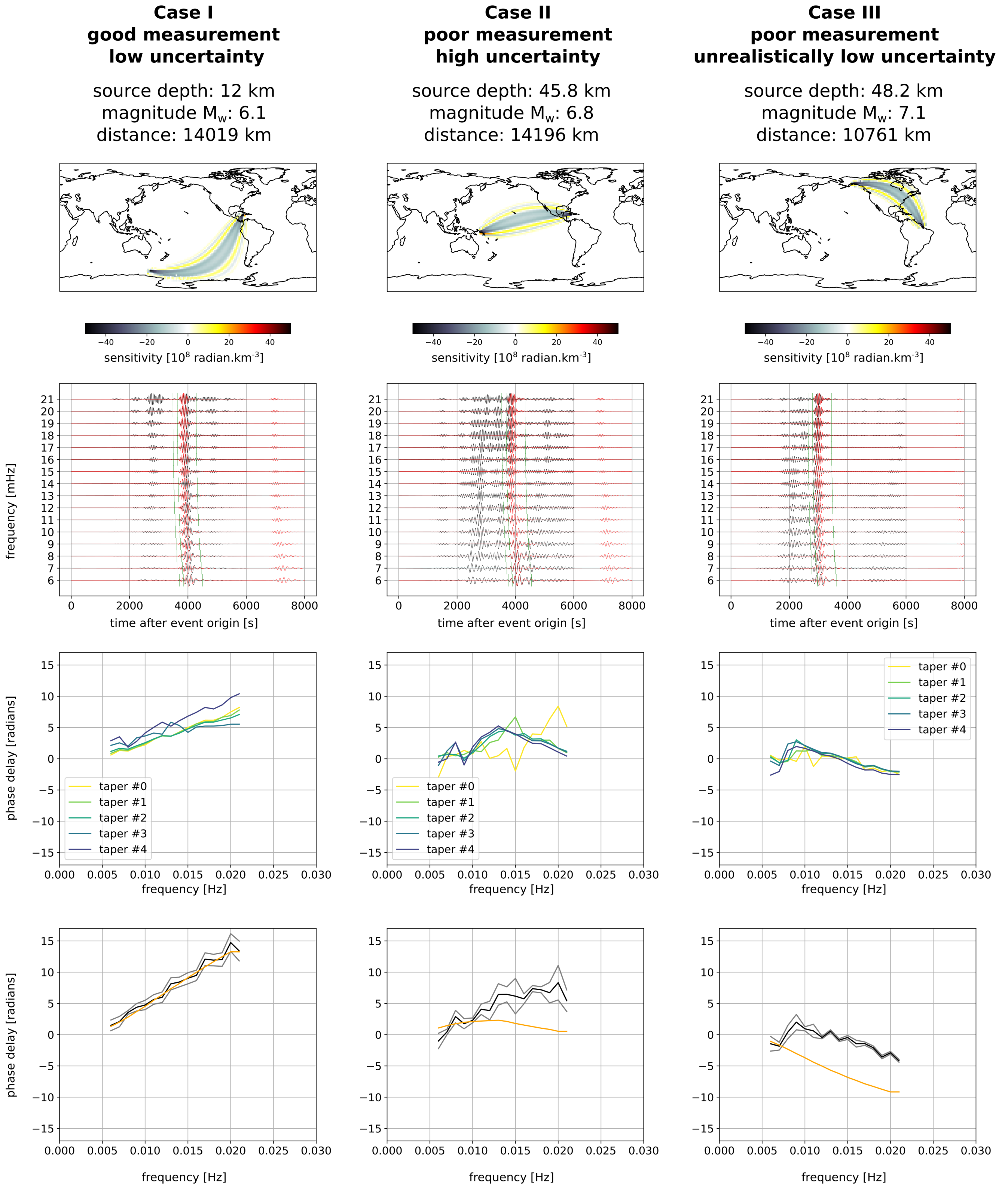


Figure 6 Example dispersion measurements, showcasing three typical cases. For each case (column), we include the sensitivity kernel at 16 mHz, plotted at 112 km depth (top row); the seismic traces (second row) for 8000 s after the event origin time (reference in black, SEM in red), filtered around each measurement frequency, and the green vertical lines indicate the start and end times of the applied tapers, around the predicted group arrival time; the measured dispersion for each taper (third row); and the final dispersion measurement (bottom row) averaged over all tapers (black) with the estimated uncertainty (grey), compared with the analytical prediction (orange). In the last row, the crustal correction is also applied to the measurements.

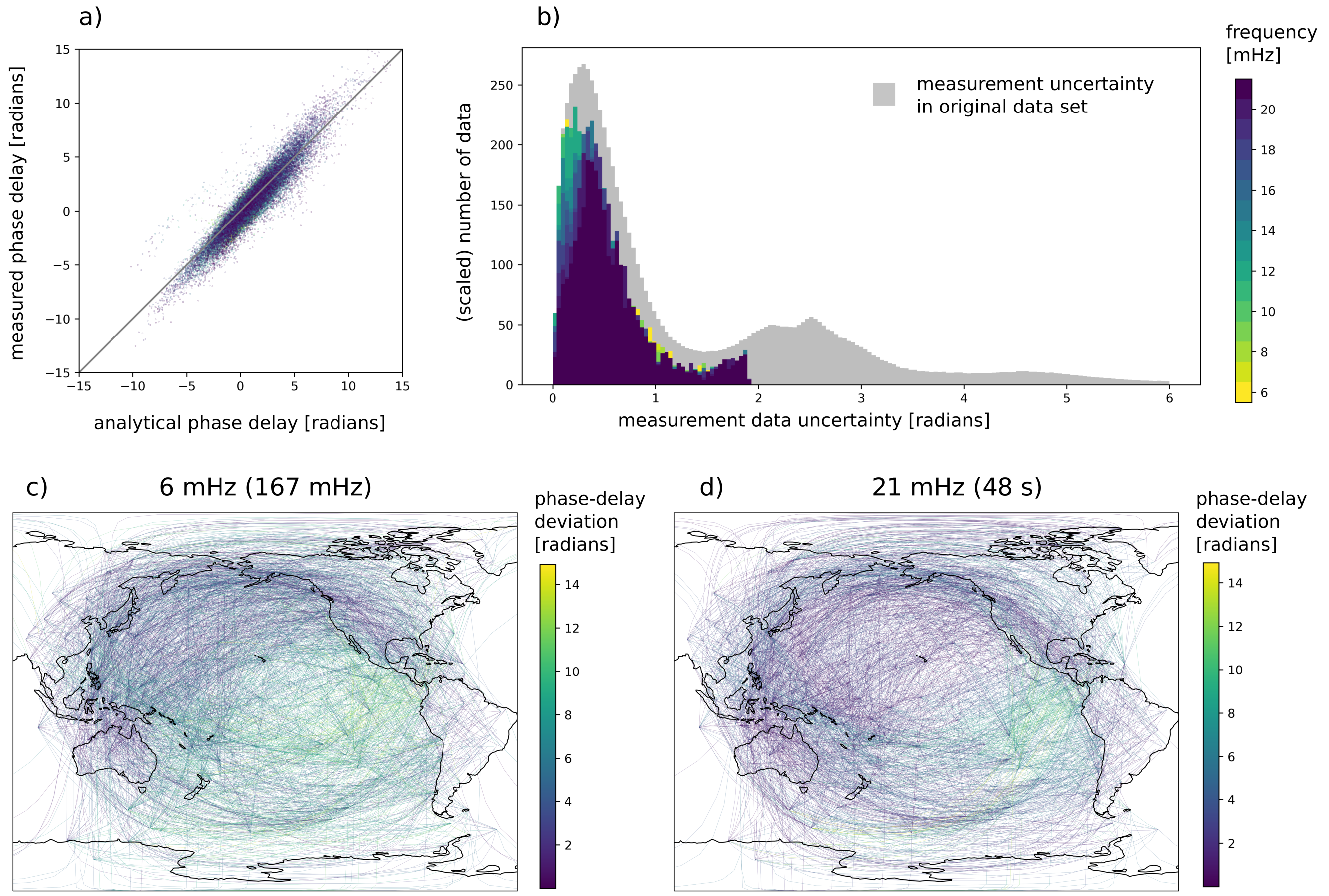


Figure 7 Summary of data and measurement uncertainty. a): Cross-plot of the measured phase delay (after crustal correction) versus the analytical phase delay prediction, coloured by frequency. Positive phase-delays typically indicate slow velocity anomalies. b) Distribution of measurement data uncertainty (coloured by frequency) before (grey) and after applying several selection criteria. Our selection criteria include a threshold for the data uncertainty (lower than 1.9 radians), as visible in the plot. The distribution of the measurement uncertainty before applying the selection criteria is scaled by 0.003 to enhance its visibility. c) and d) ray-path distribution coloured by the deviation between analytical and measured phase delays at 6 mHz and 21 mHz respectively.

ure 8d) arise due to a combination of both measurement and theoretical errors. Only the former have been taken into account in the model uncertainty map shown in Figure 8c. Note how the edges of the model solution appear rough. This is because we invert only for model parameters where the data sensitivity is higher than a certain threshold (depending on depth); this is possible due to the point-wise nature of the SOLA inversion.

5.1 Qualitative proof of concept: velocity models

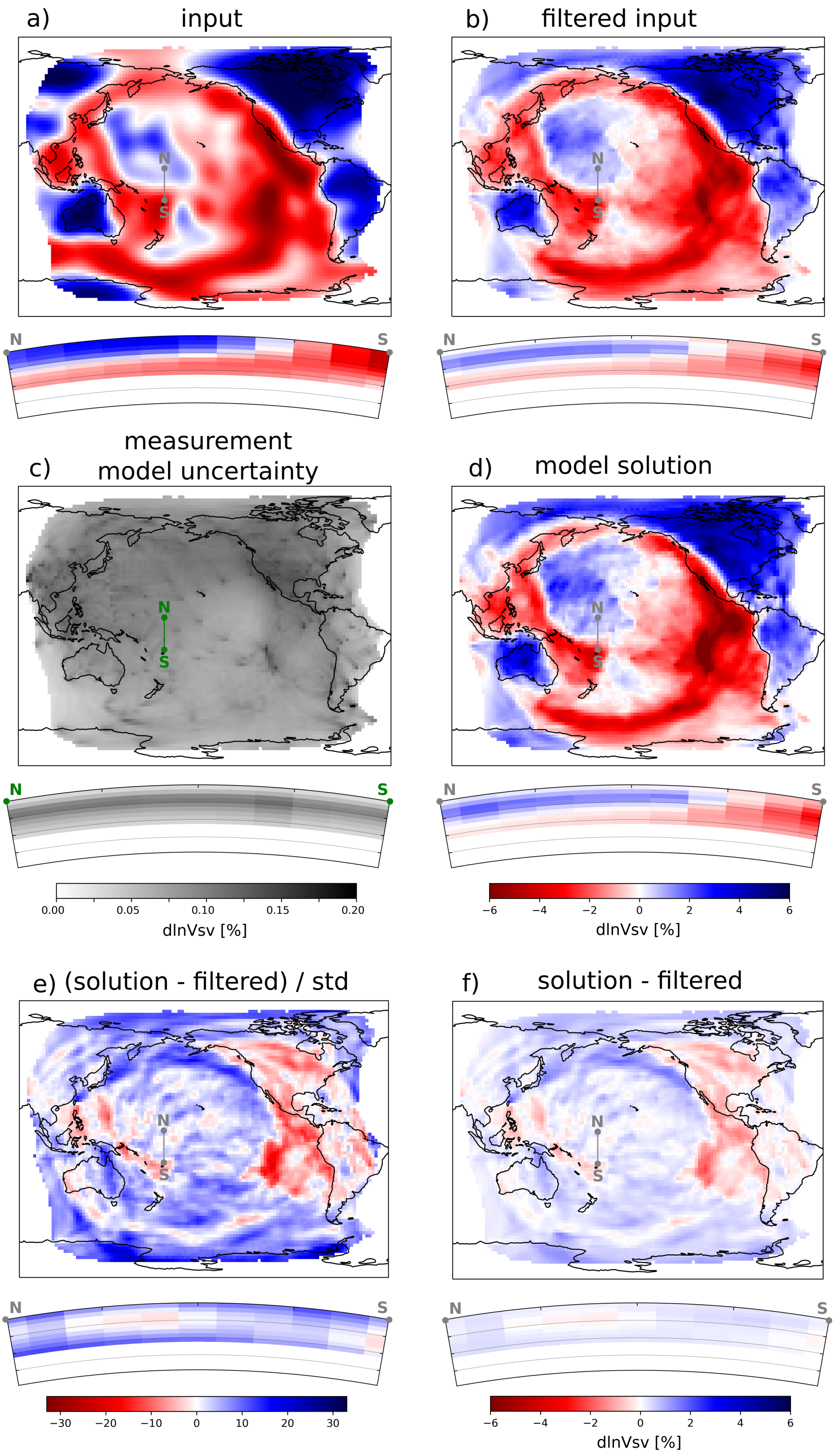
The features in the input model (Figure 8a) are also mostly present in the filtered model (Figure 8b). This indicates that the model resolution is good, at least at 112 km depth. For example, we retrieve mid-ocean ridges (low velocities at the East-Pacific rise, Pacific-Antarctic ridge, the edges of the Nazca plate), the lithosphere cooling effect (increasing velocity with distance from the ridge), the ring of fire (low velocity in the back-arc regions behind subduction zones such as the Aleutian trench, Okhotsk trench, edges of the Philippine sea plate and the Tonga-Kermadec trench), and cratons (fast velocities within the Australian and North American continents). Note that *S362ANI* is a relatively smooth model, and we would probably miss smaller-scale features in a rougher model.

The amplitudes of the velocity anomalies in the filtered model are lower than in the input model. This is expected since the filtered model represents (unbiased) local averages (Zaroli et al., 2017). The filtered model is also rougher on short length scales compared to the input model. This can be explained by the local nature of SOLA inversions where each model parameter is inverted independently from the others. In this case, we notice this particularly because the input model itself is very smooth. Some artefacts appear such as the fast velocity anomaly of SW Australia extending through the slow velocity of the Australian-Antarctic ridge. Some striations also appear in the fast velocity region in the NW Pacific, trending in the SW-NE direction. These artefacts are probably the result of anisotropic ray coverage, with many sources in East-Asia mostly recorded by stations in North-America. In addition to these artefacts, some local features disappear in the filtered model, such as the low velocity finger extending southward from the Aleutian trench, or the branch extending north-westward from Hawaii. Overall, the filtered input resembles the 'true' input model well, as also reflected in the cross-sections underneath.

The resulting model solution based on SEM seismograms (Figure 8d) appears very similar to the filtered input (Figure 8b), with differences between them shown in Figure 8e and f. Compared to the input and filtered input models described above, the model solution appears somewhat rougher due to the propagation of data errors into the model solution (Figure 8d). The striations observed in the NW Pacific in the filtered model are also stronger in the model solution than in the filtered input. The strongest spatially coherent discrepancies appear close to the East Pacific Rise, the North American Craton, and along the ocean-continent boundaries. These locations correlate well with the locations of ray paths of the most discrepant measurements (Figure 7). Finally, the cross-section indicates a good agreement between the filtered model and our model solution.

5.2 Quantitative proof of concept: uncertainty

Our model measurement uncertainty map (Figure 8c) is very similar to the 'uncertainty propagation factor' map in Figure 5. Uncertainty is typically higher where there are clusters of stations and at isolated stations with linear features following great circle paths. Uncertainty peaks at ~ 87 km depth and decreases strongly at greater depth. This uncertainty only stems from the data uncertainty, and is lacking the contribution from the theoretical uncertainty.


a) input
b) filtered input
N
S
c) measurement model uncertainty
d) model solution
0.00
0.05
0.10
0.15
0.20
dlnVsv [%]
−6
−4
−2
0
2
4
6
dlnVsv [%]
e) (solution - filtered) / std
f) solution - filtered
−30
−20
−10
0
10
20
30
−6
−4
−2
0
2
4
6
dlnVsv [%]

Theoretical errors arise from a multitude of approximations as discussed in the Introduction. How much these contribute to the data uncertainty is generally difficult to determine, but using our setup we try to obtain some insights into the theoretical uncertainty and to inform future studies.

We propose the following strategy to estimate the magnitude of the theoretical model uncertainty. Let $\boldsymbol{m}^{\mathbf{input}}$ and $\widetilde{\boldsymbol{m}}^{\mathbf{output}}$ be the input model and model solution respectively. Any discrepancy between the input model and model solution arises from the limited resolution and propagation of data uncertainty into model uncertainty. To rule out the effect of limited resolution, we apply the resolution to the input model to obtain the 'filtered' input model $\boldsymbol{Rm}^{\mathbf{input}}$. Therefore, in this synthetic setup, it is only the propagation of measurement and theoretical errors into model errors that explains the discrepancy between the 'filtered' input model and the obtained model solution. This is confirmed by the fact that the model solution based on error-free analytical data reproduces the filtered input exactly. Let us define the model misfit normalised by the model uncertainty as:

$$\xi_{\widetilde{m}} = \sqrt{\frac{1}{\sum_{k \in \mathcal{P}} V_k} \sum_{k \in \mathcal{P}} V_k \frac{[(\widetilde{\boldsymbol{m}}^{\mathbf{output}})_k - (\boldsymbol{Rm}^{\mathbf{input}})_k]^2}{(\boldsymbol{\sigma}_{\widetilde{\boldsymbol{m}}})_k^2}}, \tag{8}$$

where k refers to the model parameter index, V_k is the volume of voxel k, $\mathcal{P}$ is the set of model parameters considered for the analysis, and $\boldsymbol{\sigma}_{\widetilde{\boldsymbol{m}}}$ refers to the model uncertainty estimate.

If the data uncertainty is well-estimated, then $\xi^2_{\widetilde{m}} = 1$. As an experiment, we add random noise with a known distribution to the analytical data (i.e. to those obtained using $\boldsymbol{d}^{\mathbf{analytical}} = \boldsymbol{Gm}^{\mathbf{input}}$). In this case, the simulated data uncertainty is perfectly known and we obtain exactly $\xi^2_{\widetilde{m}} = 1$. In the case of our synthetic tomography with phase delays measured on SEM waveforms, we obtain $\xi^2_{\widetilde{m}} \approx 33 \gg 1$ when we only consider the propagation of data measurement uncertainty into model measurement uncertainty. This model uncertainty estimate is dramatically under-estimated as we may have underestimated the data measurement uncertainty and/or lack the theoretical uncertainty. We thus need to either upscale or add another component to the model uncertainty to account for this. We can write:

$$\sigma^{\mathrm{total}^2}_{\widetilde{m}^{(k)}} = \alpha^2 \sigma^{\mathrm{measurement}^2}_{\widetilde{m}^{(k)}} + \beta^2 \tag{9}$$

Here, α is the factor needed to upscale the model measurement uncertainty to account for the fact the measurement uncertainty itself might be underestimated. β is the theoretical uncertainty term that appears as an added component. We can now vary α and β independently and investigate for which combinations we obtain $\xi^2_{\widetilde{m}} = 1$. Note that in this analysis the scaling factor α and the added uncertainty component β are both assumed to be constant over all model parameters involved (consisting here of all model parameters for V_{SV} at 112 km depth).

Figure 9 shows the evolution of $\xi^2_{\widetilde{m}}$ for various combinations of α and β. We use this plot to illustrate three distinct cases. (i) The model measurement uncertainty serves as total model uncertainty, i.e. no upscaling nor added

---

Figure 8 ***(preceding page)*** Summary of synthetic inversion results, comparing a) input model S362ANI, b) input model S362ANI filtered using our resolution matrix, c) the model measurement uncertainty (propagated from data measurement uncertainty), and d) the model solution retrieved using the measured data values (based on the SEM seismograms), f) the difference between the model solution in d) and filtered input model in b), and e) same as f) but normalised by the model uncertainty. All maps represent depth slices at 112 km depth, as in Figure 3. Below each map is a N-S vertical cross-section with the location indicated by the grey or green line on the maps, and the dotted lines indicate depths of 100, 200 and 300 km.

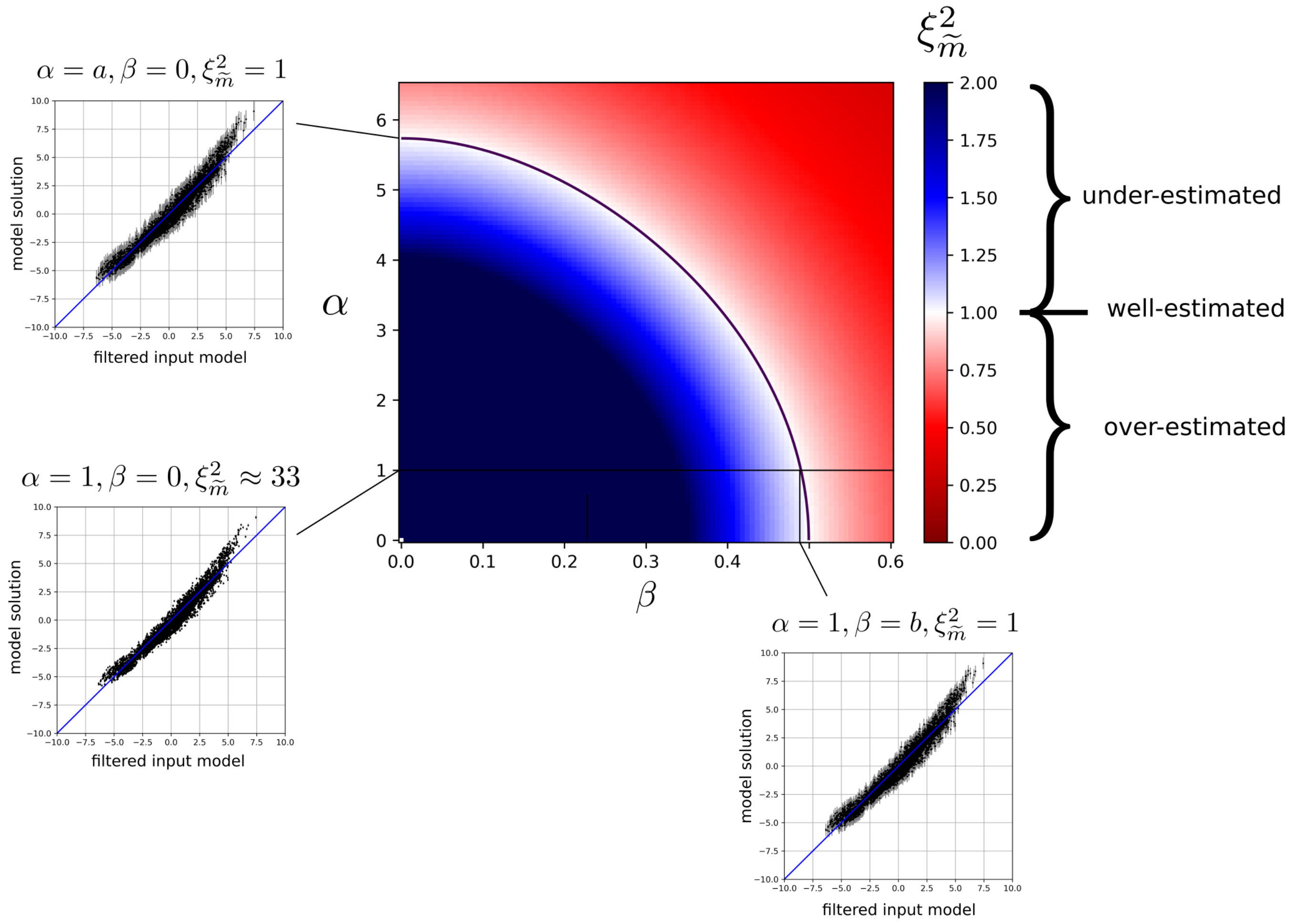


Figure 9 Model uncertainty analysis. The central plot shows the value of $\xi^2_{\widetilde{m}}$ (the misfit between the model solution and the filtered input model, normalised by the model uncertainty) for various combinations of the scaling factor α and added theoretical component β. In general, one should aim to find values of α and β that lead to $\xi^2_{\widetilde{m}} = 1$ (the black line in the white area). For small values of both α and β (blue region, or lower-left part of the plot), $\xi^2_{\widetilde{m}} > 1$, meaning that the model uncertainty is under-estimated, while the red regions indicate the model uncertainty is overestimated. The three cross-plots show the velocity variations in the model solution versus those in the filtered input model for three cases: (i) upscaled measurement uncertainty and no added component (upper-left), (ii) no upscaling nor added component (lower-left), and (iii) an added component, but no upscaling (lower-right). Note that only the error bars representing the total model uncertainty for various combinations of α and β change between these plots.

component, i.e. $\alpha = 1$ and $\beta = 0$. In this case, $\xi^2_{\widetilde{m}} \approx 33$ falls in the under-estimated uncertainty region. (ii) We only upscale the model measurement uncertainty to obtain $\xi^2_{\widetilde{m}} = 1$, with $\beta = 0$, which requires $\alpha \approx 5.74$. (iii) We add an uncertainty component without upscaling the model measurement uncertainty to obtain $\xi^2_{\widetilde{m}} = 1$, with $\alpha = 1$, which requires $\beta \approx 0.49$. This shows that the model measurement uncertainty explains only a small part of the discrepancy between the filtered input and the model solution. For comparison, the mean measurement model uncertainty is 0.09 (without upscaling). This means that the theoretical model uncertainty that needs to be added to the measurement uncertainty for a correct total model uncertainty is $0.49/0.09 \approx 5.5$ times the model measurement uncertainty (without any upscaling). Therefore, in this case, the total model uncertainty is dominated by what we refer to as theoretical uncertainty. In other words, the uncertainty provided by the measurement algorithm explains only a small fraction of the total magnitude of the uncertainty.

6 Discussion

The SOLA-finite-frequency framework for surface-wave tomography we present in this study makes it possible to obtain 3D resolution and uncertainty estimates in surface-wave tomography. Here, we discuss our findings regarding resolution and uncertainty in more detail and discuss possible future directions.

6.1 Full 3D resolution

While our setup does not handle non-linearity, it offers many advantages related to the seismic model resolution: we obtain the full resolution matrix in a computationally efficient way; the resolution is fully 3D; it is unbiased by construction (local averaging weights sum to 1) as demonstrated by Zaroli et al. (2017); and we have to some extent direct control over the resolution we obtain by choosing the target kernels. This is in contrast with most other studies that typically have assessed the resolution through inverting synthetic input models (e.g. French et al., 2013), checkerboard tests (e.g. Zhou et al., 2006; Auer et al., 2014; Rawlinson and Spakman, 2016), point spread functions (Ritsema et al., 2004; Bonadio et al., 2021), using the Hessian in the context of full-waveform inversion (e.g. Fichtner and Trampert, 2011), statistical methods using Monte Carlo approaches or transdimensional tomography (e.g. An, 2012; Bodin et al., 2012b; Sambridge et al., 2013), or other algebraic manipulations (e.g. Fichtner and Zunino, 2019; Shapiro et al., 2005; French and Romanowicz, 2014). Since surface-wave tomography is often based on a two-step approach, estimates for the resolution have typically been only 2D (lateral) or 1D (vertical), but there are some recent examples of 3D applications, for example using transdimensional tomography (Zhang et al., 2018, 2020)

In this synthetic study, we find that the resolution is laterally good enough to qualitatively retrieve the main features of the input model (compare Figure 8a and b). These large-scale or strong anomalies are features most surface-wave tomography models agree on. This may be surprising given the small number of data in our inversion (47 700). We believe there are three main reasons for this: (i) we carefully select our input data; (ii) finite-frequency theory provides improved constraints compared to ray theory since one 3D sensitivity kernel constrains more model parameters than a thin ray, while also being more accurate (e.g. Zhou et al., 2005); and (iii) the SOLA inversion performs well in optimally using the data sensitivities. Point (ii) shares some similarities with adjoint methods used in full waveform inversion, given the volumetric nature of the adjoint sensitivity kernels (e.g. Monteiller et al., 2015).

The SOLA method consists of individual inversions for each model parameter without imposing any global constraint on all model parameters together. Therefore, the fact that we recover large-scale structures in the filtered model and model solution that are consistent with the input model is encouraging (Zaroli, 2016). The global consistency of the model is provided indirectly by the overlap between the averaging kernels. However, compared to the input model, some short-scale variability arises in the filtered input, where adjacent cells show relatively strong differences. This is due to the point-wise nature of the SOLA inversion, combined with the absence of a smoothness criterion, and the smooth nature of the input model itself. Using a coarser target resolution would produce a smoother model, but would also filter out heterogeneities that are informative. Even though we present our results by plotting the mean of our model parameters in adjacent voxels (to visualise them as a tomographic model), it is important to remember that these are local average estimates.

In the above, we typically assess the performance of the resolution by comparing the filtered model to the input model. In doing this, we must keep in mind that our ability to retrieve the input model depends on the roughness of the input model itself. In particular, if the input model had contained shorter scale structure, we might not have been able to resolve it. While the resolution itself remains reliable, the comparison of input *versus* output models depends on the input itself; this bears some similarity with the inherent limitations of checkerboard tests (e.g. Lévêque et al., 1993; Rawlinson and Spakman, 2016). The full resolution itself remains necessary for robust model interpretations.

Since the data sensitivity and the resolution are fully 3D, we can confidently interpret the model resolution and uncertainty at all depths. This is a great advantage compared to our earlier 2D work (Latallerie et al., 2022), where the data sensitivity was imposed based on the lateral ray coverage (assuming ray theory). As a consequence, this study was likely too optimistic about the resolution at greater depth and therefore it was not possible to clearly state up to what depth the resolution and uncertainty estimates could be robustly interpreted. Moreover, since our resolution is fully 3D, we can investigate vertical resolution effects here. In addition to the well-known lateral smearing that arises in surface-wave tomography (discussed by Latallerie et al. (2022)), our averaging kernels indicate also significant vertical smearing (or depth leakage) in the cross-sections (Figures 3 and 4). Similar observations have been made in the context of full waveform inversion through assessment of the Hessian (e.g. Fichtner and Trampert, 2011). For some model parameters, the averages we recover relate primarily to structure above or below the 'true' location as the averaging kernel is shifted upward or downward relative to the target kernel. In particular, the structure obtained at greater depth tends to be an average over shallower structure, with the effect becoming stronger with depth. Ignoring this full 3D resolution could thus lead to biased interpretations of surface-wave tomography, for example in studies of the age-depth trends of the oceanic lithosphere (e.g. Ritzwoller et al., 2004; Priestley and Mckenzie, 2006; Maggi et al., 2006b; Isse et al., 2019). This synthetic study thus emphasises the importance of taking vertical resolution into account when interpreting surface-wave tomography models and provides a quantitative way to estimate the depth to which a surface-wave tomography model should be interpreted. Within the SOLA approach, the depth leakage could potentially be reduced by varying the trade-off parameter with depth, and by adding a directionality to the trade-off parameter. We could also use a full covariance matrix or include a weighing matrix in the optimisation problem of Equation 5, to give more weight to low frequency data (which would improve the resolution at greater depth).

Resolution and uncertainty are closely related: regions with high resolution tend to have high uncertainty, and *vice versa*. In this study, we find that the propagation of uncertainty decreases with depth (Fig. 5). This might be counter-intuitive as we expect the sensitivity of surface waves to decrease with depth. However, this observation has also been noted in other studies (e.g. Zhang et al., 2018; Earp et al., 2020; Latallerie et al., 2022). Our 3D resolution provides a robust explanation for the the decrease of uncertainty with depth. As depth increases, the resolution typically degrades, in the sense that it does not represent the average focused around the target location. It rather tends to represent an average over regions with high data sensitivity (averages are estimated over larger volumes and are shifted spatially with respect to their associated target location), leading to lower uncertainties. This illustrates that a combined analysis of uncertainty and 3D resolution is necessary to fully understand the limitations of surface-wave tomographic models.

6.2 Robust uncertainty estimates?

In this study, we estimate model uncertainty by propagating data uncertainty into model uncertainty using SOLA, which works for linear(ised) inverse problems. Other studies have used Bayesian approaches (e.g. Bodin et al., 2012b; Sambridge et al., 2013; Zhang et al., 2018), recently helped by machine learning approaches (e.g. Earp et al., 2020), where the posterior probability density function for the model can be interpreted as a measure of uncertainty. The Hessian has also been used in full waveform inversions (e.g. Fichtner and Trampert, 2011). However, in non-linear problems, the interpretation becomes more difficult. In general, we are left with the problem of estimating robust data uncertainties, which in the Bayesian philosophy entails finding the right prior probability distribution (though in this case non-informative priors could be used or compared with the posteriors).

We have estimated the measurement uncertainty with repeated sampling, changing the time window using the multi-taper technique. This is not dissimilar to previous studies, which have used summary rays, bootstrapping or perturbation methods to estimate the data mean and measurement uncertainty (e.g. Maggi et al., 2006b; Amiri et al., 2023; Asplet et al., 2020). Summary rays are not useful in our case as the sensitivity kernels depend on source mechanisms. However, future studies could compare the uncertainty we obtain with the multitaper technique to estimates using bootstrapping. Bootstrapping could also provide a range of sub-datasets with differing levels of uncertainty that could be used to investigate the effect on the model solution using SOLA. This would however have a significant computational cost.

In general, model uncertainty appears to be underestimated. This is clear from meta-analyses of published tomography models that show that the discrepancies are stronger than the typical error bars (e.g. Hosseini et al., 2018; Marignier et al., 2020; De Viron et al., 2021). This has led authors to use simple *ad hoc* criteria for upscaling the measurement uncertainty. For example, Latallerie et al. (2022) use a least-squares χ-test to upscale the uncertainty by a factor up to 3.4, while Lin et al. (2009) multiply their random error uncertainty estimates by 1.5 to obtain a more realistic model uncertainty estimate. While the measurement uncertainty might indeed be underestimated (which led us to define the factor α in section 5.2), the total uncertainty also needs to account for additional theoretical uncertainty (the factor β in section 5.2). Theoretical errors are technically deterministic, but for mathematical convenience we have treated them as random variables.

Theoretical uncertainty has typically been estimated using Monte-Carlo approaches in synthetic tests, during which input parameters are varied and the range of recovered data values is recorded as uncertainty. For example, for surface-wave dispersion measurements, Bozdağ and Trampert (2008) investigated the theoretical errors induced by imperfect crustal corrections, while Amiri et al. (2023) estimated the theoretical error induced by source mislocation. Similarly, Akbarashrafi et al. (2018) investigated the theoretical error produced by different coupling approximations on normal mode measurements, finding that reported data uncertainties need to be at least doubled to account for the errors due to theoretical omissions. In this work, we instead estimated the effect of the theoretical uncertainties on the model using a synthetic tomography setup that included many sources of theoretical uncertainty simultaneously. The effect of resolution was removed by filtering the input model so that discrepancies between our model estimate and the filtered input model represent the total uncertainty. After propagating the data measurement uncertainty into model measurement uncertainty, we noticed that these need to be upscaled by ~ 5.5 to obtain a ξ^2

of 1. This means that the theoretical model uncertainty is ~ 5.5 times larger than the model measurement uncertainty, assuming that the data measurement uncertainty is estimated correctly. The theoretical model uncertainty is thus larger than previously proposed factors of 1.5–3.4 (Lin et al., 2009; Latallerie et al., 2022), providing further evidence that the model uncertainty is indeed severely underestimated if we only propagate the data measurement uncertainty. Whether there is a need to upscale the measurement uncertainty naturally also depends on the specifics of the study and on the reliability of the measurement uncertainty estimate itself.

The main aim of this study is to provide a framework for surface-wave tomography with robust model statistics, including both the 3D resolution and total uncertainty. However, we still suffer from several drawbacks. For instance, although our measurement uncertainty should account for contamination by other phases or higher modes and cycle skipping, visual inspection indicates that this is not always the case (Figure 6). In the case of poor measurements (e.g. due to a missed cycle skip) with low uncertainty, we underestimate the measurement uncertainty and consequently overestimate the theoretical uncertainty. This is the rationale behind the factor α to upscale the measurement uncertainty in Section 5.2 and illustrates the difficulty of correctly estimating the measurement uncertainty. An interesting alternative approach was presented by several studies (Bodin and Sambridge, 2009; Bodin et al., 2012a; Zhang et al., 2020; Del Piccolo et al., 2024), which use a hierarchical transdimensional Bayesian approach where the data uncertainty is an output of the inverse process itself, rather than an input.

Another drawback of our approach is that our estimates of theoretical uncertainty depend on the input model used, i.e. S362ANI (Kustowski et al., 2008). The validity of the forward theory depends on several assumptions (e.g. forward scattering, paraxial approximation) whose applicability depends on the properties of the medium in which waves propagate (e.g. Liu and Zhou, 2013; Parisi et al., 2015). It is therefore important to perform our analysis in an Earth-like model and further work could investigate the dependency on the input model. Additionally, the scaling factor α (upscaling of the measurement uncertainty) and the added component β (representing the theoretical uncertainty) need to be determined for a sufficiently large number of model parameters for the results to be statistically significant (here we considered all model parameters at 112 km depth). In particular, we would recommend to determine these parameters for each depth in the model independently, as velocity structure and the magnitudes of measurement and theoretical uncertainties likely change with depth.

Furthermore, the theoretical model uncertainty is estimated in the model space, and therefore may depend in a non-trivial way on the model resolution. This would be reflected by a dependency of ξ^2 on the model resolution. This means that while the theoretical model uncertainty is accurately estimated for this particular solution, it may not apply to another inverse solution with a different resolution. One way to obtain the theoretical model uncertainty for models with different resolution without having to repeat their estimation in the same way, could be to compute the contribution of theoretical uncertainty on the data themselves using the sensitivity matrix, and then to propagate this contribution for models with different resolution using their respective generalised inverse matrices.

We further assume the data uncertainties to be uncorrelated, whereas in reality we expect them to be correlated to some extent – e.g an error in the source location or mechanism will impact several measurements. In theory, it is possible to account for correlations between data uncertainties, but estimating these correlations remains a challenge in surface-wave tomography. The addition of the theoretical uncertainty contribution to measurement uncertainty relies on the assumption that they are normally distributed. Furthermore, the assumption of a zero-mean Gaussian distribution for the data errors seems reasonable, but the use of more general probability distributions could also be investigated (e.g. Tarantola, 2005). Note that the off-diagonal terms of the model covariance matrix are also non-zero (even with a diagonal data covariance matrix). In SOLA we do not consider them explicitly because the information they carry is already embedded in the resolution.

Lastly, we estimate the theoretical uncertainty from the discrepancy between the filtered input model and the model solution based on measurements on SEM seismograms. Since the crustal model we assume for the crustal corrections is exactly the same as in the input model, and the source parameters used for generating the reference seismograms are exactly the same as for the SEM seismograms, there is no theoretical error associated with errors in the crustal model or source solution in our synthetic framework. Nevertheless, these two components likely introduce non-negligible errors in reality (e.g. Marone and Romanowicz, 2007; Bozdağ and Trampert, 2008; Panning et al., 2010; Ferreira et al., 2010; Liu and Zhou, 2013; Latallerie, 2022; Amiri et al., 2023). Additionally, we base our kernels on the reference model *stw105*, which is already optimal for the input model *S362ANI* that we aim to retrieve. This inherently limits the magnitude of theoretical errors arising due to non-linearity in this study. Additionally, non-linearities are expected to be stronger in the real Earth than in the relatively smooth input model S362ANI. In future, additional work could be done to estimate the model uncertainty related to these components, which could be incorporated in the proposed theoretical uncertainty estimate. In addition, we use spectral element modelling (SEM) to provide the ground truth, but any deviation from SEM in reality would lead to additional theoretical errors in a data-based study.

The restriction of SOLA inversion to linear problems remains an important overall drawback of the method. Here we treat non-linearity as an additional component in the uncertainty. Accounting for non-linearities with iterative inversion schemes can improve the models significantly (e.g. Thrastarson et al., 2024; Rodgers et al., 2024) and would allow for a better representation of the crust (e.g. Marone and Romanowicz, 2007; Bozdağ and Trampert, 2008; Panning et al., 2010; Liu and Zhou, 2013; Chen and Romanowicz, 2024). However, non-linearities would also make the computation and interpretation of the resolution and uncertainty more complicated. The extension of Backus-Gilbert theory to non-linear inverse problems as proposed by Snieder (1991) could help to better account for non-linearities with SOLA and should be the subject of future work.

Despite the drawbacks outlined above, we believe that our study provides a valuable starting point to obtain 3D resolution and to estimate theoretical model uncertainty in surface-wave tomography, upon which future work can build. This information is vital for robust model interpretations and to reconcile existing discrepancies between published tomography models (e.g. Hosseini et al., 2018; Marignier et al., 2020; De Viron et al., 2021).

6.3 Future directions

The depth sensitivity and thus resolution in this study is limited by the restriction to fundamental-mode surface-wave data. This can be mitigated by adding measurements for surface-wave overtones. In theory, including these in the presented framework is trivial, but it will be important to carefully estimate the data uncertainty for these new measurements. The resolution and uncertainty produced in our setup can be used to inform other tomographic studies. Our 3D resolution maps indicate how well certain model parameters are constrained depending on their position and particularly with depth. Based on this, we may choose sets of source-receiver paths and frequencies that best suit a certain target. For example, to better homogenise the resolution with depth, we may want to increase the number and/or the relative weight of low frequency data.

The obvious next step is to apply the approach presented here to real data, using the lessons learned in this synthetic study. As noted here, the depth leakage at depths greater than ~100 km becomes extremely strong for a dataset that is restricted to the fundamental mode. This suggests that including overtones will be necessary to obtain a model that is well-resolved deeper down in the mantle. In general, the information on 3D resolution and uncertainty obtained using SOLA would be particularly useful for testing geodynamic predictions (Freissler et al., 2022). In addition, this information would ensure that we only interpret the tomographic models to their limits, and not beyond, being aware of potential resolution artefacts, especially with depth.

There are many other directions for further development. For example, it is possible to extend the SOLA-finite-frequency framework for surface-wave tomography to other data and physical parameters, e.g. amplitude measurements to study anelasticity in the upper-mantle (e.g. Zhou, 2009b). These could be investigated independently, or through a joint approach, thus reducing theoretical uncertainty due to neglecting the effect of other physical parameters.

Conclusion

In this contribution, we have combined the Backus-Gilbert-based SOLA inverse method with finite-frequency theory in a synthetic study of the Pacific upper mantle. Our 3D modelling and inversion framework enables us to control and produce uncertainty and resolution information together with the surface-wave tomography model. We have used a synthetic framework to demonstrate the reliability of our approach and to investigate the effect of 3D resolution, laterally and vertically, in surface-wave tomography. We find that the limited resolution induces well-known artefacts, including lateral smearing effects where data coverage is poor or highly anisotropic. More importantly, we show that limited vertical resolution can induce strong artefacts with model parameters potentially representing averages of 'true' Earth properties at much shallower depth. Knowledge of this full 3D resolution is crucial for robust interpretations of surface-wave tomography models. Our synthetic setup allows us to also explore the reliability of model uncertainty estimates. We find that the theoretical uncertainty, required to match the filtered input model, might be much larger than the measurement uncertainty in the data. This demonstrates the need to account for both measurement and theoretical uncertainty in surface-wave tomography. We believe that our study is a starting point towards better use and interpretation of surface-wave tomography models.

Acknowledgments

The authors are grateful to the editor, Chiara Civiero, and to Andreas Fichtner and an anonymous reviewer whose detailed comments have helped to improve this manuscript. This research is supported by the UKRI NERC Large grant "Mantle Circulation Constrained (MC2): A multidisciplinary 4D Earth framework for understanding mantle upwellings" (NE/T01248X/1). PK acknowledges financial support from a Royal Society University Research Fellowship (URF \1 \180377). CZ acknowledges financial support from ITES (Institut Terre et Environnement de Strasbourg, UMR 7063) for a research visit to Oxford. This study used the ARCHER2 UK National Supercomputing Service (https://www.archer2.ac.uk). For this study we made extensive use of GNU/Linux and Python (including packages Scipy, Numpy, Matplotlib, Pandas and Multiprocessing). For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising.

Data and code availability

Seismic source solutions were downloaded from the Global Centroid Moment Tensor (GCMT) Catalog (Dziewonski et al., 1981; Ekström et al., 2012). The facilities of the EarthScope Consortium were used to access waveforms and related metadata and derived data products. These services are funded through the National Science Foundation's Seismological Facility for the Advancement of Geoscience (SAGE) Award under Cooperative Agreement EAR-1724509. All waveforms used in this study are SEM synthetics from the GlobalShakeMovie project (Tromp et al., 2010), and were obtained through IRIS DMC (Hutko et al., 2017; IRIS DMC, 2012). To compute the finite-frequency sensitivity kernels, we used software provided by Ying Zhou (Zhou, 2009b), available via their webpage. To compute the reference seismograms in a 1D radial Earth model using normal modes summation, we used MINEOS 1.0.2 (Masters et al., 2011) published under the GPL2 license. We thank the Computational Infrastructure for Geodynamics (http://geodynamics.org), which is funded by the National Science Foundation under awards EAR-0949446, EAR-1550901, and EAR-2149126 for making the code available.

The SOLA tomography code used in this study consists of running the LSQR algorithm of Paige and Saunders (1982) with specific input matrices and vectors. These inputs can be constructed from the sensitivity matrix and target kernels as detailed in Appendix A1 of Zaroli (2016). The LSQR code is freely downloadable from the webpage of the Systems Optimisation Laboratory (Stanford University): https://web.stanford.edu/group/SOL/software/lsqr/. A preconstructed software package for SOLA tomography is available from Christophe Zaroli (c.zaroli@unistra.fr) upon e-mail request.

References

Akbarashrafi, F., Al-Attar, D., Deuss, A., Trampert, J., and Valentine, A. P. Exact free oscillation spectra, splitting functions and the resolvability of Earth's density structure. *Geophysical Journal International*, 213(1):58–76, Apr. 2018. doi: 10.1093/gji/ggx539.

Amiri, S., Maggi, A., Tatar, M., Zigone, D., and Zaroli, C. Rayleigh wave group velocities in North-West Iran: SOLA Backus-Gilbert vs. Fast Marching tomographic methods. *Seismica*, 2(2), Dec. 2023. doi: 10.26443/seismica.v2i2.1011.

An, M. A simple method for determining the spatial resolution of a general inverse problem. *Geophysical Journal International*, 191(2): 849–864, Nov. 2012. doi: 10.1111/j.1365-246X.2012.05661.x.

Asplet, J., Wookey, J., and Kendall, M. A potential post-perovskite province in D" beneath the Eastern Pacific: evidence from new analysis of discrepant SKS–SKKS shear-wave splitting. *Geophysical Journal International*, 221(3):2075–2090, June 2020. doi: 10.1093/gji/ggaa114.

Auer, L., Boschi, L., Becker, T. W., Nissen-Meyer, T., and Giardini, D. *Savani* : A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets. *Journal of Geophysical Research: Solid Earth*, 119(4):3006–3034, Apr. 2014. doi: 10.1002/2013JB010773.

Backus, G. and Gilbert, F. The Resolving Power of Gross Earth Data. *Geophysical Journal International*, 16(2):169–205, Oct. 1968. doi: 10.1111/j.1365-246X.1968.tb00216.x.

Backus, G. E. and Gilbert, F. Uniqueness in the inversion of inaccurate gross Earth data. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 266(1173):74, 1970. doi: https://doi.org/10.1098/rsta.1970.0005.

Backus, G. E. and Gilbert, J. F. Numerical Applications of a Formalism for Geophysical Inverse Problems. *Geophysical Journal International*, 13(1-3):247–276, July 1967. doi: 10.1111/j.1365-246X.1967.tb02159.x.

Barmin, M. P., Ritzwoller, M. H., and Levshin, A. L. A Fast and Reliable Method for Surface Wave Tomography. *Pure appl. geophys.*, 158:25, 2001.

Bassin, C., Laske, G., and Masters, G. The Current Limits of Resolution for Surface Wave Tomography in North America. *Eos, Transactions American Geophysical Union*, 81(F897), 2000.

Boaga, J., Vignoli, G., and Cassiani, G. Shear wave profiles from surface wave inversion: the impact of uncertainty on seismic site response analysis. *Journal of Geophysics and Engineering*, 8(2):162–174, June 2011. doi: 10.1088/1742-2132/8/2/004.

Boaga, J., Vignoli, G., and Cassiani, G. Reply to comment on 'Shear wave profile from surface wave inversion: the impact of uncertainty on seismic site response analysis'. *Journal of Geophysics and Engineering*, 9(2):244–246, Apr. 2012. doi: 10.1088/1742-2132/9/2/244.

Bodin, T. and Sambridge, M. Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, 178(3):1411–1436, Sept. 2009. doi: 10.1111/j.1365-246X.2009.04226.x.

Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P. Transdimensional tomography with unknown data noise. *Geophysical Journal International*, 189(3):1536–1556, June 2012a. doi: 10.1111/j.1365-246X.2012.05414.x.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., and Rawlinson, N. Transdimensional inversion of receiver functions and surface wave dispersion: TRANSDIMENSIONAL INVERSION OF RF AND SWD. *Journal of Geophysical Research: Solid Earth*, 117(B2):n/a–n/a, Feb. 2012b. doi: 10.1029/2011JB008560.

Bonadio, R., Lebedev, S., Meier, T., Arroucau, P., Schaeffer, A. J., Licciardi, A., Agius, M. R., Horan, C., Collins, L., O'Reilly, B. M., Readman, P. W., and Ireland Array Working Group. Optimal resolution tomography with error tracking and the structure of the crust and upper mantle beneath Ireland and Britain. *Geophysical Journal International*, 226(3):2158–2188, June 2021. doi: 10.1093/gji/ggab169.

Bozdağ, E. and Trampert, J. On crustal corrections in surface wave tomography. *Geophysical Journal International*, 172(3):1066–1082, Mar. 2008. doi: 10.1111/j.1365-246X.2007.03690.x.

Chen, L.-W. and Romanowicz, B. On accounting for the effects of crust and uppermost mantle structure in global scale full-waveform inversion. *Geophysical Journal International*, 2024.

Dahlen, F. A. Fre¨ chet kernels for ¢nite-frequency traveltimesöI. Theory. page 18, 2000. doi: https://doi.org/10.1046/j.1365-246X.2000.00070.x.

Dahlen, F. A. and Tromp, J. *Theoretical Global Seismology*. Princeton University Press, Dec. 1999. doi: 10.1515/9780691216157.

De Viron, O., Van Camp, M., Grabkowiak, A., and Ferreira, A. M. G. Comparing global seismic tomography models using varimax principal component analysis. *Solid Earth*, 12(7):1601–1634, July 2021. doi: 10.5194/se-12-1601-2021.

Debayle, E. SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion. *Geophysical Journal International*, 137(3):747–754, June 1999. doi: 10.1046/j.1365-246x.1999.00832.x.

Debayle, E. and Kennett, B. L. N. The Australian continental upper mantle: Structure and deformation inferred from surface waves. *Journal of Geophysical Research: Solid Earth*, 105(B11):25423–25450, Nov. 2000. doi: 10.1029/2000JB900212.

Debayle, E. and Lévêque, J. J. Upper mantle heterogeneities in the Indian Ocean from waveform inversion. *Geophysical Research Letters*, 24(3):245–248, Feb. 1997. doi: 10.1029/96GL03954.

Debayle, E. and Sambridge, M. Inversion of massive surface wave data sets: Model construction and resolution assessment. *Journal of*

Geophysical Research: Solid Earth, 109(B2), Feb. 2004. doi: 10.1029/2003JB002652.

Del Piccolo, G., VanderBeek, B. P., Faccenda, M., Morelli, A., and Byrnes, J. S. Imaging Upper-Mantle Anisotropy with Transdimensional Bayesian Monte Carlo Sampling. *Bulletin of the Seismological Society of America*, 114(3):1214–1226, June 2024. doi: 10.1785/0120230233.

Dziewonski, A. M., Chou, T.-A., and Woodhouse, J. H. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research: Solid Earth*, 86(B4):2825–2852, Apr. 1981. doi: 10.1029/JB086iB04p02825.

Earp, S., Curtis, A., Zhang, X., and Hansteen, F. Probabilistic neural network tomography across Grane field (North Sea) from surface wave dispersion data. *Geophysical Journal International*, 223(3):1741–1757, Oct. 2020. doi: 10.1093/gji/ggaa328.

Ekström, G. A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25-250 s: Global dispersion model GDM52. *Geophysical Journal International*, 187(3):1668–1686, Dec. 2011. doi: 10.1111/j.1365-246X.2011.05225.x.

Ekström, G., Tromp, J., and Larson, E. W. F. Measurements and global models of surface wave propagation. *Journal of Geophysical Research: Solid Earth*, 102(B4):8137–8157, Apr. 1997. doi: 10.1029/96JB03729.

Ekström, G., Nettles, M., and Dziewoński, A. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200-201:1–9, June 2012. doi: 10.1016/j.pepi.2012.04.002.

Ferreira, A. M. G., Woodhouse, J. H., Visser, K., and Trampert, J. On the robustness of global radially anisotropic surface wave tomography. *Journal of Geophysical Research: Solid Earth*, 115(B4):2009JB006716, Apr. 2010. doi: 10.1029/2009JB006716.

Fichtner, A. and Trampert, J. Resolution analysis in full waveform inversion: Resolution in full waveform inversion. *Geophysical Journal International*, 187(3):1604–1624, Dec. 2011. doi: 10.1111/j.1365-246X.2011.05218.x.

Fichtner, A. and Zunino, A. Hamiltonian Nullspace Shuttles. *Geophysical Research Letters*, 46(2):644–651, Jan. 2019. doi: 10.1029/2018GL080931.

Foulger, G. R., Panza, G. F., Artemieva, I. M., Bastow, I. D., Cammarano, F., Evans, J. R., Hamilton, W. B., Julian, B. R., Lustrino, M., Thybo, H., and Yanovskaya, T. B. Caveats on tomographic images. *Terra Nova*, 25(4):259–281, Aug. 2013. doi: 10.1111/ter.12041.

Freissler, R., Zaroli, C., Lambotte, S., and Schuberth, B. S. A. Tomographic filtering via the generalized inverse: a way to account for seismic data uncertainty. *Geophysical Journal International*, 223(1):254–269, Oct. 2020. doi: 10.1093/gji/ggaa231.

Freissler, R., Schuberth, B. S., and Zaroli, C. The relevance of full 3D-wavefield simulations for the tomographic filtering of geodynamic models, Mar. 2022. https://meetingorganizer.copernicus.org/EGU22/EGU22-11686.html.

Freissler, R., Schuberth, B. S. A., and Zaroli, C. A concept for the global assessment of tomographic resolution and uncertainty. *Geophysical Journal International*, page ggae178, May 2024. doi: 10.1093/gji/ggae178.

French, S., Lekic, V., and Romanowicz, B. Waveform Tomography Reveals Channeled Flow at the Base of the Oceanic Asthenosphere. *Science*, 342(6155):227–230, Oct. 2013. doi: 10.1126/science.1241514.

French, S. W. and Romanowicz, B. A. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography. *Geophysical Journal International*, 199(3):1303–1327, Dec. 2014. doi: 10.1093/gji/ggu334.

Greenfield, T., Gilligan, T., Pilia, S., Cornwell, D. G., Tongkul, F., Widiyantoro, S., and Rawlinson, N. Post-Subduction Tectonics of Sabah, Northern Borneo, Inferred From Surface Wave Tomography. *Geophysical Research Letters*, 49(3), Feb. 2022. doi: 10.1029/2021GL096117.

Hjörleifsdóttir, V. *Earthquake Source Characterization Using 3D Numerical Modeling*. PhD thesis, California Institute of Technology, Apr. 2007. https://resolver.caltech.edu/CaltechETD:etd-03212007-170259. Medium: PDF Version Number: Final.

Hosseini, K., Matthews, K. J., Sigloch, K., Shephard, G. E., Domeier, M., and Tsekhmistrenko, M. SubMachine: Web-Based Tools for Exploring Seismic Tomography and Other Models of Earth's Deep Interior. *Geochemistry, Geophysics, Geosystems*, 19(5):1464–1483, May 2018. doi: 10.1029/2018GC007431.

Hutko, A. R., Bahavar, M., Trabant, C., Weekly, R. T., Fossen, M. V., and Ahern, T. Data Products at the IRIS-DMC: Growth and Usage. *Seismological Research Letters*, 88(3):892–903, May 2017. doi: 10.1785/0220160190.

IRIS DMC. Data Services Products: ShakeMovieSynthetics 1D & 3D synthetic seismograms from the Global ShakeMovie project, 2012. https://doi.org/10.17611/DP/SYN.1.

Isse, T., Kawakatsu, H., Yoshizawa, K., Takeo, A., Shiobara, H., Sugioka, H., Ito, A., Suetsugu, D., and Reymond, D. Surface wave tomography for the Pacific Ocean incorporating seafloor seismic observations and plate thermal evolution. *Earth and Planetary Science Letters*, 510: 116–130, Mar. 2019. doi: 10.1016/j.epsl.2018.12.033.

Komatitsch, D. and Tromp, J. Spectral-element simulations of global seismic wave propagation-II. Three-dimensional models, oceans, rotation and self-gravitation. *Geophysical Journal International*, 150(1):303–318, July 2002. doi: 10.1046/j.1365-246X.2002.01716.x.

Komatitsch, D. and Vilotte, J.-P. The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America*, 88(2):368–392, Apr. 1998. doi: 10.1785/BSSA0880020368.

Kustowski, B., Ekström, G., and Dziewoński, A. M. Anisotropic shear-wave velocity structure of the Earth's mantle: A global model. *Journal of Geophysical Research*, 113(B6):B06306, June 2008. doi: 10.1029/2007JB005169.

Laske, G. and Masters, G. Constraints on global phase velocity maps from long-period polarization data. *Journal of Geophysical Research: Solid Earth*, 101(B7):16059–16075, July 1996. doi: 10.1029/96JB00526.

Laske, G., Masters, G., and Zu¨rn, W. Frequency-dependent polarization measurements of long-period surface waves and their implications for global phase-velocity maps. *Physics of the Earth and Planetary Interiors*, 84(1-4):111–137, July 1994. doi: 10.1016/0031-9201(94)90037-X.

Latallerie, F. *Seismic tomography of plume-like upwellings in the French Polynesia region using Backus-Gilbert inversion*. Theses, Université de Strasbourg, Dec. 2022. https://theses.hal.science/tel-04193813. Issue: 2022STRAH012.

Latallerie, F., Zaroli, C., Lambotte, S., and Maggi, A. Analysis of tomographic models using resolution and uncertainties: a surface wave example from the Pacific. *Geophysical Journal International*, 230(2):893–907, Apr. 2022. doi: 10.1093/gji/ggac095.

Lebedev, S. and Nolet, G. Upper mantle beneath Southeast Asia from *S* velocity tomography: UPPER MANTLE BENEATH SE ASIA. *Journal of Geophysical Research: Solid Earth*, 108(B1), Jan. 2003. doi: 10.1029/2000JB000073.

Lin, F.-C., Ritzwoller, M. H., and Snieder, R. Eikonal tomography: surface wave tomography by phase front tracking across a regional broadband seismic array. *Geophysical Journal International*, 177(3):1091–1110, June 2009. doi: 10.1111/j.1365-246X.2009.04105.x.

Liu, K. and Zhou, Y. Effects of crustal thickness variations on surface wave phase delays. *Geophysical Journal International*, 192(2):773–792, Feb. 2013. doi: 10.1093/gji/ggs048.

Liu, K. and Zhou, Y. Global Rayleigh wave phase-velocity maps from finite-frequency tomography. *Geophysical Journal International*, 205 (1):51–66, Apr. 2016a. doi: 10.1093/gji/ggv555.

Liu, K. and Zhou, Y. Travelling-wave Green tensor and near-field Rayleigh-wave sensitivity. *Geophysical Journal International*, 205(1):134–145, Apr. 2016b. doi: 10.1093/gji/ggv564.

Lévêque, J., Debayle, E., and Maupin, V. Anisotropy in the Indian Ocean upper mantle from Rayleigh- and Love-waveform inversion. *Geophysical Journal International*, 133(3):529–540, June 1998. doi: 10.1046/j.1365-246X.1998.00504.x.

Lévêque, J. J., Rivera, L., and Wittlinger, G. On the use of the checker-board test to assess the resolution of tomographic inversions. *Geophysical Journal International*, 115(1):313–318, Oct. 1993. doi: 10.1111/j.1365-246X.1993.tb05605.x.

Ma, Z., Masters, G., Laske, G., and Pasyanos, M. A comprehensive dispersion model of surface wave phase and group velocity for the globe. *Geophysical Journal International*, 199(1):113–135, Oct. 2014. doi: 10.1093/gji/ggu246.

Maggi, A., Debayle, E., Priestley, K., and Barruol, G. Azimuthal anisotropy of the Pacific region. *Earth and Planetary Science Letters*, 250(1-2):

53–71, Oct. 2006a. doi: 10.1016/j.epsl.2006.07.010.

Maggi, A., Debayle, E., Priestley, K., and Barruol, G. Multimode surface waveform tomography of the Pacific Ocean: a closer look at the lithospheric cooling signature. *Geophysical Journal International*, 166(3):1384–1397, Sept. 2006b. doi: 10.1111/j.1365-246X.2006.03037.x.

Magrini, F., Diaferia, G., El-Sharkawy, A., Cammarano, F., van der Meijde, M., Meier, T., and Boschi, L. Surface-Wave Tomography of the Central-Western Mediterranean: New Insights Into the Liguro-Provençal and Tyrrhenian Basins. *Journal of Geophysical Research: Solid Earth*, 127(3), Mar. 2022. doi: 10.1029/2021JB023267.

Marignier, A., Ferreira, A. M. G., and Kitching, T. The Probability of Mantle Plumes in Global Tomographic Models. *Geochemistry, Geophysics, Geosystems*, 21(9), Sept. 2020. doi: 10.1029/2020GC009276.

Marone, F. and Romanowicz, B. Non-linear crustal corrections in high-resolution regional waveform seismic tomography. *Geophysical Journal International*, 170:460–467, Feb. 2007. doi: 10.1111/j.1365-246X.2007.03399.x.

Marquering, H., Nolet, G., and Dahlen, F. Three-dimensional waveform sensitivity kernels. *Geophysical Journal International*, 132(3):521–534, Mar. 1998. doi: 10.1046/j.1365-246X.1998.00426.x.

Masters, G., Woodhouse, J. H., and Freeman, G. Mineos v1.0.2, 2011. https://geodynamics.org/cig, 2011.

Menke, W. *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, 1989.

Montagner, J.-P. Upper mantle low anisotropy channels below the Pacific Plate. *Earth and Planetary Science Letters*, 202(2):263–274, Sept. 2002. doi: 10.1016/S0012-821X(02)00791-4.

Monteiller, V., Chevrot, S., Komatitsch, D., and Wang, Y. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method. *Geophysical Journal International*, 202(2):811–827, Aug. 2015. doi: 10.1093/gji/ggv189.

Moulik, P., Lekic, V., Romanowicz, B., Ma, Z., Schaeffer, A., Ho, T., Beucler, E., Debayle, E., Deuss, A., Durand, S., Ekström, G., Lebedev, S., Masters, G., Priestley, K., Ritsema, J., Sigloch, K., Trampert, J., and Dziewonski, A. M. Global reference seismological data sets: multimode surface wave dispersion. *Geophysical Journal International*, 228(3):1808–1849, Dec. 2021. doi: 10.1093/gji/ggab418.

Nolet, G. Solving or resolving inadequate and noisy tomographic systems. *Journal of Computational Physics*, 61(3):463–482, Dec. 1985. doi: 10.1016/0021-9991(85)90075-0.

Nolet, G. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*. Cambridge University Press, 1 edition, Sept. 2008. doi: 10.1017/CBO9780511984709.

Nolet, G., Montelli, R., and Virieux, J. Explicit, approximate expressions for the resolution and *a posteriori* covariance of massive tomographic systems. *Geophysical Journal International*, 138(1):36–44, July 1999. doi: 10.1046/j.1365-246x.1999.00858.x.

Ouattara, Y., Zigone, D., and Maggi, A. Rayleigh wave group velocity dispersion tomography of West Africa using regional earthquakes and ambient seismic noise. *Journal of Seismology*, 23(6):1201–1221, Nov. 2019. doi: 10.1007/s10950-019-09860-z.

Paige, C. C. and Saunders, M. A. LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Transactions on Mathematical Software (TOMS)*, 8(1):43–71, Mar. 1982. doi: 10.1145/355984.355989.

Panning, M. P., Lekić, V., and Romanowicz, B. A. Importance of crustal corrections in the development of a new global model of radial anisotropy. *Journal of Geophysical Research*, 115(B12):B12325, Dec. 2010. doi: 10.1029/2010JB007520.

Parisi, L., Ferreira, A. M. G., and Capdeville, Y. Validity domain of the Born approximation for seismic waveform modelling in realistic 3-D Earth structure. *Geophysical Journal International*, 200(2):910–916, Feb. 2015. doi: 10.1093/gji/ggu446.

Park, J., Lindberg, C. R., and Vernon, F. L. Multitaper spectral analysis of high-frequency seismograms. *Journal of Geophysical Research*, 92 (B12):12675, 1987a. doi: 10.1029/JB092iB12p12675.

Park, J., Vernon, F. L., and Lindberg, C. R. Frequency dependent polarization analysis of high-frequency seismograms. *Journal of Geophysical Research*, 92(B12):12664, 1987b. doi: 10.1029/JB092iB12p12664.

Parker, R. L. Understanding Inverse Theory. *Annual Review of Earth and Planetary Sciences*, 5(1):35–64, May 1977. doi: 10.1146/annurev.ea.05.050177.000343.

Percival, D. B. and Walden, A. T. *Spectral Analysis for Physical Applications*. Cambridge University Press, 1 edition, June 1993. doi: 10.1017/CBO9780511622762.

Pijpers, F. P. and Thompson, M. J. Faster formulations of the optimally localized averages method for helioseismic inversions. *Astronomy and Astrophysics*, 262:L33–L36, Sept. 1992.

Pijpers, F. P. and Thompson, M. J. The SOLA method for helioseismic inversion. *Astronomy and Astrophysics*, 281:231–240, Jan. 1994.

Priestley, K. Seismic evidence for a moderately thick lithosphere beneath the Siberian Platform. *Geophysical Research Letters*, 30(3):1118, 2003. doi: 10.1029/2002GL015931.

Priestley, K. and Mckenzie, D. The thermal structure of the lithosphere from shear wave velocities. *Earth and Planetary Science Letters*, 244 (1-2):285–301, Apr. 2006. doi: 10.1016/j.epsl.2006.01.008.

Rawlinson, N. and Spakman, W. On the use of sensitivity tests in seismic tomography. *Geophysical Journal International*, 205(2):1221–1243, May 2016. doi: 10.1093/gji/ggw084.

Rawlinson, N., Fichtner, A., Sambridge, M., and Young, M. Seismic Tomography and the Assessment of Uncertainty. *Advances in Geophysics*, 55:1–76, 2014. doi: https://doi.org/10.1016/bs.agph.2014.08.001.

Restelli, F., Zaroli, C., and Koelemeijer, P. Robust estimates of the ratio between S- and P-wave velocity anomalies in the Earth's mantle using normal modes. *Physics of the Earth and Planetary Interiors*, 347:107135, Feb. 2024. doi: 10.1016/j.pepi.2023.107135.

Ritsema, J., van Heijst, H. J., and Woodhouse, J. H. Global transition zone tomography: GLOBAL TRANSITION ZONE TOMOGRAPHY. *Journal of Geophysical Research: Solid Earth*, 109(B2), Feb. 2004. doi: 10.1029/2003JB002610.

Ritsema, J., McNamara, A. K., and Bull, A. L. Tomographic filtering of geodynamic models: Implications for model interpretation and large-scale mantle structure. *Journal of Geophysical Research*, 112(B01303), Jan. 2007. doi: 10.1029/2006JB004566.

Ritzwoller, M. H., Shapiro, N. M., and Zhong, S.-J. Cooling history of the Pacific lithosphere. *Earth and Planetary Science Letters*, 226(1-2): 69–84, Sept. 2004. doi: 10.1016/j.epsl.2004.07.032.

Rodgers, A. J., Doody, C. D., and Fichtner, A. WUS324: Multiscale Full Waveform Inversion Approaching Convergence Improves Waveform Fits While Imaging Seismic Structure of the Western United States. *Geophysical Research Letters*, 51(20):e2024GL110911, Oct. 2024. doi: 10.1029/2024GL110911.

Ruan, Y. and Zhou, Y. The effects of 3-D anelasticity (*Q*) structure on surface wave phase delays. *Geophysical Journal International*, 181(1): 479–492, Apr. 2010. doi: 10.1111/j.1365-246X.2010.04514.x.

Sambridge, M., Bodin, T., Gallagher, K., and Tkalčić, H. Transdimensional inference in the geosciences. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110547, Feb. 2013. doi: 10.1098/rsta.2011.0547.

Scales, J. A. and Snieder, R. To Bayes or not to Bayes? *GEOPHYSICS*, 62(4):1045–1046, July 1997. doi: 10.1190/1.6241045.1.

Schuberth, B. S. A., Bunge, H., and Ritsema, J. Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone? *Geochemistry, Geophysics, Geosystems*, 10(5):2009GC002401, May 2009. doi: 10.1029/2009GC002401.

Seredkina, A. I. Surface Wave Tomography of the Arctic from Rayleigh and Love Wave Group Velocity Dispersion Data. *Izvestiya, Physics of the Solid Earth*, 55(3):439–450, May 2019. doi: 10.1134/S106935131903008X.

Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M. H. High-Resolution Surface-Wave Tomography from Ambient Seismic Noise. *Science*, 307(5715):1615–1618, Mar. 2005. doi: 10.1126/science.1108339.

Simmons, N. A., Schuberth, B. S. A., Myers, S. C., and Knapp, D. R. Resolution and Covariance of the LLNL-G3D-JPS Global Seismic Tomography Model: Applications to Travel time Uncertainty and Tomographic Filtering of Geodynamic Models. *Geophysical Journal International*, 217(3):1543–1557, June 2019. doi: 10.1093/gji/ggz102.

Simons, F. J., van der Hilst, R. D., Montagner, J.-P., and Zielhuis, A. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle. *Geophysical Journal International*, 151(3):738–754, Dec. 2002. doi: 10.1046/j.1365-246X.2002.01787.x.

Slepian, D. Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-V: The Discrete Case. *Bell System Technical Journal*, 57 (5):1371–1430, May 1978. doi: 10.1002/j.1538-7305.1978.tb02104.x.

Snieder, R. 3-D linearized scattering of surface waves and a formalism for surface wave holography. *Geophysical Journal International*, 84 (3):581–605, Mar. 1986. doi: 10.1111/j.1365-246X.1986.tb04372.x.

Snieder, R. An extension of Backus-Gilbert theory to nonlinear inverse problems. *Inverse Problems*, 7(3):409–433, June 1991. doi: 10.1088/0266-5611/7/3/008.

Snieder, R. and Nolet, G. Linearized scattering of surface waves on a spherical Earth. *Journal of Geophysics*, 61(1):55–63, 1987.

Socco, L. V., Foti, S., Comina, C., and Boiero, D. Comment on 'Shear wave profiles from surface wave inversion: the impact of uncertainty on seismic site response analysis'. *Journal of Geophysics and Engineering*, 9(2):241–243, Apr. 2012. doi: 10.1088/1742-2132/9/2/241.

Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Jan. 2005. doi: 10.1137/1.9780898717921.

Tarantola, A. and Valette, B. Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics*, 20(2): 219, 1982. doi: 10.1029/RG020i002p00219.

Thomson, D. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982. doi: 10.1109/PROC.1982.12433.

Thrastarson, S., Van Herwaarden, D.-P., Noe, S., Josef Schiller, C., and Fichtner, A. REVEAL: A Global Full-Waveform Inversion Model. *Bulletin of the Seismological Society of America*, 114(3):1392–1406, June 2024. doi: 10.1785/0120230273.

Tian, Y., Zhou, Y., Sigloch, K., Nolet, G., and Laske, G. Structure of North American mantle constrained by simultaneous inversion of multiple-frequency *SH* , *SS* , and Love waves. *Journal of Geophysical Research*, 116(B2):B02307, Feb. 2011. doi: 10.1029/2010JB007704.

Trampert, J. Global seismic tomography: the inverse problem and beyond. *Inverse Problems*, 14(3):371–385, June 1998. doi: 10.1088/0266-5611/14/3/002.

Tromp, J., Komatitsch, D., Hjörleifsdóttir, V., Liu, Q., Zhu, H., Peter, D., Bozdag, E., McRitchie, D., Friberg, P., Trabant, C., and Hutko, A. Near real-time simulations of global CMT earthquakes: Near real-time simulations of CMT earthquakes. *Geophysical Journal International*, 183 (1):381–389, Oct. 2010. doi: 10.1111/j.1365-246X.2010.04734.x.

Wiggins, R. A. The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure. *Reviews of Geophysics*, 10(1):251, 1972. doi: 10.1029/RG010i001p00251.

Yomogida, K. Fresnel zone inversion for lateral heterogeneities in the earth. *pure and applied geophysics*, 138(3):391–406, Sept. 1992. doi: 10.1007/BF00876879.

Yoshizawa, K. and Kennett, B. L. N. Multimode surface wave tomography for the Australian region using a three-stage approach incorporating finite frequency effects: THREE-STAGE SURFACE WAVE TOMOGRAPHY. *Journal of Geophysical Research: Solid Earth*, 109(B2), Feb. 2004. doi: 10.1029/2002JB002254.

Yoshizawa, K. and Kennett, B. L. N. Sensitivity kernels for finite-frequency surface waves. *Geophysical Journal International*, 162(3):910–926, Sept. 2005. doi: 10.1111/j.1365-246X.2005.02707.x.

Zaroli, C. Global seismic tomography using Backus-Gilbert inversion. *Geophysical Journal International*, 207(2):876–888, Nov. 2016. doi: 10.1093/gji/ggw315.

Zaroli, C. Seismic tomography using parameter-free Backus-Gilbert inversion. *Geophysical Journal International*, 218(1):619–630, July 2019. doi: 10.1093/gji/ggz175.

Zaroli, C., Koelemeijer, P., and Lambotte, S. Toward Seeing the Earth's Interior Through Unbiased Tomographic Lenses. *Geophysical Research Letters*, 44(22):11,399–11,408, Nov. 2017. doi: 10.1002/2017GL074996.

Zhang, X., Curtis, A., Galetti, E., and de Ridder, S. 3-D Monte Carlo surface wave tomography. *Geophysical Journal International*, 215(3):1644–1658, Dec. 2018. doi: 10.1093/gji/ggy362.

Zhang, X., Roy, C., Curtis, A., Nowacki, A., and Baptie, B. Imaging the subsurface using induced seismicity and ambient noise: 3-D tomographic Monte Carlo joint inversion of earthquake body wave traveltimes and surface wave dispersion. *Geophysical Journal International*, 222(3):1639–1655, Sept. 2020. doi: 10.1093/gji/ggaa230.

Zhou, Y. Multimode surface wave sensitivity kernels in radially anisotropic earth media. *Geophysical Journal International*, 176(3):865–888, Mar. 2009a. doi: 10.1111/j.1365-246X.2008.04010.x.

Zhou, Y. Surface-wave sensitivity to 3-D anelasticity. *Geophysical Journal International*, 178(3):1403–1410, Sept. 2009b. doi: 10.1111/j.1365-246X.2009.04230.x.

Zhou, Y., Dahlen, F. A., and Nolet, G. Three-dimensional sensitivity kernels for surface wave observables. *Geophysical Journal International*, 158(1):142–168, July 2004. doi: 10.1111/j.1365-246X.2004.02324.x.

Zhou, Y., Dahlen, F. A., Nolet, G., and Laske, G. Finite-frequency effects in global surface-wave tomography. *Geophysical Journal International*, 163(3):1087–1111, Dec. 2005. doi: 10.1111/j.1365-246X.2005.02780.x.

Zhou, Y., Nolet, G., Dahlen, F. A., and Laske, G. Global upper-mantle structure from finite-frequency surface-wave tomography. *Journal of Geophysical Research*, 111(B04304), 2006. doi: 10.1029/2005JB003677.

Appendix A: The SOLA method in more detail

In this appendix, we provide more details on the SOLA method inspired by Zaroli (2016); Zaroli et al. (2017); Zaroli (2019). Here we use a slightly different notation following Latallerie et al. (2022). Let us consider N data that are gathered in a data vector $\boldsymbol{d} \in \mathcal{R}^N$. In addition, the continuous 'true' model is discretised with model parameters gathered in a model vector $\boldsymbol{m} \in \mathcal{R}^M$. Assuming linearity, the data are expressed as linear combination of the model parameters $\boldsymbol{d} = \boldsymbol{G}\boldsymbol{m}$, where the forward mapping $\boldsymbol{G} \in \mathcal{R}^{N\times M}$ contains the physical laws relating the N data to the M model parameters. This forward mapping includes theoretical errors as $\boldsymbol{G}$ does not exactly predict what we aim to measure. Additionally, the measurement introduces data errors (the measurement does not exactly measure what we aim to measure). We first discuss SOLA without theoretical and measurement errors and come back to these later on.

The inverse problem is ill-posed, i.e. $\boldsymbol{G}$ is not invertible and we cannot find a unique value for each model parameter. With SOLA, we break this non-uniqueness by instead finding a single value for a *local average* (Zaroli, 2016). Here, we define this local average as a combination of model parameters that is informative, i.e. a weighted sum of model parameters that is local to a model parameter location. The weights of such a sum is the resolution of the specific model parameter.

Let $\widetilde{m}^{(k)} \in \mathcal{R}$ be the estimate of a local average around model parameter k and let us write this estimate as a linear combination of the data $\widetilde{m}^{(k)} = \boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}\boldsymbol{d}$, where $\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})} \in \mathcal{R}^N$ is the vector containing the weights for the linear combination of the data. We use the forward equation to obtain $\widetilde{m}^{(k)} = \boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}\boldsymbol{G}\boldsymbol{m}$, which implies that the vector $\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}\boldsymbol{G}$ contains the weights specifically for the local average of model parameter k. This defines the resolution for this model parameter: $\boldsymbol{R}^{(\boldsymbol{k})} = (R_j^{(k)})_{j=1,..,M} = (\sum_{i=1}^{N} {G^{\dagger}}_i^{(k)} G_{ij})_{j=1,..,M}$. To account for varying voxel volumes, we define the averaging kernel $\boldsymbol{A}^{(\boldsymbol{k})} = (R_j^{(k)}/V_j)_{j=1,..,M}$. To find $\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}$ we design a target local average, or target kernel, $\boldsymbol{T}^{(\boldsymbol{k})} \in \mathcal{R}^M$ and minimise the squared distance between the averaging and target kernel:

$$\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})} = \arg\min_{\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}} \sum_{j=1,..,M} V_j \left[\left(\sum_{i=1,..,N} {G^{\dagger}}_i^{(k)} G_{ij}/V_j \right) - T_j^{(k)} \right]^2 \tag{10}$$

The aim of the minimisation problem in Equation 10 is to fit the target kernel given the limits imposed by the data sensitivity, i.e. the geometry of the problem. In addition, we can add a uni-modularity constraint on the resolution for the local average to be unbiased: $\sum_{ij} {G^{\dagger}}_i^{(k)} G_{ij} = 1$ (Zaroli et al., 2017). Values greater or smaller than unity imply that the local average is artificially over- or under-estimating the average of the 'true' model parameter. Note that if we compute the linear combination $\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}$ for all M model parameters, and organise them into a matrix $\boldsymbol{G}^{\dagger}$, then we can write $\widetilde{\boldsymbol{m}} = \boldsymbol{G}^{\dagger}\boldsymbol{d}$ and $\widetilde{\boldsymbol{m}} = \boldsymbol{G}^{\dagger}\boldsymbol{G}\boldsymbol{m}$, where $\widetilde{\boldsymbol{m}} \in \mathcal{R}^M$ is the collection of local average estimates. In fact, $\boldsymbol{G}^{\dagger}$ is the generalised inverse for the inverse problem, and $\widetilde{\boldsymbol{m}}$ is the model solution. This model solution can be visualised, as we have done in this study, but it is important to recall that this model solution is nothing more than a collection of local averages, not estimates of individual model parameters.

The above is incomplete as all observed data contain errors. To account for this, we can represent each datum as a Gaussian probability distribution whose mean is the measured datum (d_i) and whose standard deviation is the estimated measurement uncertainty (σ_{di}). Under this assumption, a model parameter estimate is also a Gaussian probability distribution as it is a linear combination of Gaussian probability distributions and we can easily compute its mean and standard deviation. The mean of the local average distribution is still given by $\widetilde{m}^{(k)} = \sum_{i=1}^{N} {G^{\dagger}}_i^{(k)} d_i$, while the standard deviation is given by $\sigma_{\widetilde{m}^{(k)}} = \sqrt{\sum_{i=1}^{N} {{G^{\dagger}}_i^{(k)}}^2 \sigma_{di}^2}$. Note that the model uncertainty is for a local average estimate, not an estimate for a given model parameter. The weights that specify the linear combination of data ($\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}$) also influence the propagation of data uncertainty into model uncertainty. To account for this in designing $\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}$, i.e. to find a combination of model parameters that also minimises the propagation of data uncertainty into model uncertainty, we amend the minimisation problem of Equation 10:

$$\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})} = \arg\min_{\boldsymbol{G^{\dagger}}^{(\boldsymbol{k})}} \sum_{j=1,..,M} V_j \left[\left(\sum_{i=1,..,N} {G^{\dagger}}_i^{(k)} G_{ij}/V_j \right) - T_j^{(k)} \right]^2 + {\eta^k}^2 \left(\sum_{i=1,..,N} {{G^{\dagger}}_i^{(k)}}^2 \sigma_{di}^2 \right), \quad \text{s.t.} \sum_{ij} {G^{\dagger}}_i^{(k)} G_{ij} = 1. \tag{11}$$

with η^k the trade-off parameter for the model parameter. Equation 11 leads to a set of equations for each model parameter k with its particular target resolution $\boldsymbol{T}^{(\boldsymbol{k})}$. These can be solved, as proposed by Zaroli (2016), using an

LSQR algorithm (e.g. Paige and Saunders, 1982). More details on this implementation can be found in appendix A1 of Zaroli (2016). A summary of the SOLA inversion illustrating the inputs and outputs is presented in figure 10.

Appendix B: Phase delay measurements using multi-taper technique

Let $s(\omega) = A(\omega)e^{\phi(\omega)}$ be the mathematical expression of the reference seismogram computed for the 1D reference model for a given source-receiver pair at some frequency ω, with amplitude A and phase ϕ. Let $o(\omega) = A^o(\omega)e^{\phi^o(\omega)}$ be defined equivalently for the observed seismogram, or the SEM seismogram in the case of this synthetic study. The accumulated phase results from source and receiver effects, caustics and the propagation itself (e.g. Ekström, 2011; Ma et al., 2014; Moulik et al., 2021). We typically assume the first three terms are the same for both the reference and observed seismograms. In that case, the phase delay can be directly related to the propagation and thus perturbations in the Earth model. These phase delays are what we are interested in measuring here.

Waveforms are first pre-processed (e.g. resampled at 1 Hz, instrumental response removed if necessary). As suggested by Zhou et al. (2005) and Zhou (2009a), we then use a multi-taper technique to measure the phase-delays and to obtain an estimate of the measurement uncertainty (e.g. Thomson, 1982; Park et al., 1987a,b; Laske et al., 1994; Laske and Masters, 1996; Hjörleifsdóttir, 2007). The technique uses the first few Slepians (after Slepian, 1978) defined over a 801 s window. Slepians are an infinite series of functions with optimal frequency spectrum (therefore reducing frequency leakage) that weigh different parts of the waveform (thus reducing bias in the time-domain). With a 801 s-long time-window and 1 Hz sampling rate, we should use only the first 5 Slepians (see Percival and Walden, 1993, pp. 331). To position the Slepians, we compute the predicted group arrival time at the frequency of interest, starting the Slepian time window 150 s before the expected arrival. We then apply a 4 mHz-wide bandpass filter around the frequency of interest before we compute the Fast Fourier Transform. Finally, we subtract the phase component of the tapered and filtered observed (or SEM here) waveform from the reference waveform in the frequency domain. Usually, we obtain a smooth dispersion curve, except for when the phase delay reaches $\pm\pi$, where the dispersion curve makes jumps of $\pm 2\pi$. Low frequencies are less likely to suffer from cycle-skips. Therefore, we make our measurements at increasingly higher frequency, starting at 6 mHz. When we detect these so-called cycle-skips (we use a threshold of ± 4 radians for the detection), we add or remove 2π to obtain a smooth dispersion curve and apply this correction accordingly to all higher frequencies.

For each source-receiver pair, we end up with 5 dispersion curves for the 5 Slepians, corrected for cycle-skipping. We use the average of these 5 curves as our final measurements and the standard deviation as the data measurement uncertainty. In some cases, we note an inaccurate detection of cycles-skipping (either as false-positive or false-negative). These false detections typically do not occur on all five tapers, leading to a sharp increase in measurement uncertainty. In addition, some fundamental mode measurements are contaminated by the interference of other phases or higher modes. This usually does not affect all five tapers, thus also leading to an increase in the measurement uncertainty.

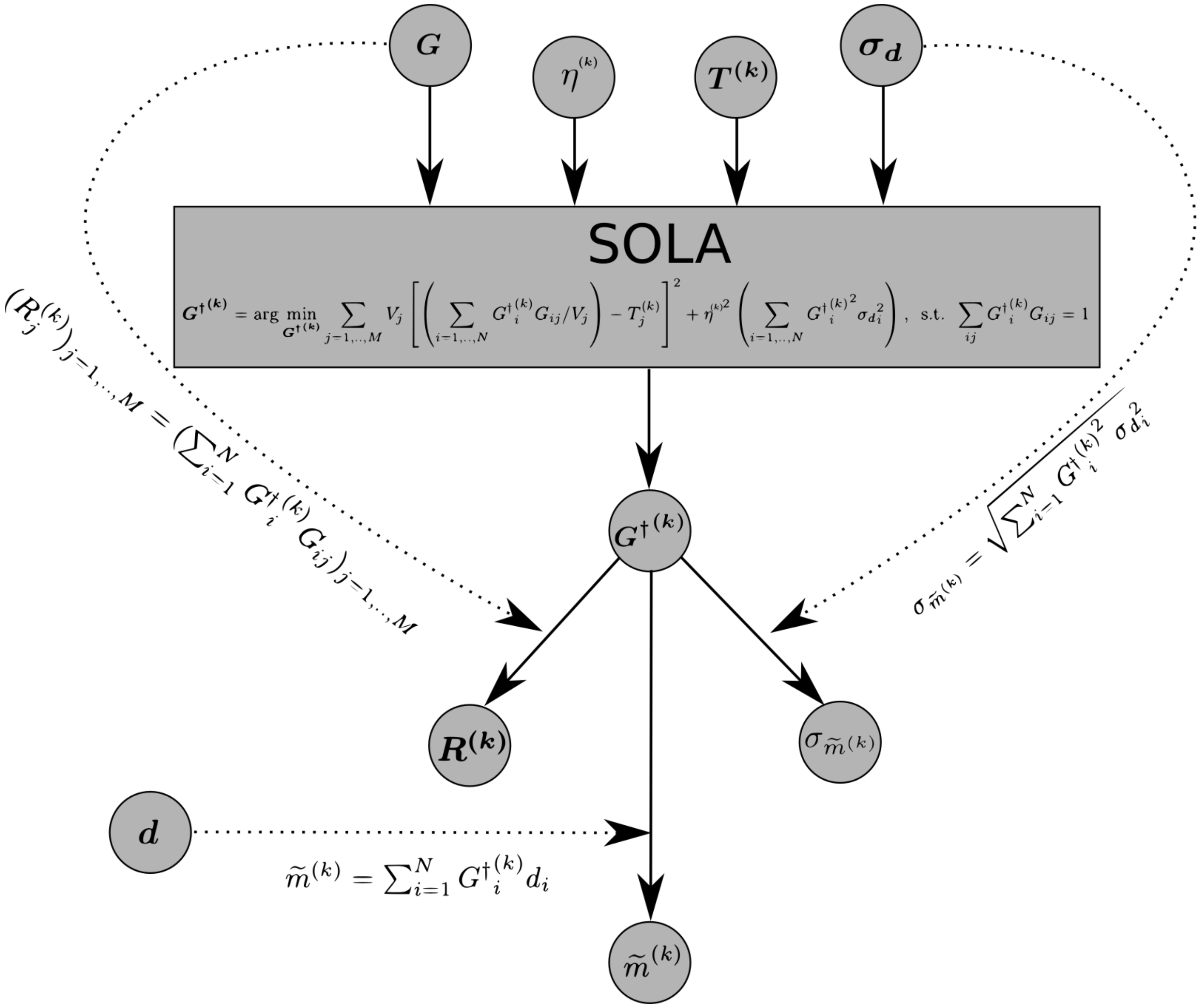


Figure 10 Illustration of the SOLA workflow. The minimisation problem at the heart of SOLA aims to find a generalised inverse matrix $\boldsymbol{G}^{\dagger(k)}$ such that the resolution is close to the target resolution and that the model uncertainty $\sigma_{\widetilde{m}}^{(k)}$ is reasonable. This minimisation problem takes four inputs: G, η^k, $\boldsymbol{\sigma_d}$ and $\boldsymbol{T^{(k)}}$. The sensitivity matrix $\boldsymbol{G}$ contains the forward theory and depends on the data geometry. The measurement uncertainties $\boldsymbol{\sigma_d}$ are estimated using the multitaper technique. For model parameter k, a target resolution is designed $\boldsymbol{T^{(k)}}$ and a trade-off parameter η^k balancing the fit to the target resolution and model uncertainty is chosen. The obtained generalised inverse allows us to compute the model uncertainty $\sigma_{\widetilde{m}}^{(k)}$ using the data uncertainty, to compute the averaging kernel $\boldsymbol{R^{(k)}}$ by combining the generalised inverse with the sensitivity matrix $\boldsymbol{G}$, and to compute the model parameter estimate $\widetilde{m}^{(k)}$ from the data values d. Note that the data values only play a role after the minimisation problem and that no a priori on the model estimate itself has been introduced. In this study, we set the measurement uncertainty to 1 as input into the SOLA minimisation problem. However, we incorporate the actual measurement uncertainty to compute the measurement model uncertainty.

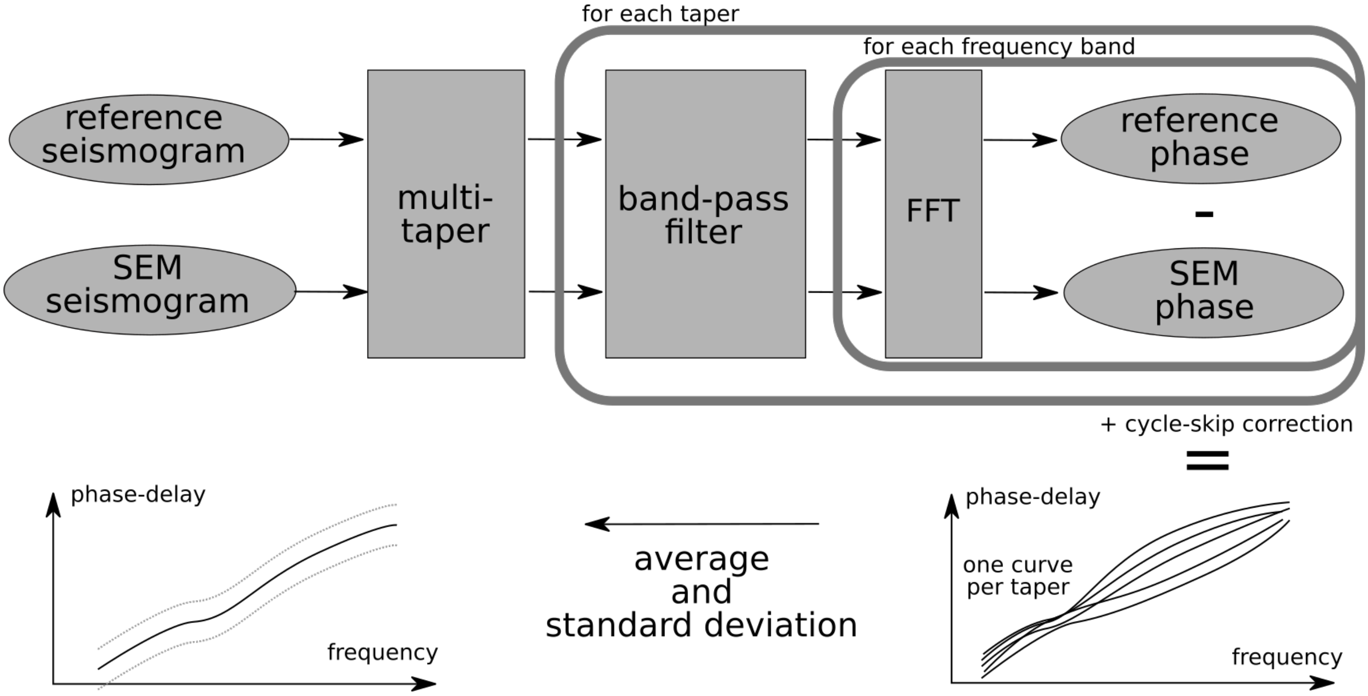


Figure 11 Overview of the measurement workflow. We compute a reference seismogram for the reference radial Earth model, which we use to measure the phase-delay of a SEM-computed seismogram (acting in this synthetic setup as observed seismogram). We apply a set of tapers (the five first Slepians), thus leading to 5 tapered traces. We filter each in a set of frequency bands, before we take the FFT. In the frequency domain, we then compute the phase difference for all frequencies for all tapers, producing a set of 5 dispersion curves. We apply a cycle-skip correction and then take the mean of all 5 tapers as the final measurement, with the measurement uncertainty given by the standard deviation of the five tapers.

Appendix C: Computational considerations

In this study, we use $N = 47\,700$ fundamental mode phase delays as data and we parameterise the spatial domain into $M = 259\,200$ voxels (cells of size $2^\circ \times 2^\circ$ laterally and 25 km depth for the first 400 km depth of the whole mantle). Therefore, the sensitivity matrix $\boldsymbol{G}$ of size $N \times M$ is reasonably large. To optimise the sparsity of the sensitivity matrix, we only consider the sensitivity kernels in the two first Fresnel zones laterally, since their amplitude is negligible further away. The sensitivity is also negligible at depths greater than 400 km depth. Our resulting matrix thus contains $645\,282\,622$ non-zero elements, i.e. the density is approximately 5.2%. The SOLA optimisation problem (Equation 5) leads to a set of normal equations taking the form of another $(M+1) \times (N-1)$ matrix $\boldsymbol{Q}$ that is less sparse than $\boldsymbol{G}$ (see Zaroli, 2016, Appendix A1). Reordering the lines of $\boldsymbol{G}$ with the sparsest row first helps to improve the sparsity of $\boldsymbol{Q}$. In this study, $\boldsymbol{Q}$ contains $657\,124\,288$ non-zero elements, i.e. sparsity is approximately 5.3%. On disk, we use a 'coordinate list' (COO) storage strategy, and $\boldsymbol{Q}$ takes up $\sim$17 GB. On RAM, we use a reversed linked-chain storage strategy to improve compute time. In this case, the $\boldsymbol{Q}$ matrix takes up $\sim$35 GB. This large memory requirement is the primary limiting factor for increasing the number of data and model parameters.

The computation time of the LSQR inversion for a single model parameter depends on the target resolution and trade-off parameter. With the choices made in this study, it takes $\sim$100 s per model parameter. As we invert for $69\,200$ model parameters, a full model estimate thus requires $\sim 692\,000$ s CPU time (or 192 CPUh). In practice, we invert for model parameters in parallel on several nodes with 128 CPU each using a multi-threading approach with OpenMP. The scaling is not fully linear due to input/output operations, but this strategy reduces the wall time to ~ 20 h.

Completed: 2025-05-27 10:49 AM
Recommendation: Accept Submission
Reviewer Comments

For author and editor

Dear authors,

Thank you very much for carefully addressing my comments! I very much value this discussion, and even though we may still disagree here and there, I am perfectly happy with the revised version of the manuscript. In my opinion, it can be published as it stands.

Best wishes

Andreas Fichtner

Completed: 2025-06-23 05:14 AM
Recommendation: Accept Submission
Reviewer Comments

For author and editor

The authors adequately addressed all of my concerns. The manuscript is now clearer and self-sustained.