Analysis of tomographic models using resolution and uncertainties: a surface wave example from the Pacific

Franck Latallerie, Christophe Zaroli, Sophie Lambotte, Alessia Maggi

submitted to *Geophys. J. Int.*

Analysis of tomographic models using resolution and uncertainties: a surface wave example from the Pacific

Franck Latallerie, Christophe Zaroli*, Sophie Lambotte, Alessia Maggi

Institut Terre et Environnement de Strasbourg, UMR 7063, Université de Strasbourg, EOST/CNRS, France

**E-mail: c.zaroli@unistra.fr*

4 January 2022

SUMMARY

Since most tomographic problems deal with imperfect data coverage and noisy data, an estimate of the seismic velocity in the Earth can only be a local average of the 'true' velocity with some attached uncertainty. We use the SOLA (Subtractive Optimally Localized Averages) method, a Backus–Gilbert-type method based on the resolution-uncertainty trade-off, to build a range of models of Rayleigh wave velocities in the Pacific upper mantle. We choose one solution and show how to analyse the model using its resolution and uncertainties. We exploit the model statistics to evaluate the significance of deviations from a theoretical prediction: a half space cooling model of the Pacific lithosphere. We investigate a slow velocity anomaly located north-east of Hawaii, at about 200 km depth, and a pattern of alternatively slow and fast velocity bands, aligned approximately north-west to south-east, between 200 and 300 km depth. According to our resolution and uncertainty analyses, both features seem to be resolved.

Key words: Seismic tomography; Inverse theory; Surface waves and free oscillations; Pacific Ocean.

1 INTRODUCTION

To understand how Earth's upper mantle behaves, we require robust estimates of its physical properties and how they vary laterally and with depth (e.g. Mégnin et al. 1997; Foulger 2011; Sim-

mons et al. 2019; Freissler et al. 2020). We can estimate one such property, S-wave velocity, using surface-wave tomography.

Most tomographic studies are performed by minimizing the squared differences between data predicted using the model and actual measured data, using one of many variations of the damped-least-squares inversion scheme (e.g. Maggi et al. 2006; Zhou et al. 2006; Nettles & Dziewoski 2008; Panning et al. 2010; Ekström 2011; French et al. 2013; Auer et al. 2014; Liu & Zhou 2016; Isse et al. 2019). The studies cited show consistent images of large scale structures (lithosphere cooling signatures, super-swells, cratons) and of structures that produce strong, localized S-wave velocity anomalies (subduction zones, ridges). These images differ, however, at shorter scales and for structures that produce weaker anomalies. For instance, the model savani of Auer et al. (2014) is 'slow' in the region of Hawaii at 200 km depth; on the contrary, this region is 'fast' in the model SEMum2 of French et al. (2013) (see for example fig. S7 from Isse et al. (2019) showing comparisons of various models).

Although some of these discrepancies may be imputed to differences in theoretical approaches (forward or inverse methods) used by the aforementioned studies, many discrepancies simply reflect how large uncertainties in the initial seismic measurements propagate into the final tomographic models. To compare tomographic models with each other – or to compare them to some prior, tectonic-based, conceptual model of the Earth – we need reliable estimates of their uncertainties (e.g. Rawlinson et al. 2014; Simmons et al. 2019; Freissler et al. 2020). Unfortunately, because many damped-least-squares inversion schemes are optimized to invert large volumes of data to constrain large numbers of model parameters (e.g. Debayle & Sambridge 2004), they often do not keep track of how data uncertainties propagate into the tomographic models; for large scale tomographic problems, most studies using these methods do not compute explicit model uncertainties, yet robust interpretation of features requires comparing the magnitude of the anomalies with the model uncertainties. For smaller scale tomographic problems, these methods only provide at best an *a posteriori* estimate of the uncertainties and cannot control them directly.

This difficulty derives from the manner in which earthquakes and seismic stations are distributed geographically: some regions may be sampled by many independent surface-wave ray

paths; other regions may be poorly constrained. This irregular data coverage causes the inverse problem to be under-determined: within poorly covered parts of a region, the model can change widely without changing how it fits the data. Which model we choose is, in some sense, arbitrary (Scales & Snieder 1997; Nolet 2008). Some tomographic studies make use of irregular or adapted meshes to account for the heterogeneous distribution of the data (e.g. Sambridge & Rawlinson 2005). Many tomographic studies use *ad-hoc* regularization constraints to reduce the non-uniqueness of the model solution; they often minimize the model's complexity by damping the inversion (e.g. Nolet 2008, Chapter 14) – hence the name *damped*-least-squares inversion. Such inversions may underestimate seismic velocity anomalies where data coverage is poor; less intuitively, they can also overestimate velocity anomalies where data coverage is highly uneven. In either case, we call such models *locally biased* (in the sense of 'averaging bias' effects, as reported by Zaroli et al. 2017).

Half-a-century ago, Backus and Gilbert proposed an original approach to invert linear geophysical data designed to produce complete uncertainty information, optimize local resolution, and eliminate bias (Backus & Gilbert 1967, 1968, 1970). Instead of searching for a model that fits the data with some smoothness criterion, the Backus-Gilbert method searches for a model that has optimal resolution given the geometry and uncertainties of the data. Pijpers & Thompson (1992, 1993) reformulated the method, applied it to helio-seismology, and named it *Subtractive Optimally Localized Averages* (SOLA). More recently, Zaroli (2016) introduced and adapted the SOLA method to seismic tomography problems with large numbers of parameters; he then applied this new inversion scheme to obtain a body wave tomographic model of the lower mantle. The SOLA Backus-Gilbert method still suffers from a form of non-uniqueness, as model resolution and model uncertainties trade-off against each other. Despite not being designed for that purpose, the method also leads to models that fit the data. Note that SOLA tomography can deal with heterogeneous spatial distribution of the data and usually produces models with no averaging bias. Computationally, the SOLA formulation of the Backus-Gilbert approach remains tractable even when dealing with large-scale tomographic problems (Zaroli 2016; Zaroli et al. 2017). Lastly,

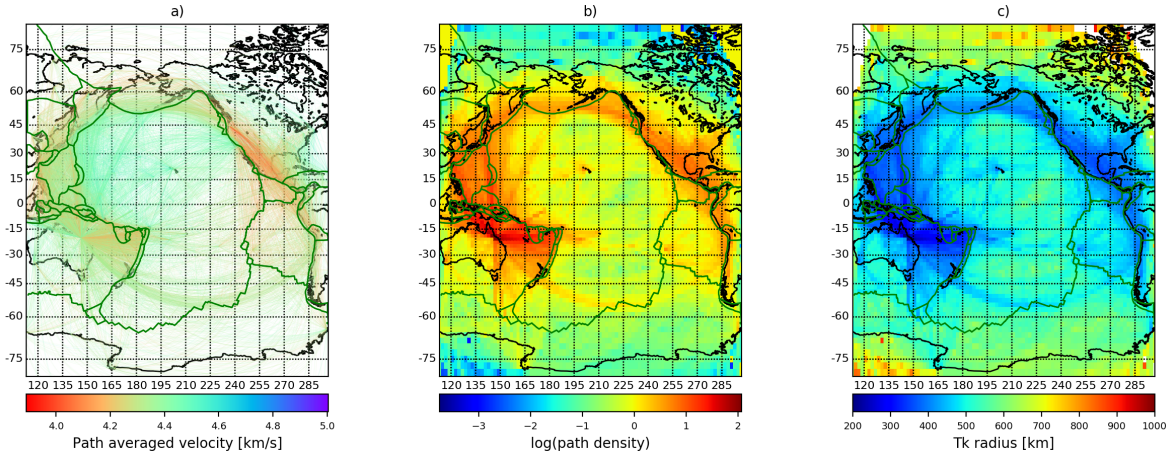

Figure 1. Data geometry and target resolution at 75 km depth. Thick green lines indicate plate boundaries. (a) Each thin line represents one of the ∼15 000 paths from Maggi et al. (2006) used as input to the SOLA tomography; the color of each path is the path-average shear wave-speed at 75 km depth in the 1D model corresponding to the path; (b) Path density (total path length that falls in each cell normalized by the total path length over the entire model); (c) Target-kernel radii.

it provides the full resolution and model uncertainties, which are the model statistics necessary to draw well-informed conclusions from the tomographic model.

In this study, we use the SOLA inversion to produce tomographic images of the Pacific upper mantle from path-averaged shear-wave velocity profiles obtained from surface-waves. After discussing how resolution and uncertainty trade-off in our models, we illustrate how to use them to perform meaningful comparisons of tomographic images with a geodynamic prediction. Taking as a reference a simple half-space cooling model (Parker & Oldenburg 1973), we point out anomalous regions in our tomographic model and we argue about their statistical significance and whether they are resolved or not.

2 DATA AND METHOD

Most surface-wave tomography studies that produce 3D shear-wave velocity models proceed in a sequence of two steps: either they first make 2D phase- or group-velocity maps at various frequencies (linear tomographic inversion), then combine them to obtain depth-dependent shear-wave velocity profiles at each point in the map (non-linear inversion) – examples are Ekström (2011) or Liu & Zhou (2016); or they first invert for 1D shear-wave profiles as a function of depth along each

source-station path (non-linear inversion), then combine these path-averaged velocity profiles into maps at each depth (linear tomographic inversion) – examples are Maggi et al. (2006) or Isse et al. (2019). Each sequence contains both a linear and a non-linear inversion. Only the linear inversion can be performed using the SOLA method.

2.1 Path-averaged velocity and uncertainty

As we aim to discuss how SOLA can improve the way we interpret seismic tomography images, we have chosen to apply the method on a pre-existing data-set of path-averaged shear-wave velocity profiles obtained from Rayleigh waves (Maggi et al. 2006). These profiles range from 50 km to 450 km depth, and were obtained by inverting over 56 000 multi-mode surface-waves whose paths crossed the Pacific Ocean, using the Debayle (1999) automation of the Cara & Lévêque (1987) secondary observables method. After obtaining a shear-wave velocity profile for each path, Maggi et al. (2006) gathered similar paths (those whose end-points were within 200 km of each other) into $\sim$15 000 clusters, obtained the average velocity profile of each cluster (Fig. 1a), and estimated the uncertainty of the velocities using the standard-deviation in each cluster. These clustered path-averaged shear-wave velocities are the input data for our SOLA inversion.

The robustness of the analysis later in this paper relies on the reliability of the data uncertainty estimates. The approach of Maggi et al. (2006) to estimate the data uncertainty captures the stochastic uncertainty: it measures the variability between adjacent paths within one cluster due, for example, to random measurement errors or to random errors in the location of the earthquakes in the cluster. However, it does not account for errors that would affect identically all paths in a cluster; for example, neglecting off-great circle paths due to some lateral heterogeneity or mislocating all earthquakes in the same direction due to lateral heterogeneities or network configuration. Data uncertainties from Maggi et al. (2006) decrease at larger depths because paths along which long-period and higher-mode data were unavailable were damped toward the background model, a smoothed version of PREM (Dziewonski & Anderson 1981). As all the path-averaged models tend toward the same values when the sensitivity decreases, the stochastic uncertainties also decrease.

Since the Maggi et al. (2006) approach considers only stochastic uncertainties, it underesti-

Table 1. χ^2_{red} values at various depths for the damped-least-squares model of Maggi et al. (2006).

depth [km]	50	75	100	150	200	250	300	350	400
χ^2_{red}	3.378	3.199	2.630	2.216	1.648	1.561	1.542	1.579	1.497

mates the full data uncertainties. Because the data uncertainties influence the SOLA inversion and propagate directly into model uncertainties (see section 2.2), we needed more reliable estimates of them. We calculated the χ^2_{red} for the final model of Maggi et al. (2006) (e.g. Nolet 2008):

$$\chi^2_{\text{red}}(\widetilde{m}) = \frac{1}{N}\sum_{i=1}^{N}\frac{[(\sum_{j=1}^{M} G_{ij}\widetilde{m}_j) - d_i]^2}{\sigma^2_{d_i}}, \tag{1}$$

where $\widetilde{m}_j$ is the j-th model parameter, M the number of model parameters, d_i the i-th datum, N the number of data, G_{ij} the elements of the sensitivity matrix of size $N \times M$ such that $\sum_{j=1}^{M} G_{ij}\widetilde{m}_j$ is the i-th predicted data, and σ_{d_i} the i-th data uncertainty. The values of χ^2_{red} at all depths are much greater than one (see Table 1). Two reasons may explain these values. First Maggi et al. (2006) may have chosen an overdamped solution to ensure that all features in their model were interpretable. Second, the data uncertainties may be underestimated because they do not account for systematic bias. To obtain more reliable data uncertainty estimates, we assume that the final model of Maggi et al. (2006) is a model that explains the data so that the χ^2_{red} values should be one, then we rescaled the data uncertainties by the factor required to bring $\chi^2_{\text{red}} = 1$, i.e. we multiplied each data uncertainty by $(\chi^2_{\text{red}})^{1/2}$ at the corresponding depth (see Table 1).

2.2 Regionalization with SOLA

We divide the Pacific region into independent depth layers, each of them divided into cells of 2° in latitude and longitude. Each model parameter m_j ($j = 1, ..., M$) represents the shear-wave velocity in cell j. Each datum d_i ($i = 1, ..., N$) represents the average shear-wave velocity along the i-th source-receiver path. The sensitivity matrix $\boldsymbol{G}$ linearly relates the data to the velocities as follows:

$$d_i = \sum_{j=1}^{M} G_{ij}m_j + n_i, \tag{2}$$

where n_i denotes a noise term (we assume uncorrelated zero mean Gaussian noise). The tomographic problem is ill-posed because some regions are under-sampled and contradictions may arise from the noise in the data. Therefore the linear sensitivity matrix $\boldsymbol{G}$ is not invertible and we have to seek a 'generalized' inverse matrix $\boldsymbol{G}^{\dagger}$, such that the k-th parameter estimate $\widetilde{m}_k$ can be written as a linear combination of the data:

$$\widetilde{m}_k = \sum_{i=1}^{N} G_i^{\dagger(k)} d_i = \sum_{i=1}^{N} \sum_{j=1}^{M} G_i^{\dagger(k)} G_{ij} m_j + \sum_{i=1}^{N} G_i^{\dagger(k)} n_i. \tag{3}$$

Here k refers to the k-th grid cell, $\boldsymbol{G}^{\dagger(\boldsymbol{k})} = (G_i^{\dagger(k)})_{i=1,..,N}$ represents the k-th row of the $\boldsymbol{G}^{\dagger}$ matrix (k may vary from 1 to M) and $\boldsymbol{R} = \boldsymbol{G}^{\dagger}\boldsymbol{G}$ is called the resolution matrix. We define the k-th averaging (or resolving) kernel as:

$$\boldsymbol{A}^{(\boldsymbol{k})} = (A_j^{(k)} = R_j^{(k)}/\mathcal{S}_j)_{j=1,...,M}, \tag{4}$$

where $R_j^{(k)} = \sum_{i=1}^{N} G_i^{\dagger(k)} G_{ij}$ are the elements of the k-th row of the resolution matrix. The factor $\mathcal{S}_j$ is, in our case, the cell surface area; it appears because the model parameterization has cells with different sizes.

If the problem were well-posed, the sensitivity matrix would be invertible, the resolution matrix would be the identity matrix, and the model estimate $\widetilde{\boldsymbol{m}}$ would equal the true model $\boldsymbol{m}$. In practice, the model estimates are local averages of the true Earth. The k-th row of the resolution matrix $\boldsymbol{R}^{(\boldsymbol{k})}$ (closely related to the k-th averaging kernel through the term $\mathcal{S}_j$) relates $\widetilde{m}_k$ to the true parameters ($\widetilde{m}_k \approx \sum_{j=1}^{M} R_j^{(k)} m_j = \sum_{j=1}^{M} \mathcal{S}_j A_j^{(k)} m_j$). In the Backus–Gilbert approach, we want each averaging kernel to be centered and peak-shaped around the location of its corresponding model parameter.

The model estimate $\widetilde{m}_k$ also contains the term $\boldsymbol{G}^{\dagger(\boldsymbol{k})}\boldsymbol{n}$ that reflects propagation of data noise into model space. Note that smaller averaging kernel sizes lead to higher model uncertainties (Menke 1989). This can be understood intuitively: fewer data can be averaged within smaller averaging kernels, leading to fewer chances for data errors to cancel out. There is, therefore, a trade-off between good resolution and low model uncertainties.

The model uncertainty is defined statistically as:

$$\sigma_{\widetilde{m}_k} = \sqrt{\sum_{i=1}^{N} {G_i^{\dagger(k)}}^2 \sigma_{d_i}^2}. \tag{5}$$

In this study we assume uncorrelated zero-mean Gaussian data uncertainty. Thus the data covariance matrix is diagonal: $\boldsymbol{C_d} = diag(\sigma_{d_i}^2), i = 1, ..., N$. Note that it would be possible to consider non-diagonal data covariance matrix in the SOLA approach (Pijpers & Thompson 1992) but we ignore off-diagonal terms for simplicity.

The key idea of the SOLA method is to specify an *a priori* target form, $\boldsymbol{T^{(k)}} = (T_j^{(k)})_{j=1,...,M}$, for each averaging kernel, $\boldsymbol{A^{(k)}}$. We specify some *a priori* information on the model resolution, which is fundamentally different from specifying *a priori* information on the model itself (as in data-fitting methods such as Bayesian methods and damped-least-squares). Rather than minimizing the spread (spatial extent) of each averaging kernel, $\boldsymbol{A^{(k)}}$, as in the original Backus-Gilbert approach, the SOLA variant aims to minimize the integrated squared difference between each averaging kernel, $\boldsymbol{A^{(k)}}$, and its corresponding target kernel, $\boldsymbol{T^{(k)}}$. A minimization problem directly computes each generalized inverse vector for each k-th parameter:

$$\begin{cases} \underset{\boldsymbol{G^{\dagger(k)}} \in R^N}{\arg\min} \ \sum_{j=1}^{M} \mathcal{S}_j (A_j^{(k)} - T_j^{(k)})^2 + \eta^2 \sigma_{\widetilde{m}_k}^2 \\ \text{s.t.} \sum_{j=1}^{M} \mathcal{S}_j A_j^{(k)} = 1, \end{cases} \tag{6}$$

where η is a trade-off parameter to give more weight to the resolution or to the model uncertainties in the optimization. Equation 6 is independent for each model parameter therefore efficient parallel computations are possible. We could adapt the trade-off parameter η for each model parameter but this would reduce the optimization of calculation; here we adapt only the target resolution. The constraint that the averaging kernels should be unimodular means that SOLA tomographic models should be free of averaging bias; a condition that cannot usually be guaranteed with inversion methods incorporating regularization (Zaroli et al. 2017).

The target resolution for a parameter is non-zero only in a circular region around that parameter's location; the circle's radius is computed from the path density:

$$r(\rho) = r_{max} - (r_{max} - r_{min}) \left(\frac{\log_{10}(\rho) - \log_{10}(\rho_{min})}{\log_{10}(\rho_{max}) - \log_{10}(\rho_{min})} \right), \tag{7}$$

where r is the target kernel radius, ρ is the path density, ρ_{min} and ρ_{max} are the minimal and maximal path densities, and r_{min} and r_{max} are the minimal and maximal target kernel radii. The logarithms lower the spatial variations of the target kernel radii compared to those of the path density. This smoothing is important to avoid tomographic images that would be difficult to interpret if nearby averaging kernels had widely different sizes. Other formulae could be chosen to reduce even further the spatial variability of the target kernel radii. Figs 1(b) and (c) show the path densities and target resolution length calculated for the path distribution in Fig. 1(a). We show several target kernels in Fig. 2(a).

The optimization problem (equation 6) involves the rows of the generalized inverse through the averaging kernels $\boldsymbol{A^{(k)}}$ (equation 4) and model uncertainties $\sigma_{\widetilde{m}_k}$ (equation 5). It leads to a set of independent equations whose unknowns are the rows of the generalized inverse. Derivation of these equations from equation 6 can be found in Zaroli (2016). These equations are solved using the LSQR algorithm of Paige & Saunders (1982) as proposed by Nolet (1985) and Zaroli (2016). Once the k-th row of the generalized inverse $\boldsymbol{G^{\dagger(k)}}$ has been computed, one obtains the associated model solution $\widetilde{m}_k$ from equation 3, the model uncertainty $\sigma_{\widetilde{m}_k}$ from equation 5, and the averaging kernel $\boldsymbol{A^{(k)}}$ from equation 4. For further details, the reader is referred to Zaroli (2016), Zaroli et al. (2017), and Zaroli (2019).

3 RESULTS

In this section, we present the effects of varying the trade-off parameter η in equation 6, choose a value for η, describe in more detail the features of the resulting tomographic model, and illustrate how to take resolution and uncertainty into account when comparing a tomographic model to the theoretical predictions of a simple geodynamic model (half-space cooling).

3.1 Resolution, uncertainty, and model estimate for various trade-off parameters

All tomographic inversions have trade-offs. In damped-least-squares inversions, model smoothness trades-off against data fit (smooth models fit the data poorly, but rough models risk fitting the noise part in the data). In SOLA, model resolution trades-off against model uncertainty. Fig. 2

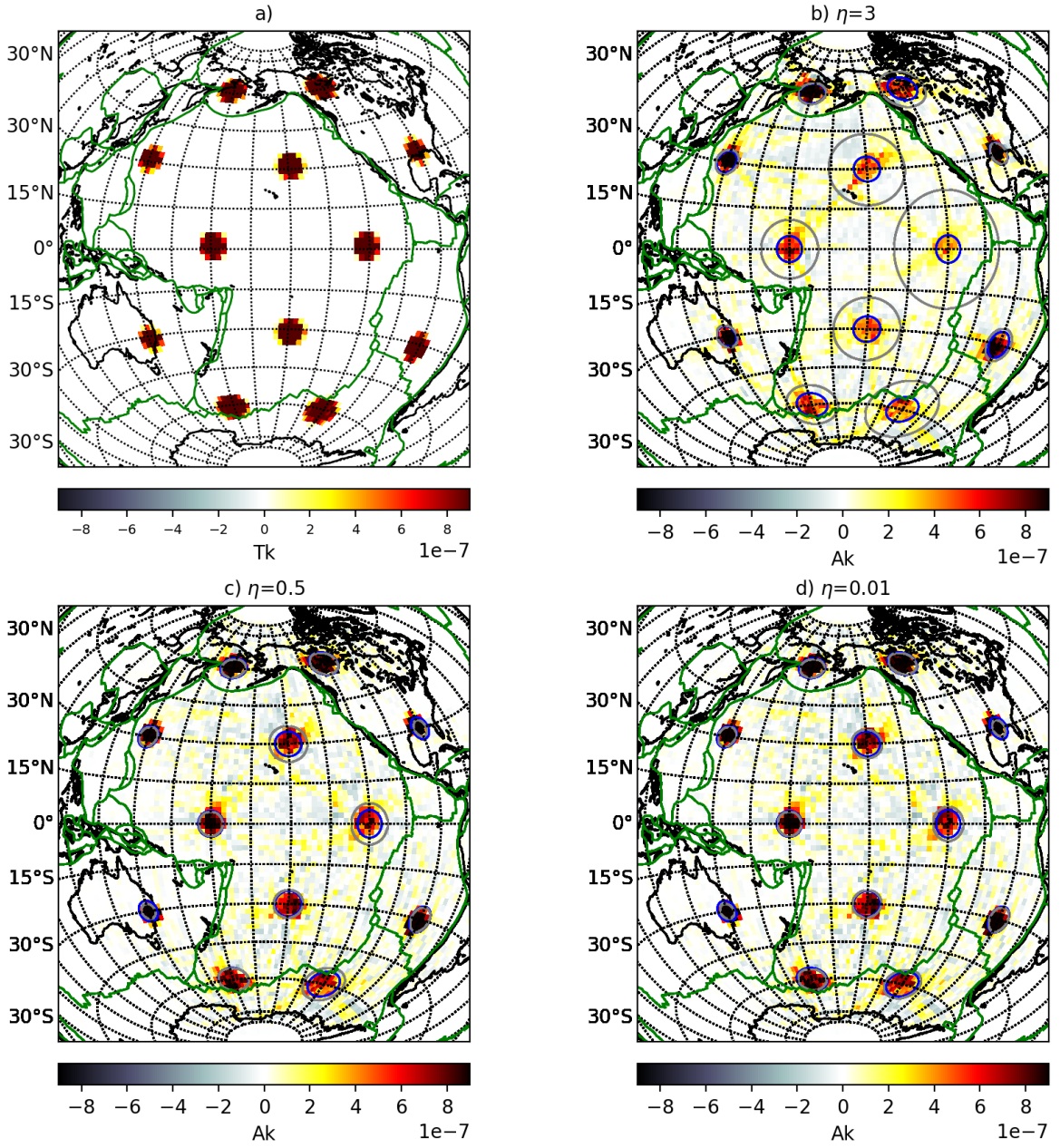

Figure 2. Examples of SOLA target kernels $\boldsymbol{T}^{(k)}$ (a) and corresponding averaging kernels $\boldsymbol{A}^{(k)}$ (b-d) at 75 km depth for η=3, 0.5 and 0.01. Blue circles indicate the edge of the target kernel disks; grey circles contain 68% of the averaging kernel amplitude. For convenience, several target (or averaging) kernels are plotted on the same map.

shows a selection of target kernels $\boldsymbol{T}^{(k)}$ and averaging kernels $\boldsymbol{A}^{(k)}$ for three values of the trade-off parameter η. The averaging kernels have dome-like centers and ramified extremities and are better focused for lower trade-off parameter values. They look like the results of the synthetic spike or point-spread tests used in some tomographic studies to approximate the resolution of their models

(e.g. Rawlinson & Spakman 2016). For a trade-off that favors low model uncertainties but poor resolution, Fig. 2(b), the ramifications of the averaging kernels extend far from the model parameter location, their magnitudes are non negligible, and they follow the non-uniform azimuthal path coverage. This is the well-known smearing effect, visible in the averaging kernel in the north-east of Hawaii, where paths have a predominant north-east to south-west orientation. In regions where the azimuthal path coverage is better, we do not see such artifacts. This is the case, for example, in the north-east of Japan where the averaging kernels are compact and nearly circular. Improving the resolution by lowering η makes the averaging kernels more circular and decreases the length and magnitude of the ramifications. The averaging kernels may have a small negative component, which is physically meaningless. A strong negative component would indicate that the target resolution was badly chosen and invalidate the results; in our case, it is small enough to be ignored.

A complete set of averaging kernels – one map for each model parameter – would fully represent the resolution of a SOLA inversion, but would be cumbersome and difficult to use when interpreting tomographic models. To simplify this information, we compute for each model parameter k a *resolution length* L_k corresponding to the radius of a circle that contains 68% of the averaging kernel – a proxy to the standard deviation for a 2D Gaussian (gray circles in Fig. 2). This proxy to the resolution is not ideal if the averaging kernel has a very complex shape, especially in case of smearing. For example, the gray circles on Fig. 2(b) do not represent well the averaging kernels when the ramifications are long and strong (especially for those in the middle of the Pacific plate). Note that other simplifications of the resolution could be used; for instance we could fit an ellipse to the averaging kernels to obtain a main direction in case of anisotropic resolution (smearing). However, since our averaging kernels are mostly circular, we use the circular approximation for simplicity. The top row of Fig. 3 shows resolution maps for three values of the trade-off parameter. In all maps the resolution is bad at the model borders and in the central Pacific Ocean, where path coverage is poor. As the trade-off parameter decreases, resolution improves in the central Pacific, though the improvement seems to level off between $\eta = 0.5$ and $\eta = 0.01$.

Before choosing a value for η, we must also examine the influence on the model uncertainties. Note that in this paper we express the uncertainties in the 1σ sense: there is a 68% probability

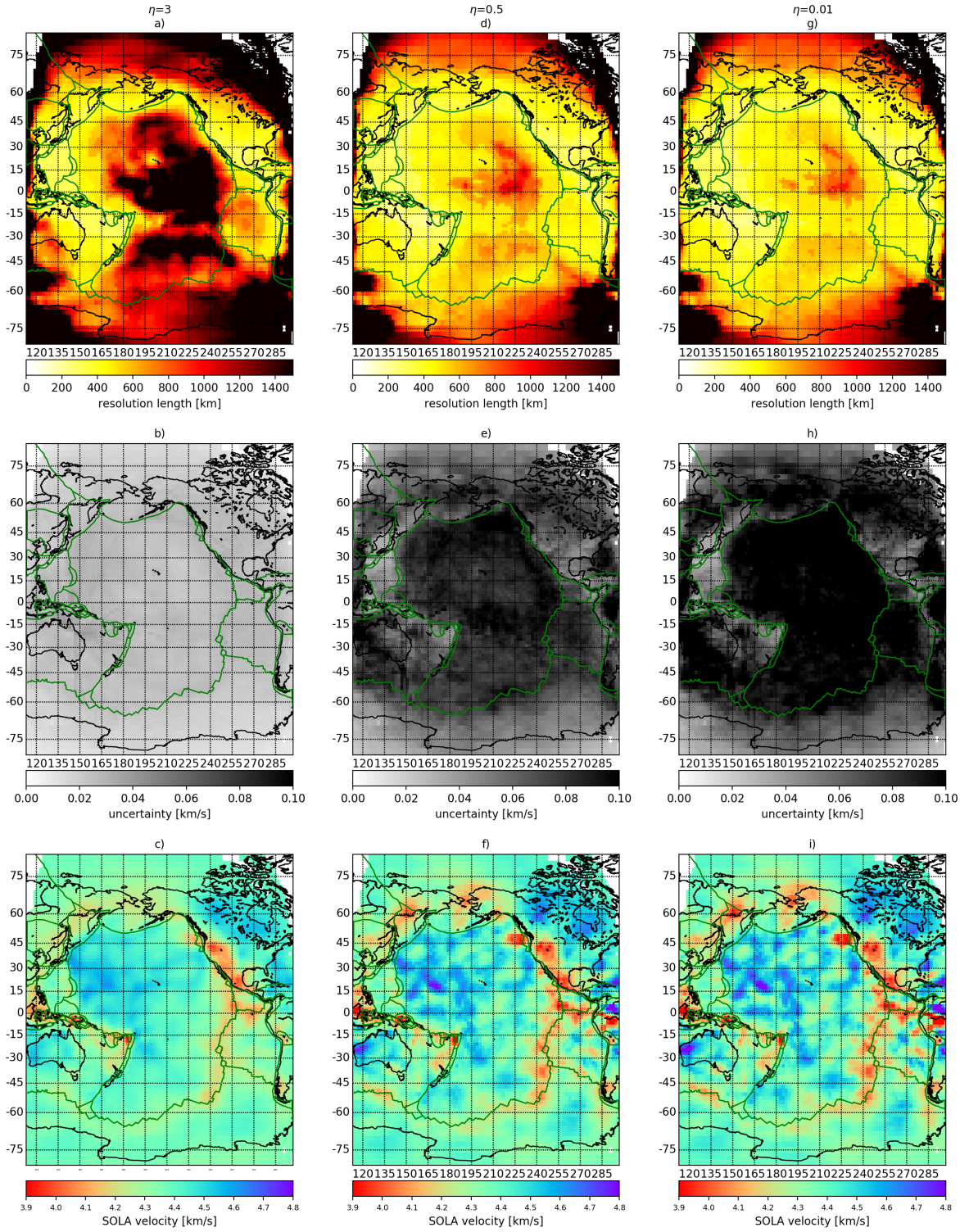

Figure 3. SOLA inversion results at 75 km depth for three values of trade-off parameters: η=3, 0.5, 0.01. Each column corresponds to a value of η and from top to bottom, maps show resolution lengths (L_k) (a, d, g), model uncertainties (b, e, h) and shear-wave velocities (c, f, i).

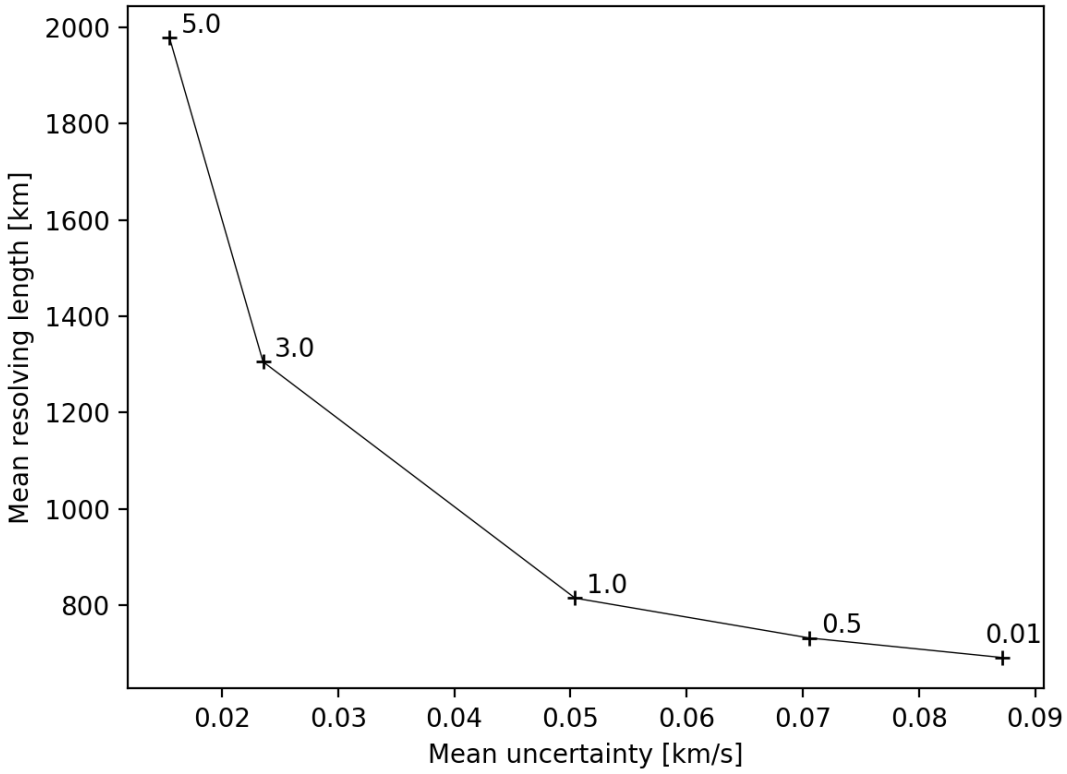

Figure 4. Mean resolution versus mean uncertainty of SOLA results for five values of the trade-off parameter η (5, 3, 1, 0.5, 0.01) indicated on the curve. The slope from $\eta = 3$ to $\eta = 0.5$ is more negative than from $\eta = 0.5$ to $\eta = 0.01$. This is consistent with the apparent variations of the model statistics on Fig. 3. As η changes, model resolution and uncertainties follow an L-curve.

that the true locally averaged shear-wave velocity lies within the uncertainty interval around the estimated value. The middle row of Fig. 3 shows model uncertainties for the same three values of η. As predicted by the nature of the trade-off, model uncertainties worsen as the trade-off parameter decreases, whereas model resolution improves. Decreasing η from 3 to 0.5 improves resolution greatly while increasing model uncertainties only slightly; however, decreasing η again from 0.5 to 0.01 improves resolution only slightly while increasing model uncertainties greatly. As η changes, model resolution and uncertainties follow a well-known L-curve (see Backus & Gilbert 1970). Fig. 4 shows mean resolution ($\frac{1}{M}\sum_{k=1}^{M} L_k$) versus mean uncertainty ($\frac{1}{M}\sum_{k=1}^{M} \sigma_{\widetilde{m}_k}$) for five values of the trade-off parameter, including 3, 0.1 and 0.01.

The bottom row of Fig. 3 shows tomographic models at 75 km depth for the same three η values

discussed above. As expected, the tomographic models are smoother where resolution lengths are large and uncertainties low, and rougher where resolution lengths are low and uncertainties high. No single model is intrinsically better than another: the poor resolution, low uncertainty model in Fig. 3(c) would be appropriate if we were interested in high precision estimates of the average shear-wave velocity over large patches; the better resolution, higher uncertainty models in Fig 3(f) or (i) would be appropriate if we were interested in smaller anomalies we expect to be strong enough to be detectable despite the increased uncertainties.

In the next subsection, we describe in more detail the $\eta = 0.5$ model, whose resolution lengths are on average short enough that the model is not too smooth, and whose uncertainties are on average low enough not to swamp all its features.

3.2 Detailed analysis of the resolution, uncertainty and model estimate at a given trade-off

For any tomographic inversion, resolution, model uncertainty, and model smoothness are intrinsically linked to path coverage and data quality. The best-sampled regions for our inversion are the eastern, northern and western borders of the Pacific ocean (Fig. 1): these regions are strongly seismogenic and host many seismic stations. The interior of the Pacific ocean and its southern boundary are sparsely sampled, except near Hawaii and French Polynesia where Maggi et al. (2006) exploited both permanent stations and temporary deployments (Barruol 2002) to increase local path coverage.

The resolution map in Fig. 3(d) shows that the well-sampled regions have the best lateral resolution: the radii of the circles that approximate the averaging kernels in these regions are between 300 km and 500 km. The Pacific ocean itself has a poor resolution, around 800 km on average. The regions west of Hawaii and west of the French Polynesia have better resolution (550 km) because the seismic networks installed within those archipelagos record earthquakes that occur on the western boundary of the Pacific plate. The best resolved region is northeast Australia, where the resolution length is below 300 km; the worst resolved region away from the model's edges is located southeast of Hawaii, where the resolution length is greater than 1000 km.

But the resolution is not only related to the path coverage: the data uncertainties also influence the achievable resolution and model uncertainties.

The uncertainty map in Fig. 3(e) shows that model uncertainties and model resolution do not always correlate. In some well-resolved regions, such as the Philippine Sea, Tonga-Kermadek, and central America, the locally-averaged shear-wave velocities are known to within 0.02 km/s. In these regions, we can resolve and interpret small features, even those that generate only moderate shear-wave velocity signatures. In some poorly-resolved regions, such as the central Pacific Ocean, shear-wave velocities are known to within greater bounds (0.07 km/s on average); here it becomes harder to interpret features unless they are large and generate strong shear-wave velocity signatures. Well-resolved regions may also display high model uncertainties: in the northern Pacific and off the coast of north-east Australia, the resolution length is close to 300 km, yet the uncertainties are similar to those in the central Pacific. In such regions, we can interpret features, regardless of their size, only if they generate strong shear-wave velocity signatures. This is a reminder that we need both resolution and uncertainty information for interpreting tomographic models.

The shear-wave velocity maps at 75 km depth (Fig. 3f), 125 km depth (Fig. 5c), and 400 km depth (Fig. 5f) show all the large-scale geological features we may expect in tomographic images in oceanic regions: low shear-wave velocities down to $\sim$100 km depth beneath the East Pacific Rise and the Pacific-Antarctic Ridge; fast velocities in the subduction zones (at least down to 200 km depth) and lower velocities in their back-arc regions; high velocities from 75 to $\sim$200 km depth in the North American and Australian cratons; velocities that increase with distance from the mid-ocean ridges down to depths of $\sim$150 km. Unsurprisingly, given we have used their shear-wave velocity profiles, Maggi et al. (2006) also saw similar features, though with slightly poorer resolution, as their damped-least-squares tomographic inversion used a single smoothing parameter for all locations in the model.

The resolution and uncertainty maps at 125 km and 400 km depth (Fig. 5) look like those at 75 km depth (Figs 3d and e). This is not what we would expect given the behaviour of surface waves. Deeper regions of the upper mantle can only be resolved by longer-period or higher-mode surface-waves, both of which have long wavelengths (so poor lateral resolution) and cannot be

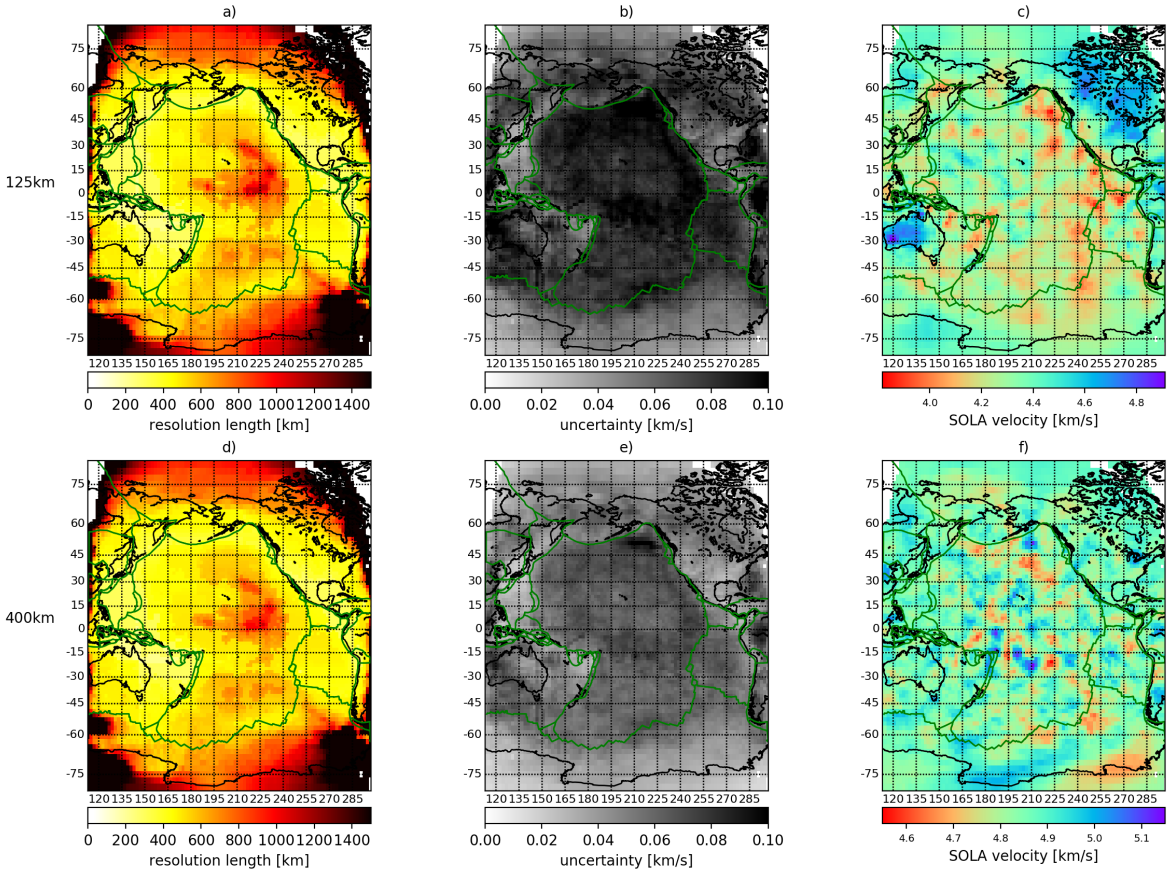

Figure 5. SOLA inversion results at 125 km (first row) and 400 km depth (second-row) for $\eta = 0.5$: (a, d) resolution lengths; (b, e) model uncertainties; (c, f) shear-wave velocities.

observed clearly on all paths; therefore, we would expect the resolution lengths to increase with depth. The reason those in Fig. 5 stay constant stems from the data we used in the SOLA inversions: path-averaged shear-wave velocity profiles from Maggi et al. (2006), each one the result of a non-linear inversion of surface-wave measurements. Where long-period and higher-mode data were available, shear-wave velocity profiles were constrained by data down to 400 km depth; where these data were unavailable, the profiles were damped, at depth, towards the background model: a smoothed version of PREM (Dziewonski & Anderson 1981). It is likely, therefore, that many of the velocity profiles used for the SOLA inversion are uninformative below $\sim$300 km depth; we could not remove them from our inversion without redoing Maggi et al. (2006)'s entire analysis, which was outside the scope of this study. In the following, therefore, bear in mind that the resolution maps (and for a similar reason the uncertainty maps) in Fig. 5 are overly optimistic at depths beyond $\sim$300 km.

3.3 Comparing tomographic images with theoretical predictions

The tomographic maps of Fig 3 show an increase in seismic velocity with distance from the mid-oceanic ridges. This is connected with the well-known phenomenon of plate cooling (e.g. Ritzwoller et al. 2004; Faul & Jackson 2005; Priestley & Mckenzie 2006; Isse et al. 2019). Armed with complete resolution and uncertainty information, we investigated if any features of these tomographic images deviated significantly from the predictions of a theoretical cooling model and if such deviations were well-resolved.

3.3.1 The reference model

We chose to use the simple half-space cooling model of Parker & Oldenburg (1973) to illustrate our process of comparing tomographic images to theoretical predictions, not because we believed it to be the best thermodynamic fit to the ocean-cooling problem, but because it has the fewest adjustable parameters. Even with an unsuitable reference model, the analysis would still be relevant as it would simply lead us to state that the tomographic velocities differ significantly from the reference model.

The half-space cooling model describes the lithosphere as an infinite half-space of thermal diffusivity k that cools from a starting temperature Θ_m. It predicts the temperature of the lithosphere as a function of age t and depth z from the following expression:

$$\Theta(t,z) = \Theta_m \mathrm{erf}\left(z\sqrt{\frac{1}{4kt}}\right), \tag{8}$$

where we take $\Theta_m = 1300°C$ and $k = 1.10^{-6}m^2s^{-1}$, as suggested by Ritzwoller et al. (2004). We turned this expression into temperature maps at each depth by relating age to geographic position using the age model of Müller et al. (2008), then projected them onto our tomographic grid: the predicted temperature in cell j at depth z is given by $\Theta_j(z) = \Theta(t_j, z)$, where t_j is the average age of cell j.

We then had to relate temperatures predicted as a function of position with tomographic images that contained velocity estimates representing finite-resolution local averages. In order to make meaningful comparisons, we needed to account for this resolution and consider equivalent local

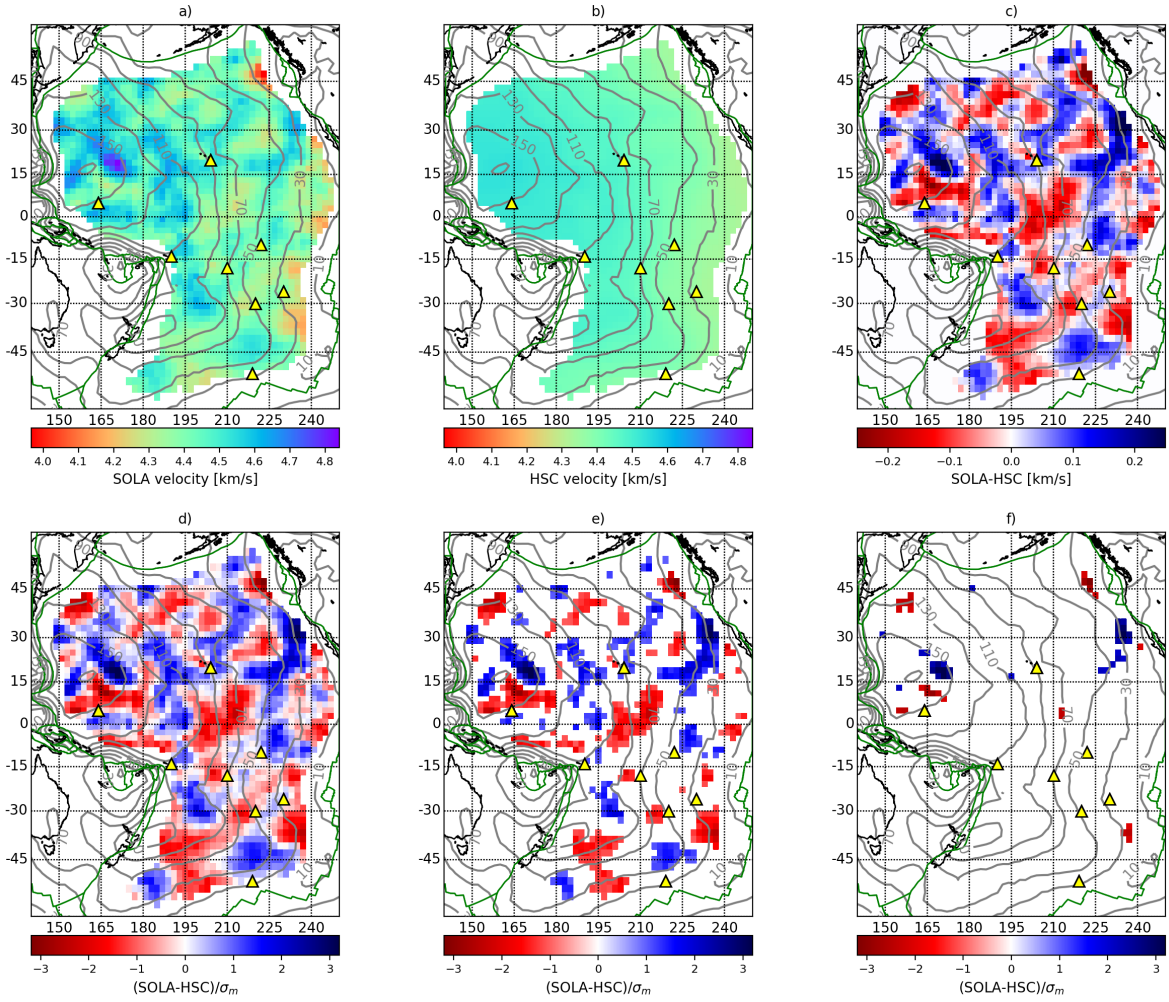

Figure 6. Comparison at 75 km depth between the SOLA tomographic model and the predictions of the half-space cooling (HSC) model from Parker & Oldenburg (1973) on the Pacific plate, excluding ridges and subduction zones. a) shear-wave velocities from SOLA; b) shear-wave velocities predicted for the HSC model; c) difference between SOLA and HSC; d) difference between SOLA and HSC scaled by the tomographic uncertainties σ_m; e) and f) same as d) but with masks to remove deviations smaller than $\pm 1\sigma_m$ or $\pm 2\sigma_m$ respectively. Grey contour lines indicate lithosphere age from Müller et al. (2008) and yellow triangles locate hot-spots from Courtillot et al. (2003).

averages of the predicted temperatures rather than the temperatures themselves. We obtained these local temperature averages using the resolution matrix : $\widetilde{\Theta_k}(z) = \sum_{j=1}^{M} R_j^{(k)}(z)\Theta_j(z)$, a procedure known as *tomographic filtering* (e.g. Ritsema et al. 2007; Simmons et al. 2019).

In order to convert the locally-averaged predicted temperatures to shear-wave velocities, we needed to make some assumptions about the relationship between these two physical quantities that were at least partially supported by rock physics. We assumed that, in regions distant from

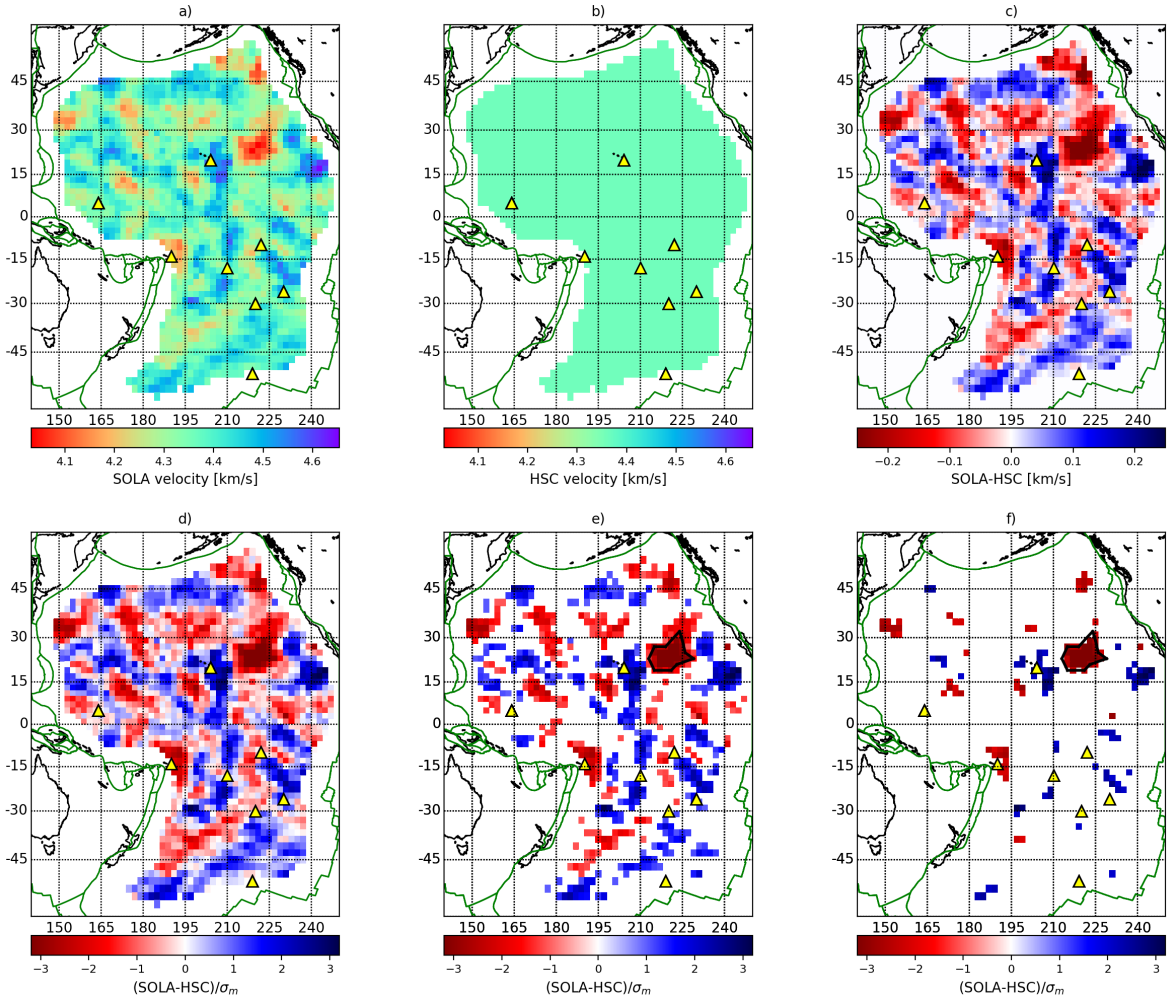

Figure 7. Same as Fig. 6 but at 200 km depth (note that at this depth the HSC model shows no age-dependence). The black line indicates the contour of the anomaly located to the north-east of Hawaii.

both ridges and subduction zones, mantle temperature and shear-wave velocities were linearly related (e.g. Chen et al. 1996; Foulger 2011, section 5.1.2 pp. 147). We therefore performed a linear regression between the locally-averaged temperature maps and our tomographic images to obtain locally-averaged predicted shear-wave velocities. In the following, we will refer to this predicted shear-wave model as the HSC (Half Space Cooling) reference model.

Figs 6(a) and (b), 7(a) and (b), and 8(a) and (b) show comparisons between the SOLA tomography results and the HSC model predictions at 75 km, 200 km, and 275 km depth on the Pacific plate. Unsurprisingly, the HSC predictions look like the long-wavelength component of the tomographic images. The lithospheric cooling signature is visible at 75 km depth, but absent below the deepest extent of the oceanic lithosphere ($\sim$150 km).

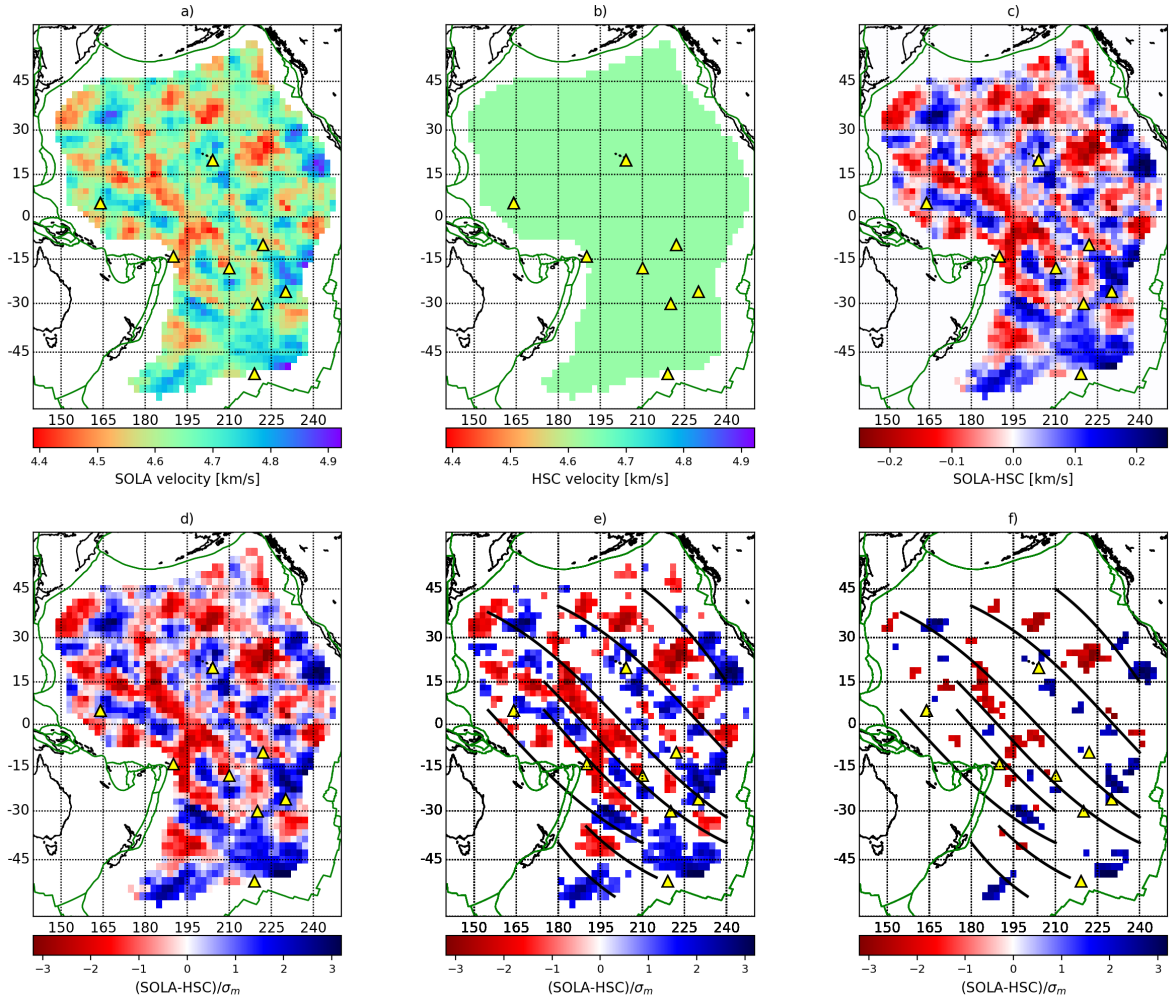

Figure 8. Same as Fig. 6 but at 275 km depth. The black lines indicate low velocity anomalous bands (the lines follow approximately the zeroes between anomalous bands).

3.3.2 Deviations from the reference model

Figs 6(c), 7(c), and 8(c) show the differences between the tomographic model and the HSC reference model at 75 km, 200 km, and 275 km depth. Red anomalies correspond to areas where the tomographic model is slower than HSC; blue anomalies correspond to areas where the to-

Table 2. Proportion of points remaining after masking values within the $1\sigma_m$ or $2\sigma_m$ error bar(s) at various depths. For a normal distribution, we would expect proportions of 0.32 for $1\sigma_m$ and 0.05 for $2\sigma_m$.

Depth (km)	50	75	100	150	200	250	300	350	400
For $1\sigma_m$	0.32	0.33	0.33	0.33	0.38	0.41	0.46	0.49	0.43
For $2\sigma_m$	0.06	0.03	0.03	0.07	0.09	0.10	0.14	0.13	0.12

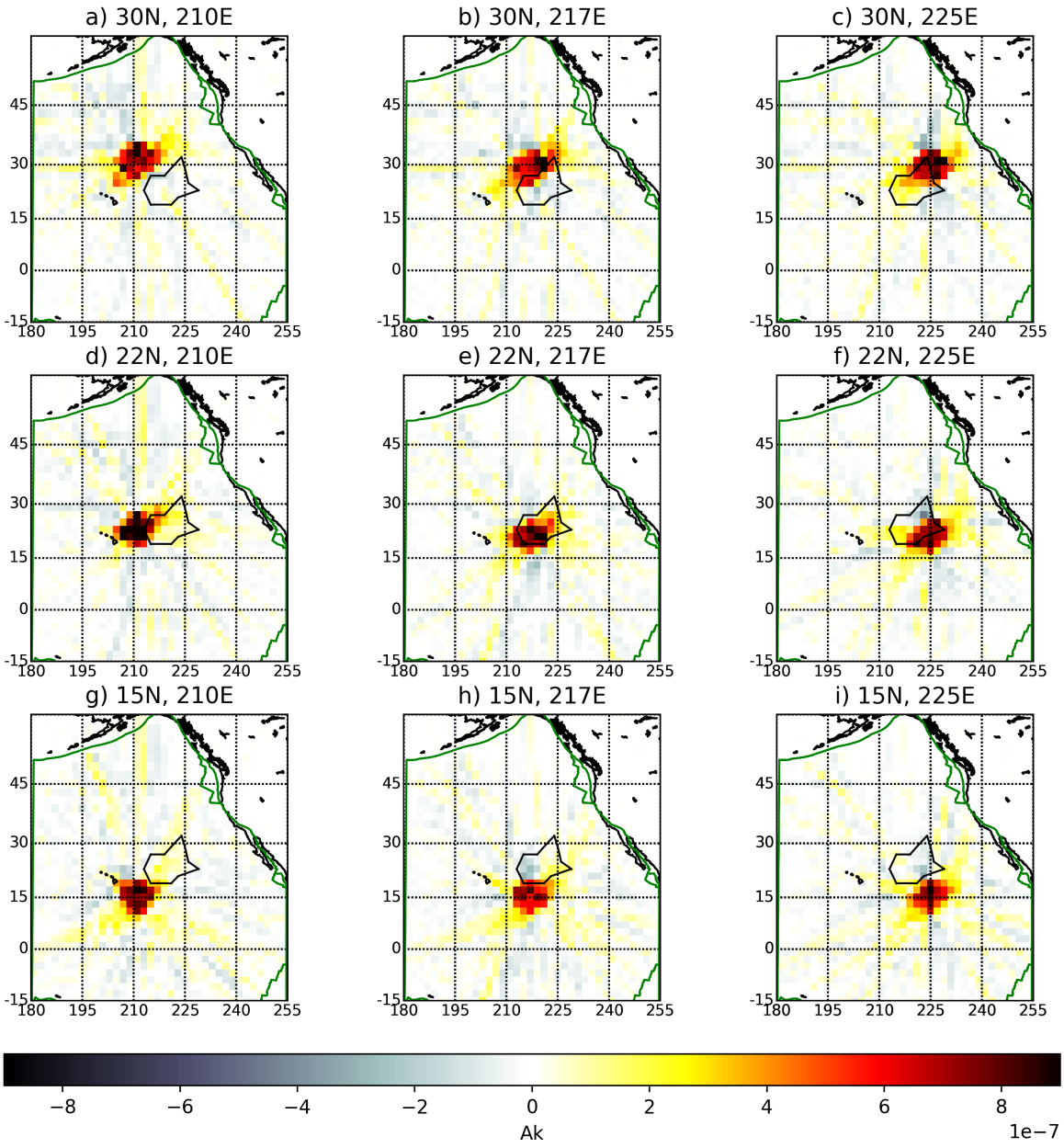

Figure 9. Selection of averaging kernels around the slow velocity anomaly located to the north-east of Hawaii (see Fig. 7) at 200 km depth. The black line represents the anomaly contour.

mographic model is faster than HSC. We used the model uncertainties to assess the significance of these anomalies. Panels (d) of the same figures show the deviations from the reference model scaled by the uncertainties. A region could be considered anomalous with a confidence threshold of 68% if the velocity difference with respect to the reference model exceeds the tomographic uncertainties by more than $\pm 1\sigma_m$ (panels (e) of the same figures). A region could be considered anomalous with a confidence threshold of 95% if the velocity difference with respect to the ref-

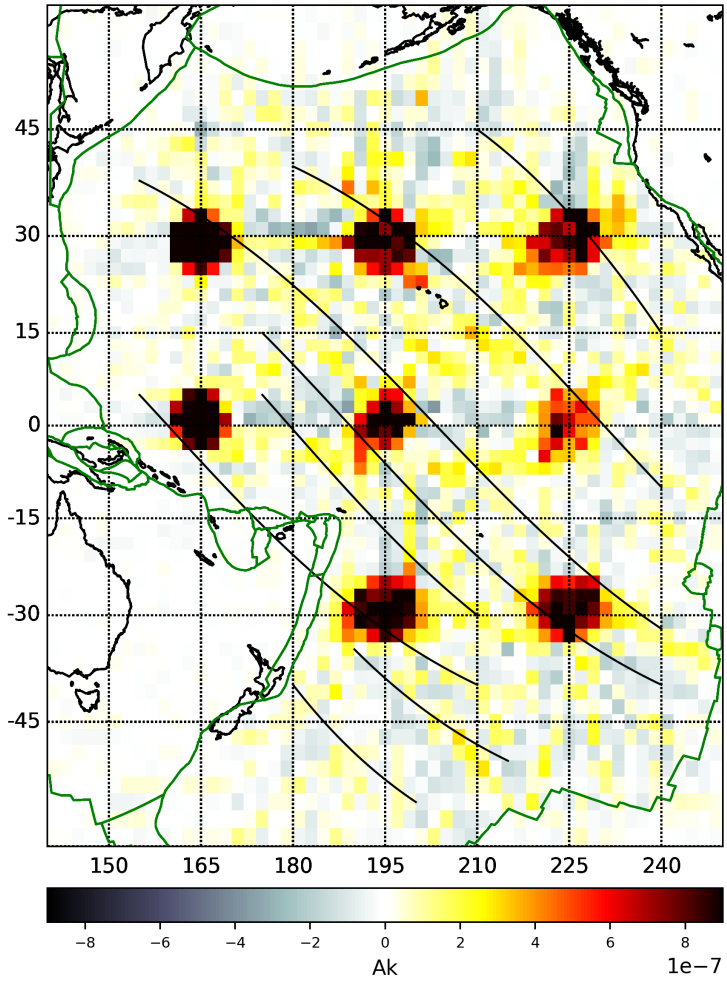

Figure 10. Selection of averaging kernels in the Pacific at 275 km depth. The black lines represent the pattern of anomalous velocity bands observed on Fig. 8. The averaging kernels are located at: (30N, 165E), (30N, 195E), (30N, 225E), (0N, 165E), (0N, 195E), (0N, 225E), (-30N, 165E), (-30N, 195E), (-30N, 225E). For convenience, we plot the averaging kernels on the same map.

erence model exceeds the tomographic uncertainties by more than $\pm 2\sigma_m$ (panels (f) of the same figures). Since the estimation of the data uncertainties could be improved, so could the model uncertainties, we present $\pm 1\sigma_m$ and $\pm 2\sigma_m$ maps to appreciate a range of uncertainties.

However, just identifying anomalies as exceeding $\pm 1\sigma_m$ or $\pm 2\sigma_m$ is not enough to declare them *significant*, because even if the Earth were in reality identical to the prediction of the half-space cooling model, we would still expect 32% of points in a tomographic model with uncertainties σ_m to exceed $\pm 1\sigma_m$ and 5% of them to exceed $\pm 2\sigma_m$. We could be justified in declaring anomalies to be *significant* only if more points than expected exceed the $\pm 1\sigma_m$ and $\pm 2\sigma_m$ thresh-

olds, or if these points organized geographically in coherent regions and these anomalous regions could indeed be resolved by the tomography (anomalies larger than the resolving lengths). This definition of significance is stricter than the one used in most tomographic studies, including Maggi et al. (2006), and underlines the importance of correctly estimating the data uncertainties that feed into the estimates of σ_m.

Table 2 shows the proportion of points that remain in the tomographic models after applying the $\pm 1\sigma_m$ and $\pm 2\sigma_m$ masks: at lithospheric depths, the proportions of such points do not exceed greatly the expected ones (0.32 and 0.05); however, as depth increases, so do the proportions of points exceeding the $\pm 1\sigma_m$ and $\pm 2\sigma_m$ levels. In section 2.1 we discussed that the data-uncertainties from Maggi et al. (2006), and hence our estimates of σ_m, may have been underestimated especially below ~ 300 km. For this reason, we have limited our analysis of significant anomalies to depths shallower than ~ 300 km.

3.3.3 Significant anomalies

At lithospheric depths (75 km), the proportion of points still visible in the tomographic image after applying the $\pm 1\sigma_m$ and $\pm 2\sigma_m$ masks is not significant (see Table 2). What about the geographic distribution of these points? Figs 6(e) and (f) shows that the positions of unmasked low-velocity anomalous regions at 75 km depth do not correlate with hot-spot locations published by Courtillot et al. (2003), and that anomalies that exceed the $\pm 2\sigma_m$ thresholds are much smaller than the correlation lengths from Fig. 3(d). We therefore concluded that the lithosphere corresponding to the Pacific plate contains no significant shear-wave anomalies with respect to predictions made by the half-space cooling model, at least given the data set of Maggi et al. (2006).

At shallow asthenospheric depths (200 km), a greater proportion of points remains in the masked tomographic image (see Table 2 and Figs 7e and f). A strong low velocity anomaly appears to the north-east of Hawaii at 20°N 220°E. This anomaly of 12° ($\sim$ 1400 km) radius is visible between 125 km and 250 km depth, and is about 0.25 km/s slower than the half-space cooling model (stronger than -3σ_m). Not only is the anomaly strong, it is also larger than the local resolution length of 750 km (Fig. 5a). Fig. 9 shows individual averaging kernels within and around

the anomaly: the kernels are all of similar size or smaller than the anomaly and are all well focused, indicating absence of smearing. Given the amplitude and size of the anomaly with respect to the tomographic model's uncertainty σ_m and its resolution length, we considered it to be *significant*. We need to carefully analyse averaging kernels when interpreting the size and shape of anomalies. Since the averaging kernels are spread in space, the velocity anomalies tend to appear larger than they really are. The shape of the anomaly in the tomographic image can also be misleading. For example the anomaly to the north-east of Hawaii shows two bumps, one to the north and one to the east, which may be due to the complex shape of the averaging kernels, or to data noise propagated into the model. Only features outside the uncertainty range with size greater than the resolution should be interpreted.

Deeper in the asthenosphere, a pattern of alternatively slow and fast velocity bands appears, oriented approximately south-east to north-west (Fig. 8). The bands are approximately 15° wide (∼1700 km), 70° (∼8000 km) long, and are visible between 275 km and 400 km depth. The bands seem to follow the absolute plate motion of the Pacific plate (e.g. Gripp & Gordon 1990). They also resemble the bands observed by French et al. (2013) at depths between 200 km and 350 km and the low anisotropy channels observed at 100 km depth by Montagner (2002). The width and length of these bands are greater than the resolution length. To exclude possible smearing, always an issue when tomographic images show elongated features, we have shown several averaging kernels in Fig. 10. Given the amplitude and size of the highly-correlated bands and the focused nature of the averaging kernels, we again considered these bands to be *significant* features of the tomographic model.

4 DISCUSSION

We have produced a tomographic model of shear-wave velocities in the Pacific upper mantle using the SOLA Backus-Gilbert method (Zaroli 2016; Zaroli et al. 2017), and have shown how to exploit the full model resolution and uncertainty information to evaluate the true significance of deviations from a theoretical prediction.

4.1 Model statistics – a rare commodity

Several authors have inverted surface-wave data to obtain velocity models of the upper mantle, either globally or in the Pacific, using different forward theories and inversion methods (e.g. Zhou et al. 2006; Ekström 2011; French et al. 2013; Liu & Zhou 2016; Isse et al. 2019). Although they agree with each other at large scales, they often display different local features. These discrepancies may be partly explained by the differing data-sets (fundamental-mode Rayleigh waves, multi-mode Rayleigh waves, or full-waveforms) and inversion schemes, but may also be due to the complex relationship between the heterogeneous path coverage of the Pacific region, irregular resolution, and data-errors.

Without full resolution and uncertainty information – model statistics – we cannot give strong quantitative arguments for or against the significance of any small-scale feature of a tomographic model (e.g. Foulger et al. 2013). Unfortunately, most damped-least-squares or iterative conjugate-gradient inversion schemes do not provide model statistics or control them during the inversion, as they focus instead on minimising the misfit between observed measurements and those predicted by the tomographic model. It is possible to approximate the model statistics for these inversion schemes (see for example Rawlinson et al. 2014; Rawlinson & Spakman 2016, for reviews of various techniques), but such methods are rarely used in practice as they are computationally costly and often provide only crude estimates of the model statistics. For the tomography problem considered here, it would indeed be possible to explicitly compute the $\boldsymbol{G}^{\dagger}$ matrix from a 'traditional' damped-least-squares inversion and then propagate the data uncertainties into model uncertainties; but the advantage of the SOLA method is that it allows a direct control on the model resolution and uncertainties.

4.2 Advantages of Backus-Gilbert type inversions

Beyond delivering model resolution and uncertainties at no extra cost, there are other important advantages of using inversion schemes based directly on the model statistics, such as those proposed by Backus & Gilbert (1967, 1968, 1970), Pijpers & Thompson (1992, 1993), Zaroli (2016), and Zaroli et al. (2017). Firstly, these inversion schemes constrain model uncertainties to vary smoothly

over the parameter space. Secondly these schemes drive the inversion to produce well-focused local averages (averaging kernels) that limit smearing along predominant paths (with SOLA this requires imposing well-focused target kernels, in our case circular ones). Thirdly, these schemes guarantee that the weights of the local averaging kernels (i.e. the rows of the resolution matrix) sum exactly to one, thereby eliminating the averaging bias that occurs with heterogeneous spatial distributions of data.

4.3 Model statistics – required for tomographic filtering and evaluating significance

Before comparing two different tomographic models, or a tomographic model to other data with different spatial resolution, we need to homogenize their resolutions, a process called *tomographic filtering* (Ritsema et al. 2007; Simmons et al. 2019). This requires the knowledge of the full resolution matrix. In our study, we considered a simple temperature model of lithosphere cooling based on the half space cooling model of Parker & Oldenburg (1973) that we filtered to the resolution of the tomographic model. Because the temperature model was very smooth, the tomographic filtering had only a minor effect. However, such filtering would be critical when comparing tomographic images with sharper models (Mégnin et al. 1997; Simmons et al. 2019; Freissler et al. 2020).

To assess if some parts of a tomographic model significantly deviate from some reference we need to know the model uncertainties at each location. Because we had this information, we were able to mask deviations from the half-space cooling predictions that were smaller than $\pm 1\sigma_m$ and $\pm 2\sigma_m$ and argue for the significance of certain anomalies. The process we illustrated in section 3 could replace statistical appraisal techniques and other tomographic resolution tests that are not always well understood (Lévêque et al. 1993; Rawlinson et al. 2014; Rawlinson & Spakman 2016).

4.4 Two significant anomalies in the Pacific region

We have focused on the significance of two anomalies: a low-velocity anomaly located to the north-east of Hawaii at 200 km depth and coherent bands of fast and slow velocities at 275 km depth.

4.4.1 Anomaly NE of Hawaii

This emerges strongly from the background with an amplitude about 0.25 km/s slower than the average shear-wave velocity of the Pacific plate, where the uncertainty is around 0.07 km/s, thus corresponding to about -3σ_m (Figs 7e and f). Despite using the same data-set, the model produced by Maggi et al. (2006) shows a much weaker anomaly at the same location, probably because it was smoothed out by their damped-least-squares inversion and regularization. This anomaly has also been seen, though at weaker amplitudes than in this study, in tomographies constructed using fundamental mode Rayleigh waves with a finite-frequency scheme (Liu & Zhou 2016) or multi-mode Love and Rayleigh waves (Isse et al. 2019), while it was entirely missed by the GDM52 model of Ekström (2011) based on fundamental mode Love and Rayleigh wave dispersion data. Given the different data-sets, forward theories, inversion and damping schemes of these studies, it is hard to pinpoint the reason for their weaker anomaly, but we speculate that the anomaly was probably over-smoothed by the regularization of their inversions.

4.4.2 Coherent SE-NW bands

We see these alternating fast and slow bands emerging from the background at 275 km depth with deviations from HSC up to ± 0.25 km/s ($\pm 1\sigma_m$ and $\pm 2\sigma_m$, see Fig. 8). Maggi et al. (2006) do not show these bands in their 275 km depth image, suggesting that the SOLA inversion scheme was able to extract more information from the same dataset. Other studies have observed similar patterns: Ekström (2011) show a very weak pattern in their isotropic results, probably because fundamental-mode surface waves are only weakly sensitive at these depths; French et al. (2013) see a stronger pattern with their full-waveform inversion; and, intriguingly, Liu & Zhou (2016) show similar bands in their dispersion maps at 30s and 50s, which are sensitive to much shallower depths.

4.5 Limits and perspectives

The SOLA inversion seems promising because it gives a way to control the model statistics (uncertainties and resolution) and to obtain them explicitly. The main problem is that it assumes

the forward problem is linear while surface-wave physics is not (the dispersion characteristics of surface-waves are non-linearly related to the shear-wave velocity as a function of depth). Surface-wave tomography is often performed in two steps: a non-linear step to relate frequency-dependent phase and/or group velocities to depth-dependent shear-wave velocity; and a linear regionalization step. Depending on the order of the two steps, the regionalization is performed either on group and phase velocities, or on 1D shear-wave velocities as in Maggi et al. (2006). Only the linear step can be performed using SOLA.

In this study, as in Maggi et al. (2006), each depth in the tomographic model was inverted independently, so vertical coherence could not be imposed or evaluated using resolution kernels. Furthermore, as the inputs to the SOLA inversion were 1D path-averaged shear-wave velocity models defined on the same depth range, the geographical coverage was identical for each depth, and did not take into account the greater sensitivity to depth of some paths (illuminated by longer-period or higher-mode surface waves) with respect to others.

To obtain a fully 3D shear-wave velocity model directly from frequency-dependent measurements of surface-waves in a single linear step with SOLA, and to take the full depth-sensitivity of individual measurements into account, we would need to adopt a finite-frequency description of surface waves to help linearize the tomographic problem (e.g. Zhou et al. 2004; Yoshizawa & Kennett 2005). This would also allow us to obtain a fully 3D model with 3D resolution kernels, and therefore be able to apply the same statistical rigor to interpreting any vertical structures within the model (e.g. plumes, the lithosphere depth).

Like any other process, the pertinence of the SOLA outputs rest upon the quality of the inputs. In particular, the model errors we have relied upon to decide the significance of anomalies are nothing more than a propagation into the model space of the data errors. While traditional damped-least-squares schemes use data uncertainties to weigh data relatively to each other and to evaluate model quality (using χ^2_{red} measures, for example), SOLA uses them more directly to evaluate uncertainties in the model which then drive the inversion through their trade-off with resolution. In order to be able to trust the magnitude of the model errors in SOLA, it is necessary for the magnitude of the data errors to be correct. This is the reason why we upscaled the data uncertainties

using the χ^2_{red} values of the damped-least-squares model of Maggi et al. (2006); however a more rigorous assessment of the data uncertainties is still required.

Our study calls for other surface wave tomography studies based on the SOLA inversion in other regions of the world, with new data sets, and particular care in estimating data uncertainties. Our tomographic model, together with its resolution and uncertainty maps, calls for new data in the Pacific to observe other seismic velocity anomalies significant enough to be interpreted robustly.

5 CONCLUSION

Rayleigh wave tomography of the Pacific upper mantle shows large scale geological features (e.g. lithosphere cooling) but both the model uncertainties and resolution are required for interpretations at smaller scales. We exploited the SOLA tomographic inversion scheme (Zaroli 2016) to propose a workflow to analyse the tomographic model using its resolution and uncertainties, that is:

(i) Define some reference, Earth-like, physical model;

(ii) Filter the reference model to the tomographic resolution;

(iii) Compute the deviations from the reference model;

(iv) Normalize the deviations by the model uncertainties;

(v) Mask model estimates that are within one or two error bars to focus only on significant seismic anomalies;

(vi) Compare those non-masked anomalies to the local resolution to discuss their resolvability and spot artifacts, if any.

In this study, we used a half space cooling description of the Pacific lithosphere as a reference model. We observed a low velocity anomaly to the north-east of Hawaii at 200 km depth and a pattern of alternatively slow and fast anomalous bands oriented approximately south-east to north-west at 275 km depth. Both features are reliable given the model statistics. Our study shows that the model statistics can be properly analysed in surface wave tomography using the SOLA inversion; the same scheme could be applied in other regions of the globe.

In this study, the data uncertainties were estimated using multiple sampling with ray paths

close to each other, then upscaled because apparently underestimated. We should aim collectively to characterize more accurately uncertainties in seismological data sets - clearly a challenging but crucial task.

DATA AND CODE AVAILABILITY

The computer codes for the SOLA inversions and the data used in this study are available from C.Z. (c.zaroli@unistra.fr) and A.M. (alessia.maggi@unistra.fr), respectively, upon reasonable request.

Acknowledgements

The authors are grateful to Huajian Yao (editor), Frederik Tilmann, and an anonymous reviewer whose constructive comments helped to improve this paper.

REFERENCES

Auer, L., Boschi, L., Becker, T. W., Nissen-Meyer, T., & Giardini, D., 2014. *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.

Backus, G. & Gilbert, F., 1968. The Resolving Power of Gross Earth Data, *Geophys J Int*, **16**(2), 169–205.

Backus, G. E. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data, *Phil. Trans. R. Soc. A.*, **266**(1173), 74.

Backus, G. E. & Gilbert, J. F., 1967. Numerical Applications of a Formalism for Geophysical Inverse Problems, *Geophys J Int*, **13**(1-3), 247–276.

Barruol, G., 2002. PLUME investigates South Pacific Superswell, *Eos Trans. AGU*, **83**(45), 511–514.

Cara, M. & Lévêque, J. J., 1987. Waveform inversion using secondary observables, *Geophys. Res. Lett.*, **14**(10), 1046–1049.

Chen, G., Spetzler, H. A., Getting, I. C., & Yoneda, A., 1996. Selected elastic moduli and their temperature derivatives for olivine and garnet with different Mg/(Mg+Fe) contents: Results from GHz ultrasonic interferometry, *Geophys. Res. Lett.*, **23**(1), 5–8.

Courtillot, V., Davaille, A., Besse, J., & Stock, J., 2003. Three distinct types of hotspots in the Earths mantle, *Earth and Planet. Sc. Lett.*, **205**(3-4), 295–308.

Debayle, E., 1999. SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion, *Geophys. J. Int.*, **137**(3), 747–754.

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

Dziewonski, A. M. & Anderson, D., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**(4), 297–356.

Ekström, G., 2011. 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.

Faul, U. & Jackson, I., 2005. The seismological signature of temperature and grain size variations in the upper mantle, *Earth and Planet. Sc. Lett.*, **234**(1-2), 119–134.

Foulger, G. R., 2011. *Plates vs plumes: a geological controversy*, John Wiley and Sons.

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., & Yanovskaya, T. B., 2013. Caveats on tomographic images, *Terra Nova*, **25**(4), 259–281.

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

French, S., Lekic, V., & Romanowicz, B., 2013. Waveform Tomography Reveals Channeled Flow at the Base of the Oceanic Asthenosphere, *Science*, **342**(6155), 227–230.

Gripp, A. E. & Gordon, R. G., 1990. Current plate velocities relative to the hotspots incorporating the NUVEL-1 global plate motion model, *Geophys. Res. Lett.*, **17**(8), 1109–1112.

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

Lévêque, J. J., Rivera, L., & Wittlinger, G., 1993. On the use of the checker-board test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, **115**(1), 313–318.

Liu, K. & Zhou, Y., 2016. Global Rayleigh wave phase-velocity maps from finite-frequency tomography, *Geophys. J. Int.*, **205**(1), 51–66.

Maggi, A., Debayle, E., Priestley, K., & Barruol, G., 2006. Multimode surface waveform tomography of the Pacific Ocean: a closer look at the lithospheric cooling signature, *Geophys. J. Int.*, **166**(3), 1384–1397.

Mégnin, C., Bunge, H.-P., Romanowicz, B., & Richards, M. A., 1997. Imaging 3-D spherical convection models: What can seismic tomography tell us about mantle dynamics?, *Geophys. Res. Lett.*, **24**(11), 1299–1302.

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

Montagner, J. P., 2002. Upper mantle low anisotropy channels below the Pacific Plate, *Earth and Planet. Sc. Lett.*, **202**(2), 263–274.

Müller, R. D., Sdrolias, M., Gaina, C., & Roest, W. R., 2008. Age, spreading rates, and spreading asymmetry of the world's ocean crust, *Geochem. Geophys. Geosyst.*, **9**(4).

Nettles, M. & Dziewoski, A. M., 2008. Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America, *J. Geophys. Res.*, **113**(B2), B02303.

Nolet, G., 1985. Solving or resolving inadequate and noisy tomographic systems, *J. Comput. Phys.*, **61**(3), 463–482.

Nolet, G., 2008. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*, Cambridge University Press, Cambridge.

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

Panning, M. P., Leki, V., & Romanowicz, B. A., 2010. Importance of crustal corrections in the development of a new global model of radial anisotropy, *J. Geophys. Res.*, **115**(B12), B12325.

Parker, R. & Oldenburg, D., 1973. Thermal Model of Ocean Ridges, *Nature Physical Science*, **242**(122), 137–139.

Pijpers, F. & Thompson, M., 1992. Faster formulations of the optimally localized averages method for helioseismic inversions, *Astron. Astrophys.*, **262**, 33–36.

Pijpers, F. & Thompson, M., 1993. The SOLA method for helioseismic inversion, *Astron. Astrophys.*, **281**, 231–240.

Priestley, K. & Mckenzie, D., 2006. The thermal structure of the lithosphere from shear wave velocities, *Earth and Planet. Sc. Lett.*, **244**(1-2), 285–301.

Rawlinson, N. & Spakman, W., 2016. On the use of sensitivity tests in seismic tomography, *Geophys. J. Int.*, **205**(2), 1221–1243.

Rawlinson, N., Fichtner, A., Sambridge, M., & Young, M., 2014. Seismic Tomography and the Assessment of Uncertainty, *Adv. Geophys.*, **55**, 1–76.

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

Ritzwoller, M. H., Shapiro, N. M., & Zhong, S.-J., 2004. Cooling history of the Pacific lithosphere, *Earth and Planet. Sc. Lett.*, **226**(1-2), 69–84.

Sambridge, M. & Rawlinson, N., 2005. Seismic tomography with irregular meshes, in *Geophysical Monograph Series*, vol. 157, pp. 49–65, eds Levander, A. & Nolet, G., American Geophysical Union, Washington, D. C.

Scales, J. A. & Snieder, R., 1997. To Bayes or not to Bayes?, *Geophysics*, **62**(4), 1045–1046.

Simmons, N. A., Schuberth, B. S. A., Myers, S. C., & Knapp, D. R., 2019. 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.

Yoshizawa, K. & Kennett, B. L. N., 2005. Sensitivity kernels for finite-frequency surface waves, *Geophys. J. Int.*, **162**(3), 910–926.

Zaroli, C., 2016. Global seismic tomography using Backus–Gilbert inversion, *Geophys. J. Int.*, **207**(2), 876–888.

Zaroli, C., 2019. Seismic tomography using parameter-free Backus–Gilbert inversion, *Geophys. J. Int.*, **218**(1), 619–630.

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

Zhou, Y., Dahlen, F. A., & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.*, **158**(1), 142–168.

Zhou, Y., Nolet, G., Dahlen, F. A., & Laske, G., 2006. Global upper-mantle structure from finite-frequency surface-wave tomography, *J. Geophys. Res.*, **111**(B04304).