Chapter 3

SOLA tomography in a nutshell

Foreword

In this chapter, I intend to give an overview – with a lecture-like flavor – of the SOLA–Backus–Gilbert inversion method (or SOLA[†] for short) that I have recently introduced and adapted to solve large-scale, linear, both discrete and continuous, tomographic problems (Zaroli, 2016, 2019; Zaroli et al., 2017). My goal is for the reader, in particular for MSc and PhD students in Earth sciences, to quickly learn:

1. What *SOLA tomography* is about;
2. Why it could (should) be preferred to other 'classical' tomographic methods.

[†] **S**ubtractive **O**ptimally **L**ocalized **A**verages

3.1 Introduction

3.1.1 Linear tomographic problems

First, the SOLA method is limited[††] to linear(ized) tomographic problems, such as:

- (finite-frequency) body-wave mantle tomography (Zaroli, 2016, 2019),
- surface-wave tomography (Latallerie, 2019),
- ambient-noise based, surface-wave group-velocity tomography (Ouattara, 2019),
- joint normal-mode and body-wave whole-mantle tomography (Dubois, 2020).

The three last examples refer to recent SOLA-based tomographic studies conducted by MSc and PhD students – co-supervised by myself. In the following, I shall illustrate most of the concepts of SOLA inversion through the example of finite-frequency[†††] body-wave time-residual tomography, while considering both settings of continuous and discrete model space. As a remark, it would be straightforward to replace, in the SOLA theoretical developments presented in this chapter, the *finite-frequency* physical framework by the simpler one of *ray theory* (see Nolet (2008) or Zaroli (2010) for an introduction to ray *vs* finite-frequency theory; see also Fig. 3.1.1).

[††] Though the B–G theory has been developed for linear problems, Snieder (1991) extended the theory to *weakly* non-linear problems. However, this theory, as it is, seems too complex to be used in practice (Snieder & Trampert, 1999).

[†††] Finite frequency wave-propagation theory (Marquering et al., 1998; Dahlen et al., 2000).

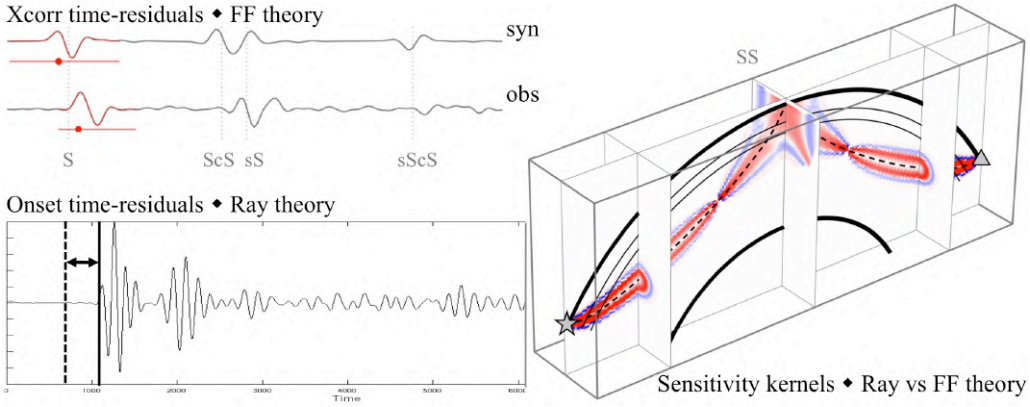


Figure 3.1.1: (Upper-left) Cross-correlating observed and synthetic S-phase waveforms, to get time-residual measurements for finite-frequency theory. (Lower-left) Picking onset time-residuals for ray theory. (Right) Example of 3–D, finite-frequency, SS-phase time-residual sensitivity kernel (units are s.km^{-3}; red/blue=negative/positive, respectively; white='no data sensitivity'); the black dashed line represents the geometrical ray path, that is, the 1–D ray-theoretical sensitivity kernel.

More specifically, let consider **linear, continuous, tomographic problems** of the form (as for finite-frequency body-wave tomography of the Earth's mantle):

$$d_i = \int_{\text{Mantle}} K_i(\mathbf{r})\, m(\mathbf{r})\, d^3\mathbf{r} \;+\; n_i\,, \quad 1 \le i \le N\,, \tag{3.1.1}$$

- $\mathbf{d} = (d_i)_{1\le i\le N}$ stands for the **data** vector, of size N[†]; each datum d_i represents say an S-wave time residual, δt_i, measured by cross-correlating a pair of observed and synthetic S-phase waveforms, filtered around a given central period (see Zaroli et al. (2010) and Fig. 3.1.1).

- $m(\mathbf{r})$ denotes the **continuous 'true' model**[††]; for instance, $m(\mathbf{r})$ represents the 3–D, isotropic, shear-wave velocity perturbation, $\frac{\delta V_S}{V_S}$, in location $\mathbf{r}$, with respect to a 1–D, radial V_S model such as IASP91 (Kennett & Engdahl, 1991).

- $K_i(\mathbf{r})$ is here a 3–D (volumetric) finite-frequency sensitivity kernel, associated to an S-wave time-residual datum d_i (see Fig. 3.1.1); such a finite-frequency kernel K_i depends on the filtering period used to measure cross-correlation time residuals (Marquering et al., 1998). More generally, K_i is referred to as the i-th **data-sensitivity kernel** (or 'sensitivity kernel' for short).

- $\mathbf{n} = (n_i)_{1\le i\le N}$ stands for the – assumed – zero-mean, normally-distributed, **noise** vector; the data covariance matrix is then diagonal: $\mathbf{C}_d = \text{diag}(\sigma_{d_i}^2)_{1\le i\le N}$. As we shall discuss further, a robust estimation of data errors (σ_{d_i}) is crucial for taking full advantage of the outcomes of the SOLA inversion approach.

[†] N stands for the number of data, although M is also used in the (American) literature.

[††] "*The true model is the model that corresponds to reality and which is only known to the gods. However, we are not aware of any definition that is operational in the sense that it provides us with a set of actions that could potentially tell us what the true model really is.*" (Snieder & Trampert, 1999)

3.1.2 Asking questions on the Earth's interior

A major goal of seismic tomography is to make inferences about the physical properties of the Earth's interior, that is, from performing a so-called 'inversion' of a set of data (seismological observables).

- **Forward problem**. Let consider that knowledge of wave-propagation physics (as the aforementioned finite-frequency theory) allows one to 'predict' data, $d_i^{\rm pred}$, given an input Earth model, $m_{\rm in}$, that is:

$$d_i^{\rm pred} \longleftarrow \int_{\rm Mantle} K_i(\mathbf{r})\; m_{\rm in}(\mathbf{r})\; d^3\mathbf{r}\,. \tag{3.1.2}$$

  This is called a – linear – forward problem.

- **Inverse problem**. In an inverse problem, one aims at reconstructing the (unknown) true model m from a set of actual data $\mathbf{d}$. The situation is sketched in Fig. 3.1.2(left). It is important to realize that in most, if not all, tomographic experiments, the model m is a *continuous* function of the space variables, while we are to deal with a *finite* amount of (noisy) data to determine m. Thus, as nicely formulated by Snieder & Trampert (1999), "*a simple count of variables shows that the data cannot carry sufficient information to determine the model uniquely*[†]". In other words, there are infinitely many model solutions, $\hat{m}$, that could explain the data equally well (even with error-free data). Therefore, any particular solution $\hat{m}$ cannot be *stricto senso* equal to the true model m.

- **Model estimation 'and' appraisal**. The estimation problem consists in reconstructing from the data, $\mathbf{d}$, a (non-unique) model estimate, $\hat{m}$, that is consistent with the data, and their error bars. In the appraisal problem one aims to determine *which* properties of the true model m can be recovered in the model estimate $\hat{m}$ – and the attached errors. See Fig. 3.1.2(right).

[†] A starting point of the theory of Backus & Gilbert (1967, 1968) was to recognize this *non unicity* aspect of inverse problems.

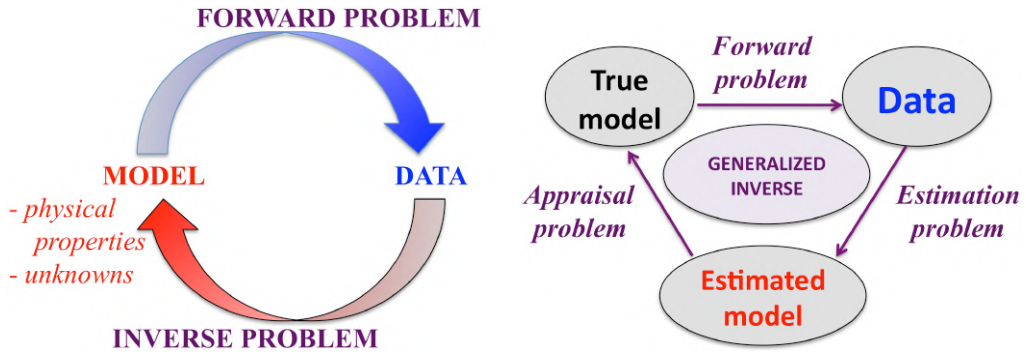


Figure 3.1.2: (Left) Forward and inverse problems. (Right) The inverse problem now viewed as a combination of an estimation problem and an appraisal problem. In any linear inverse theory, the explicit knowledge of the generalized inverse is crucial for model estimation *and* model appraisal.

- **The starting point of Backus–Gilbert**. A fundamental insight from the pioneering works by Backus & Gilbert (1967, 1968, 1970) is that tomographic problems are invariably, at least partly, underdetermined: "*the collection of Earth models which yield the physically observed values of any independent set of gross Earth data is either empty or infinite dimensional*" (Backus & Gilbert, 1967). Recognizing this fact, the Backus–Gilbert (B–G) approach[†] seeks not to construct a particular model solution, $\hat{m}$, that is, to estimate infinitely many model parameters, but instead to determine some *optimally localized averages* over the true model m. Indeed, it is possible to identify unique averages, even when the (infinitely many) parameters themselves are not uniquely defined.

[†] The linear B–G inversion scheme belongs to the class of Optimally Localized Averages (OLA) methods.

- **Asking relevant questions**. The fact that most tomographic problems are underdetermined (because of imperfect data coverage), and deal with noisy data, makes it futile to ask questions on the Earth's interior of the form: *What is the shear-wave velocity right below Paris, at 321 km depth? That is, what is the true-model value at a specific location,* $m(\mathbf{r}^{Paris})$*?* It is clear that an infinite error would be attached to any numerical answer to that question, so that nothing could be learned on the true-model properties. See Fig. 3.1.3(left).

  The Backus–Gilbert approach rather consists in asking questions of the form: *What is the average value, and attached uncertainty*[††]*, of shear-wave velocities within a given averaging volume, localized around* $\mathbf{r}^{Paris}$*, i.e.,* $< m(\mathbf{r}^{Paris}) >$[†††]*?* As illustrated in Fig. 3.1.3(right), meaningful local-average estimates spread out over a volume (in model space) that is large enough to influence the data.

  We shall see that the SOLA–Backus–Gilbert approach (Zaroli, 2016, 2019) is a tomographic method of choice for getting *quantitative answers* to the aforementioned, relevant questions to be asked on the properties of Earth.

[††] Here, uncertainties merely represent the amount of data noise that is statistically expected to propagate into the local-average estimates.

[†††] Here, the notation $< \cdot >$ denotes a general averaging process.

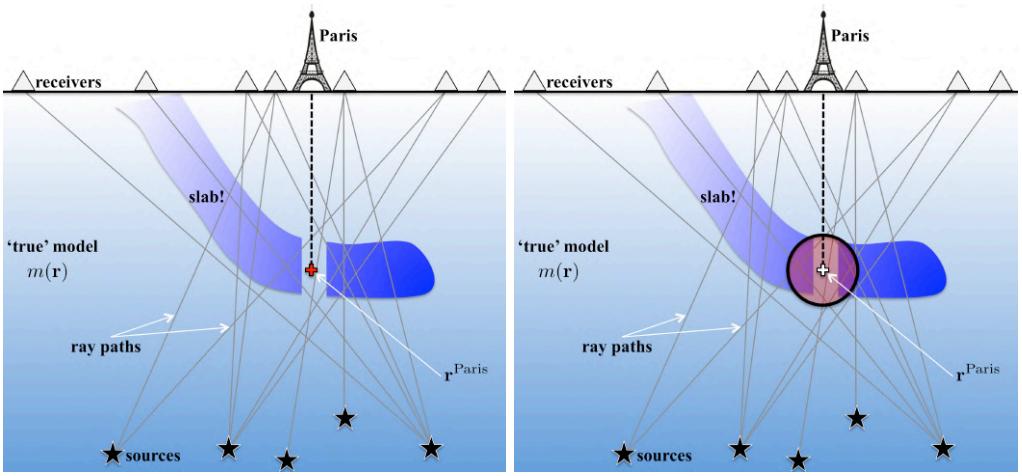


Figure 3.1.3: (Left) Futile attempt at estimating the true-model value at a 'specific' location. (Right) Meaningful attempt at estimating a 'local average' over the (continuous) true-model values, within an averaging volume (red disk) that is well localized and large enough to influence the data.

- **Two different philosophies**. Before diving into the details of the (SOLA) Backus–Gilbert approach, let us step back, and briefly review the two major philosophies which have been used (in one form or another) for solving general inverse problems in Earth sciences. That is, quoting Tarantola (2006):

  1. – *"The first*[†] *carefully avoids using any a priori information on the model parameters that could 'bias' the inferences to be drawn from the data."*
     – *"The theory developed by Backus and Gilbert is a good example of this approach."*
  2. – *"The second philosophy is clearly bayesian."*
     – *"When starting with some a priori state of information on the model parameter, which is the a posteriori state of information at which we arrive after 'assimilating' new data?"*

  [†] *"[It] is favoured by statisticians, as it can easily be formalized and presented with standard mathematical rigour."*

  A fundamental question in inverse theory is: *"Should a priori information*[††] *be used?"* (Tarantola, 2006). Clearly it is 'no' for the first philosophy (Backus–Gilbert-like), and 'yes' for the second (bayesian-like). I think that one cannot avoid a subjective reasoning, when arguing that one of these two points of view is (always) better adapted to solve inverse problems arising in (geo)sciences. On one hand, if relevant[†††] *a priori* information is available (e.g., known extreme values of model parameters or correlations between parameters), the bayesian approach can[††††] take advantage of it, while Backus–Gilbert cannot. On the other hand, if no relevant *a priori* information is available – as it happens for most seismic tomography experiments, for which *ad hoc* prior model constraints are frequently used – then the Backus–Gilbert point of view may be preferred, as it can make inferences[†††††] from the data without potential 'biases' induced by *ad hoc*, *a priori* model information (see Sect. 3.3).

  [††] I shall focus on *a priori* model information (and simply assume zero-mean data errors and a diagonal prior data covariance matrix).

  [†††] Based on physical, 'objective' considerations.

  [††††] Although the Bayesian approach may suffer from the *curse* of model dimensionality. (B–G treats model space as a continuum.)

  [†††††] Local averages with respect to the 'true' model.

3.2 SOLA–Backus–Gilbert tomography

I aim to present the concepts involved in seismic tomography using 1) continuous Backus–Gilbert inversion, 2) continuous SOLA inversion (Zaroli, 2019), and 3) discrete SOLA inversion (Zaroli, 2016). Theoretical developments will be illustrated, in the case of SOLA tomography, with examples. In particular, we shall see that the SOLA–Backus–Gilbert tomographic approach explicitly allows one to get 'quantitative answers' to questions on Earth's properties of the form: *What is the average value, and attached uncertainty, of seismic velocities within a given, localized volume?* Ultimately, that knowledge – of local averages – may help to better understand the complex structures and dynamics taking place in the Earth's interior.

3.2.1 Continuous Backus–Gilbert inversion

- **Local weighted-averages**. In the continuous Backus–Gilbert approach (Backus & Gilbert, 1967, 1968, 1970), one explicitly seeks a model estimate $\hat{m}^{(k)}$ which

represents a weighted average over the continuous 'true' model, $m(\mathbf{r})$. This averaging process takes place through an averaging kernel[†], $A^{(k)}(\mathbf{r})$, that we wish to be optimally localized around a given query point, $\mathbf{r}^{(k)}$; see Fig. 3.2.1(left). This leads to writing:

[†] Here, dimensionality of averaging kernels is km^{-3}.

$$\hat{m}^{(k)} = \underbrace{\int_{\text{Mantle}} A^{(k)}(\mathbf{r})\, m(\mathbf{r})\, d^3\mathbf{r}}_{\substack{\text{AVERAGING} \\ \text{PROCESS}}} \quad (+ \text{ propagated noise}) \,, \tag{3.2.1}$$

where the term 'propagated noise' means the part of noise, $\mathbf{n}$, in the data, $\mathbf{d}$, that *propagates* (through the inversion) into the model estimate, $\hat{m}^{(k)}$.

- **Unbiased averages**. We wish that the integral $\int A^{(k)} m$ yields *unbiased* (weighted) averages over the true model m. The averaging kernel $A^{(k)}$ should then be non-negative, and satisfy to the following condition (see Text Box 3.1):

$$\int A^{(k)}(\mathbf{r})\, d^3\mathbf{r} = 1 \,. \tag{3.2.2}$$

Text Box 3.1: WEIGHTED AVERAGE OF A CONTINUOUS FUNCTION

Let $\hat{f}$ a **weighted average** of function $f(x)$, with (non-negative) weighting function $w(x)$, with $x \in [a,b]$. It writes: $\hat{f} = \frac{\int_a^b w(x)f(x)dx}{\int_a^b w(x)dx} = \int_a^b A(x)f(x)dx$, with (non-negative) **averaging function** $A(x) = \frac{w(x)}{\int_a^b w(x)dx}$, that satisfies $\int_a^b A(x)dx = 1$. ⊡

- **Generalized-inverse operator**. As the tomographic problem ($d_i = \int K_i m + n_i$) is linear, one can seek the estimate $\hat{m}^{(k)}$ as a linear combination of the data:

$$\hat{m}^{(k)} = \sum_{i=1}^{N} x_i^{(k)} d_i \,, \tag{3.2.3}$$

[††] When inverting time residuals (s), the physical unit of generalized-inverse coefficients is s^{-1}.

where the N unknown, real-valued coefficients[††] , $\mathbf{x}^{(k)} = (x_i^{(k)})_{1\leq i\leq N}$, represent *some* generalized-inverse operator, mapping the data to the model estimate $\hat{m}^{(k)}$. This operator, and the way it is calculated, is the cornerstone of every inversion scheme; see Figs. 3.1.2(right) and 3.2.1(right). From combining the previous equations, one finds that the weighted-average estimate $\hat{m}^{(k)}$ writes:

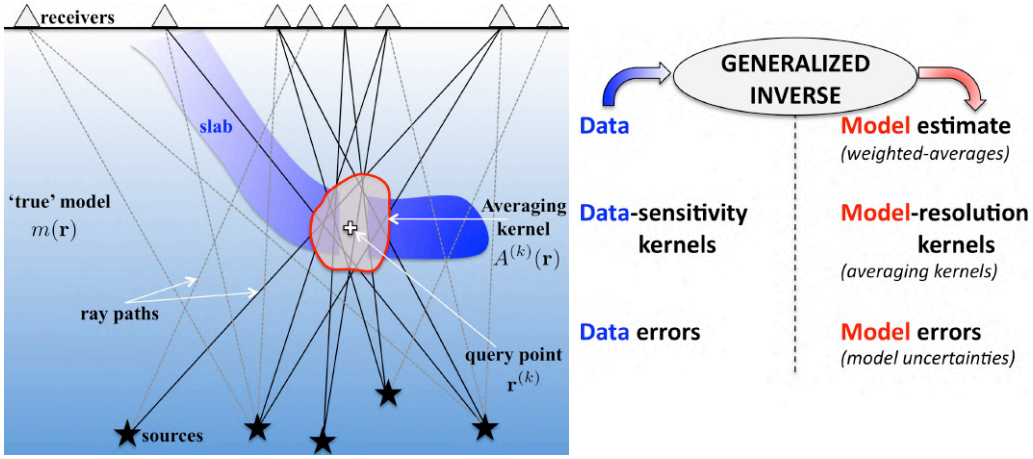


Figure 3.2.1: (Left) Illustration of an averaging process that takes place through an averaging kernel, $A^{(k)}(\mathbf{r})$, that we wish to be spatially localized around a given query point, $\mathbf{r}^{(k)}$, but also to be large enough to influence the data. Remark: In practice, the averaging kernel would likely not be constant-valued inside the reddish area (this is why one often is dealing with 'weighted' averages). (Right) Generalized-inverse operator: a 'shuttle' from data to model space.

$$\hat{m}^{(k)} = \int \underbrace{\left(\sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r})\right)}_{\substack{\text{AVERAGING KERNEL} \\ A^{(k)}(\mathbf{r})}} m(\mathbf{r}) d^3\mathbf{r} + \underbrace{\sum_{i=1}^{N} x_i^{(k)} n_i}_{\substack{\text{PROPAGATED} \\ \text{NOISE}}} \quad . \tag{3.2.4}$$

Note that, in the Backus–Gilbert approach, an averaging kernel, $A^{(k)}$, is explicitly defined as a linear combination of data-sensitivity kernels, K_i, that is:

$$A^{(k)}(\mathbf{r}) = \sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r}) \,, \tag{3.2.5}$$

see Fig. 3.2.2. Statistically, the weighted-average estimate $\hat{m}^{(k)}$ therefore writes:

$$\underbrace{\hat{m}^{(k)}}_{\substack{\text{weighted-average} \\ \text{ESTIMATE}}} \equiv \underbrace{\int A^{(k)}(\mathbf{r}) m(\mathbf{r}) \, d^3\mathbf{r}}_{\substack{\text{AVERAGING} \\ \text{PROCESS}}} + \underbrace{\mathcal{N}(0, \sigma^2_{\hat{m}^{(k)}})}_{\substack{\text{statistical propagation} \\ \text{of noise (normal distribution)}}} \quad , \tag{3.2.6}$$

where the standard deviation $\sigma_{\hat{m}^{(k)}}$ is referred to, in this work terminology, as the *uncertainty* attached to the estimate $\hat{m}^{(k)}$, and can be expressed as:

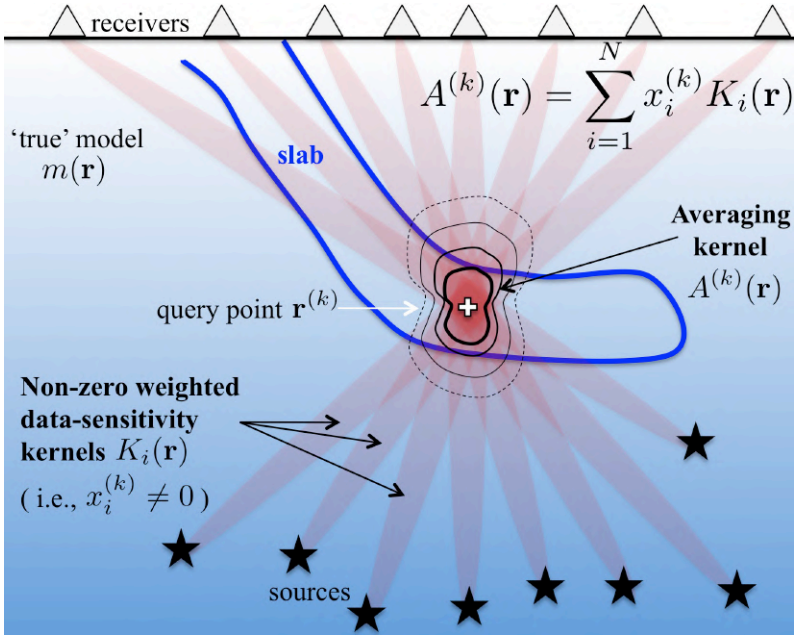


Figure 3.2.2: In the B–G approach, an averaging kernel is explicitly defined as a linear combination of data-sensitivity kernels, whose weights are given by the generalized-inverse coefficients. Data-sensitivity kernels too 'far away' from the query point (not shown here) should be given zero weights (i.e., $x_i^{(k)} = 0$), for the averaging kernel to be optimally localized close to the query point.

$$\sigma_{\hat{m}^{(k)}} = \sqrt{\sum_{i=1}^{N} \left(x_i^{(k)}\right)^2 (\sigma_{d_i})^2}\,. \tag{3.2.7}$$

As a remark, the 'uncertainty' $\sigma_{\hat{m}^{(k)}}$ merely represents the amount of noise that may propagate into the weighted-average estimate $\hat{m}^{(k)}$. That is, $\sigma_{\hat{m}^{(k)}}$ cannot inform us on how much the estimate $\hat{m}^{(k)}$ may differ from the true model value $m(\mathbf{r}^{(k)})$, at the specific query-point location $\mathbf{r}^{(k)}$. Finally, from the generalized-inverse coefficients, $\mathbf{x}^{(k)}$, one can directly infer the weighted-average estimate $\hat{m}^{(k)}$, averaging kernel $A^{(k)}$, and uncertainty $\sigma_{\hat{m}^{(k)}}$, that is:

$$\mathbf{x}^{(k)} \Longrightarrow \begin{cases} \sum_{i=1}^{N} x_i^{(k)} d_i & \longrightarrow \quad \hat{m}^{(k)} \\ \sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r}) & \longrightarrow \quad A^{(k)}(\mathbf{r}) \\ (\sum_{i=1}^{N} (x_i^{(k)} \sigma_{d_i})^2)^{1/2} & \longrightarrow \quad \sigma_{\hat{m}^{(k)}}\,. \end{cases} \tag{3.2.8}$$

- **Resolution / uncertainty**. Firstly, the spatial extent of an averaging kernel (sometimes also referred to as a resolving kernel), $A^{(k)}(\mathbf{r})$, informs us on the

size of 'true' model features, around a query point $\mathbf{r}^{(k)}$, that could be recovered (resolved) through the associated averaging process. Secondly, the uncertainty value, $\sigma_{\hat{m}^{(k)}}$, informs us on the amount of noise that could contaminate a weighted-average estimate, $\hat{m}^{(k)}$. Both averaging kernels *and* uncertainties must be taken into account to appraise (weighted-average) model estimates. Finally, as shown in Fig. 3.2.3, there is a well-known (Menke, 1989) trade-off[†] between resolution (i.e., size of $A^{(k)}$) and uncertainty (i.e., value of $\sigma_{\hat{m}^{(k)}}$).

[†] In general, the larger the volume of the resolving kernel $A^{(k)}$ (grey shape), the more the number of data influenced by this volume (red rays), and then the more the 'propagated noise' term, i.e., $(\mathbf{x}^{(k)})^T\mathbf{n}$, tends to cancel out, and so for the uncertainty $\sigma_{\hat{m}^{(k)}}$.

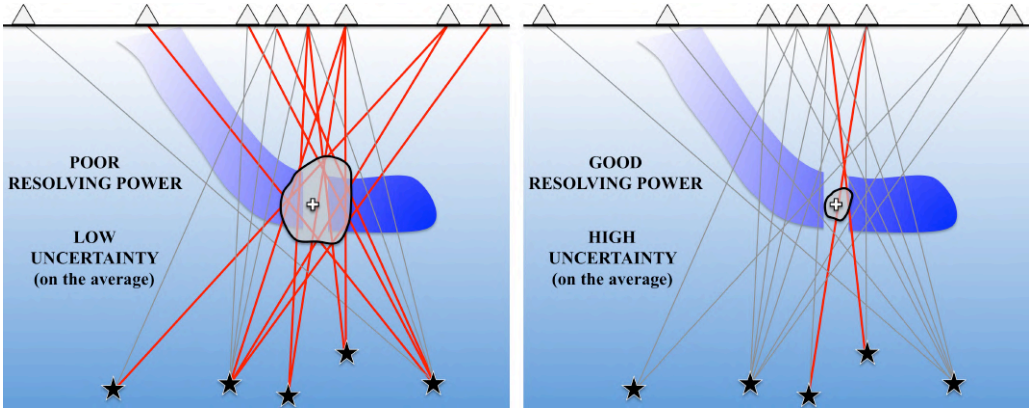


Figure 3.2.3: Illustration of trade-off between poor/good resolution (large/small spatial extent of averaging kernel $A^{(k)}$; see the grey shape), and low/high value of uncertainty $\sigma_{\hat{m}^{(k)}}$, respectively.

- **Continuous Backus–Gilbert minimization problem**. In the B–G approach, a minimization problem is set up to directly find out each generalized-inverse vector, $\mathbf{x}^{(k)}$, for each query point, $\mathbf{r}^{(k)}$, which writes (e.g., Nolet, 1985):

$$\begin{cases} \underset{\mathbf{x}^{(k)}\in\mathbb{R}^N}{\arg\min}: & \underbrace{\frac{56\pi}{9}\times\int\left[A^{(k)}\left(\mathbf{r}\right)\right]^2|\mathbf{r}-\mathbf{r}^{(k)}|^4 d^3\mathbf{r}}_{\substack{\text{Measure of the SPATIAL EXTENT}^{\dagger\dagger}\text{ of } A^{(k)}\\ \text{(Favoring peak-shaped averaging kernel)}}} + \underbrace{[\eta^{(k)}]^2}_{\substack{\text{Trade}\\ \text{off}}}\underbrace{\sigma^2_{\hat{m}^{(k)}}}_{\substack{\text{VARIANCE of } \hat{m}^{(k)}\\ \text{(Moderating uncertainty)}}} \\ \text{subject to} & \underbrace{\int A^{(k)}(\mathbf{r})d^3\mathbf{r}=1\,.}_{\substack{\text{AVERAGING}\\ \text{CONDITION}}} \end{cases} \tag{3.2.9}$$

[††] $\frac{56\pi}{9}$ is a scaling factor (see Text Box 3.2), while the power 4 ensures the correct dimensionality.

This leads to a compromise – via the trade-off parameter $\eta^{(k)}$ – between favoring a *peak-shaped* averaging kernel $A^{(k)}(\mathbf{r})$, peaked at $\mathbf{r}^{(k)}$, while also moderating the uncertainty $\sigma_{\hat{m}^{(k)}}$, on the weighted-average estimate (i.e., trade-off between resolution and uncertainty, see Fig. 3.2.3). In addition, the minimization problem is constrained for the averaging kernel $A^{(k)}$ to honor the (unbiased) 'averaging condition' (3.2.2). Finally, (3.2.9) writes (see Text Box 3.3):

$$\begin{cases} \text{solve} \quad \mathbf{F}^{(k)}\,\mathbf{x}^{(k)} = \mathbf{0}, \\ \text{subject to} \quad \sum_{i=1}^{N} x_i^{(k)} \int K_i(\mathbf{r}) d^3\mathbf{r} = 1\,, \end{cases}$$

$$\text{with:} \quad F_{ii'}^{(k)} = \tfrac{56\pi}{9} \int K_i(\mathbf{r}) K_{i'}(\mathbf{r}) |\mathbf{r} - \mathbf{r}^{(k)}|^4 d^3\mathbf{r} + [\eta^{(k)}\sigma_{d_i}]^2 \delta_{ii'}^{\text{Kro}\dagger}\,, \quad 1 \le i, i' \le N\,. \tag{3.2.10}$$

† $\delta_{ii'}^{\text{Kro}}$ is the usual Kronecker symbol.

(3.2.10) can be solved with the Lagrange-multipliers method (Pijpers & Thompson, 1992) or an LSQR-based approach (Paige & Saunders, 1982; Nolet, 1985).

Text Box 3.2: On the scaling factor $\frac{56\pi}{9}$

We wish that $W^{(k)} = \mathcal{C} \times \int \left[A^{(k)}(\mathbf{r})\right]^2 |\mathbf{r} - \mathbf{r}^{(k)}|^4 d^3\mathbf{r}$ represents the *spatial extent* of the averaging kernel $A^{(k)}$, where $\mathcal{C}$ is a scaling factor. If $A^{(k)}$ is equal to a constant inside a sphere, centered in $\mathbf{r}^{(k)}$, with diameter D, and 0 outside, then this constant $= \frac{1}{\text{sphere's volume}} = \frac{1}{\frac{4}{3}\pi(\frac{D}{2})^3}$; which ensures that $\int A^{(k)} = 1$. Without loss of generality, assume $\mathbf{r}^{(k)} = \mathbf{0}$. Therefore, using spherical coordinates (r, θ, ϕ), one finds:

$$\int \left[A^{(k)}(\mathbf{r})\right]^2 |\mathbf{r} - \mathbf{r}^{(k)}|^4 d^3\mathbf{r} = \frac{1}{(\frac{4}{3}\pi(\frac{D}{2})^3)^2} \int_{r=0}^{\frac{D}{2}} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} r^6 \sin\theta\, dr d\theta d\phi = \frac{9}{56\pi} \times D\,.$$

Finally, for this example, one wants to have $W^{(k)} = D$; this means that $\mathcal{C} = \frac{56\pi}{9}$. ⊡

Text Box 3.3: Continuous Backus–Gilbert normal equations

Expanding out $A^{(k)}$ and $\sigma_{\hat{m}^{(k)}}$ in function of $\mathbf{x}^{(k)}$, the system (3.2.9) writes:

$$\begin{cases} \underset{\mathbf{x}^{(k)} \in \mathbb{R}^N}{\arg\min:} \quad \frac{56\pi}{9} \int \left[\sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r})\right]^2 |\mathbf{r} - \mathbf{r}^{(k)}|^4 d^3\mathbf{r} + [\eta^{(k)}]^2 \sum_{i=1}^{N} (x_i^{(k)})^2 (\sigma_{d_i})^2 \\ \text{subject to} \quad \sum_{i=1}^{N} x_i^{(k)} \int K_i(\mathbf{r}) d^3\mathbf{r} = 1\,. \end{cases}$$

To obtain the normal equations corresponding to the 'unconstrained part' of the previous minimization problem, we do the following (i' is fixed):

$$\frac{\partial}{\partial x_{i'}^{(k)}} \left(\frac{56\pi}{9} \int \left[\sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r})\right]^2 |\mathbf{r} - \mathbf{r}^{(k)}|^4 d^3\mathbf{r} + [\eta^{(k)}]^2 \sum_{i=1}^{N} (x_i^{(k)})^2 (\sigma_{d_i})^2 \right) = 0\,,$$

$$\implies \quad \sum_{i=1}^{N} x_i^{(k)} \tfrac{56\pi}{9} \int K_i(\mathbf{r}) K_{i'}(\mathbf{r}) |\mathbf{r} - \mathbf{r}^{(k)}|^4 d^3\mathbf{r} + x_{i'}^{(k)} [\eta^{(k)}\sigma_{d_i}]^2 = 0\,, \quad 1 \le i' \le N\,,$$

which can be written in matrix form: $\mathbf{F}^{(k)}\,\mathbf{x}^{(k)} = \mathbf{0}$ – as in (3.2.10). ⊡

- **From a collection of local weighted-averages to an image**. Let us consider a series of query-point locations $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, ...)$, spanning a region of interest[†]; see Fig. 3.2.4(left). As a first remark, the associated generalized-inverse vectors $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ...)$ can be computed in an *embarrassingly parallel* fashion (since each system (3.2.10), associated to each query point, can be solved independently). Recall that from the generalized-inverse vectors, are inferred the weighted-averages $(\hat{m}^{(1)}, \hat{m}^{(2)}, ...)$, uncertainties $(\sigma_{\hat{m}^{(1)}}, \sigma_{\hat{m}^{(2)}}, ...)$, and averaging kernels $(A^{(1)}, A^{(2)}, ...)$. Clearly, it may be useful to visualize such a collection of local weighted-averages through a conventional tomographic image. It seems natural to do the following, as illustrated in Fig. 3.2.4(right):

  1. Consider a set of query points located on a given *tomographic grid*[††].
  2. Associate the estimate $\hat{m}^{(k)}$ to the grid's element to which $\mathbf{r}^{(k)}$ lies in.
  3. Show the grid's values as a conventional tomographic image.

  Note that one should not forget about resolution and uncertainty informations ($A^{(k)}$ and $\sigma_{\hat{m}^{(k)}}$) when interpreting features in an image, see Fig. 3.2.4. This comment is valid for *any* tomographic image. The uncertainties, $\sigma_{\hat{m}^{(k)}}$, can be visualized in the same way, but not the averaging kernels, $A^{(k)}$ – those should rather be visualized *individually*, especially for complex shapes (Zaroli, 2016).

[†] In the (SOLA) Backus-Gilbert inversion scheme, the region of interest (for which we want to get a collection of local weighted-averages) does *not* need to spread over a region that is large enough to encompass entirely all the ray paths.

[††] Various choices of grids are possible: 1) Regular grids, see Fig. 3.2.4 and Zaroli et al. (2017); 2) Irregular, data-driven grids, see Zaroli (2016, 2019).

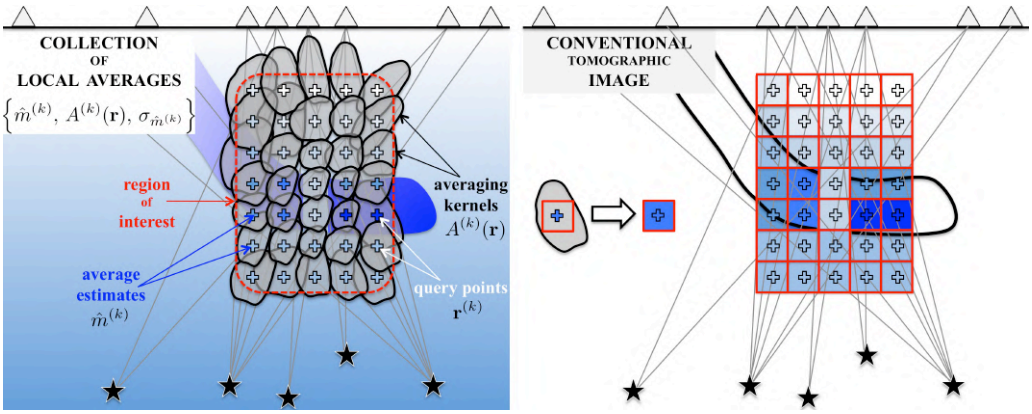


Figure 3.2.4: (Left) A collection of local weighted-averages can be displayed as (Right) a conventional tomographic image. Any structural feature – see the 'slab tear' – that shows up in such an image must be analyzed through careful investigation of the attached averaging kernels and uncertainties, to demonstrate whether it can (or not) be resolved given the data and their errors.

- **Major computational (and practical) issues**. First of all, there is a major computational issue which arises when attempting to set up Backus–Gilbert tomographic systems (3.2.10) for large-scale[†] tomographic applications. Indeed, for each newly considered query point (k), one has to set up an entirely new matrix $\mathbf{F}^{(k)}$, which is tremendous to compute, since it consists in $\sim N^2/2$

[†] High number of data and/or many query points (wide region of interest).

volumetric integrals $\int K_i(\mathbf{r})K_{i'}(\mathbf{r})|\mathbf{r}-\mathbf{r}^{(k)}|^4 d^3\mathbf{r}$ (see the explicit dependence on the query-point location $\mathbf{r}^{(k)}$).

Another computational, and practical, issue is the following: How should one select the trade-off parameters, $(\eta^{(1)}, \eta^{(2)}, ...)$, in the B–G minimization problems (3.2.9), when considering many query-point locations, $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, ...)$? How to choose them such that the – independently computed – generalized-inverse vectors, $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ...)$, lead to *globally coherent*† local averages, $(\hat{m}^{(1)}, \hat{m}^{(2)}, ...)$?

† While in the DLS approach one can easily 'see' if a given damping parameter produces a globally coherent tomographic image, in OLA methods (e.g., B–G) since only one query point (k) is treated at a time, one cannot see if each value for $\eta^{(k)}$ will, in the end, lead to a globally coherent image.

In my view, these issues certainly are some of the main reasons why, as stated by Nolet (2008): "*The Backus–Gilbert method is woefully underused in terrestrial [e.g., seismic tomography] applications*".

We shall see that the SOLA–Backus–Gilbert variant, that I have recently introduced and adapted to solve (large-scale) tomographic problems, overcomes the aforementioned issues – and thus allows one, finally (!), to experiment and exploit the Backus–Gilbert philosophy, developed half a century ago, for seismic tomography applications.

- **Additional remarks**. In my opinion, there may be a few more reasons why the B–G approach has been underused in tomographic applications.

  Firstly, when facing massive tomographic problems, Damped Least-Squares (DLS) methods still allow one to *directly compute*†† a model solution – thus avoiding the need for a 'computationally costly' generalized inverse. While in the B–G approach, to infer a model solution one cannot avoid computing a generalized inverse. However, if one is interested (and we should) in solving both the problems of model estimation *and* model appraisal – involving the computation of generalized-inverse operators – then the SOLA–B–G variant appears to be *computationally more efficient* than DLS (Zaroli et al., 2017).

  †† For example, using iterative row-action algorithms, like LSQR.

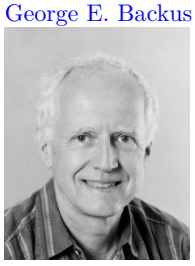
  Secondly, in the B–G approach, one may wonder whether a tomographic model assembled from a collection of – independently determined – local averages (i.e., the 'conventional images' built as previously described), may actually explain the data? Even though the B–G method does not specifically aim at minimizing the data misfit (in the way the generalized inverse is defined), I have reported that, at least for 'synthetic' seismic tomography experiments, the SOLA variant of B–G can produce *globally coherent* models, that actually do fit the data at the same level as DLS models (Zaroli et al., 2017); see Sect. 3.3.1. In my experience, the fact that SOLA tomographic models are fitting the data at a similar level as DLS models is also roughly indicated by their visual similarity, even when dealing with 'real' data sets (Zaroli, 2016; Zaroli et al., 2017; Latallerie, 2019; Ouattara, 2019; Dubois, 2020); see Sect. 3.3.1.

  As a final remark, if the region of interest does not encompass all the ray paths, as illustrated in Fig. 3.2.4, it is obvious that a (SOLA) B–G tomographic model cannot be used to predict the data, and thus one cannot formally check if it fits the data.

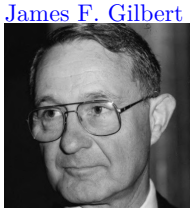
3.2.2 Continuous SOLA inversion

- **Brief historical perspective**.
  - **Late sixties** (Backus & Gilbert, 1967, 1968, 1970)
    Seminal papers by Backus & Gilbert, on a linear, continuous inversion scheme, that avoids using any *a priori* information on the model.
  - **Mid eighties** (Nolet, 1985)
    The Backus–Gilbert method was adapted by Nolet to 'discrete' linear inverse problems, and applied to a small-scale 2–D tomographic problem.
  - **Early nineties** (Pijpers & Thompson, 1992, 1994)
    The SOLA method was developed for 'helio-seismic' inversions by Pijpers & Thompson – an alternative Backus–Gilbert formulation which retains all its advantages, but is more computationally efficient and versatile in the explicit construction of averaging kernels. Pijpers & Thompson (1992) termed this variant 'SOLA'[†], though Larsen & Hansen (1997) suggest that it may have been rediscovered independently by different authors (e.g., Masters & Gilbert, 1983; Louis & Maass, 1990; Pijpers & Thompson, 1992). In this manuscript, I prefer to refer to the 'SOLA–Backus–Gilbert' approach (though I often use 'SOLA' for short). Although SOLA has been widely used in helioseismology (e.g., Pijpers, 1997; Rabello-Soares et al., 1999; Jackiewicz et al., 2012), a similar approach was used to investigate the attenuation and the resolution of density within the Earth (Masters & Gilbert, 1983; Masters & Gubbins, 2003).
  - **Before 2015**
    Many authors had considered that, in addition to a high computational cost, it could be a clumsy affair in the presence of data errors[††] to practically implement the Backus–Gilbert (B–G) method to large-scale tomographic applications (e.g., Menke, 1989; Parker, 1994; Trampert, 1998; Aster et al., 2012). To illustrate this, a few selected quotes[†††]:
    - "*The B–G theory has received more citation than actual application*" (Parker, 1994).
    - "*Being practically difficult to implement in the presence of data errors, resulted in the B–G method not finding many applications in seismic tomography*" (Trampert, 1998).
    - "*Though the B–G method receives much interest in helioseismology, it is woefully underused in terrestrial applications*" (Nolet, 2008).
    - "*[Because it is] computationally intensive the [B–G] method does not receive wide use*" (Aster et al., 2012).
  - **2016–2019** (Zaroli, 2016, 2019; Zaroli et al., 2017)
    I have introduced and adapted the SOLA–Backus–Gilbert inversion scheme to efficiently solve large-scale, linear, both discrete and continuous, 2–D and 3–D, seismic tomography problems.

George E. Backus



James F. Gilbert



[†] **S**ubtractive **O**ptimally **L**ocalized **A**verages

[††] With error-free data there would be less trouble in selecting the trade-off parameters, $(\eta^{(1)}, \eta^{(2)}, ...)$, since they would all be set to zero.

[†††] In 2015 (Sept.), when I decided (by pure chance) to work on the project of applying the B–G method to global seismic tomography, these quotes were not very encouraging, but challenged me!

- **The essence of SOLA**. The key idea of the SOLA method (Pijpers & Thompson, 1992) is to specify an *a priori* target form $T^{(k)}(\mathbf{r})$ for each averaging (resolving) kernel $A^{(k)}(\mathbf{r})$. That is, to specify some *a priori* information on the model resolution – which is fundamentally different from *a priori* information on the model itself (as in Bayesian-like methods). I often refer to those 'target resolving kernels', $T^{(k)}(\mathbf{r})$, as *target kernels* for short. In the case of our 3–D tomography example, a target kernel is then formally defined as:

$$T^{(k)}(\mathbf{r}) = \begin{cases} \frac{1}{\text{volume of } \mathbb{S}^{(k)}} & \text{if} \quad \mathbf{r} \in \mathbb{S}^{(k)} \\ \\ 0 & \text{elsewhere} \ , \end{cases} \tag{3.2.11}$$

which means that the target kernel $T^{(k)}$ is constant-valued within some (volumetric) region $\mathbb{S}^{(k)}$, which should be designed as follows:

1. $\mathbb{S}^{(k)}$ should be well 'localized' (in the model space),
2. $\mathbb{S}^{(k)}$ should be centred on the query-point location, $\mathbf{r}^{(k)}$;

see the 'yellow disk' in Fig. 3.2.5. The size of $\mathbb{S}^{(k)}$ represents the *a priori* information on the (local) resolving length. Its shape could be a ball[†] (i.e., isotropic *a priori* resolution) or a spheroid (i.e., lateral and radial *a priori* resolution). (3.2.11) implies that target kernels also satisfy to the averaging condition:

[†] If $T^{(k)}$ (i.e., $\mathbb{S}^{(k)}$) was a 3–D ball centred on $\mathbf{r}^{(k)}$, its radius would correspond to the minimum resolving length that could, at best, be reached in $\mathbf{r}^{(k)}$.

$$\int T^{(k)}(\mathbf{r})\, d^3\mathbf{r} = 1\,. \tag{3.2.12}$$

- **Continuous SOLA minimization problem**. Rather than minimizing the spread (spatial extent) of each averaging kernel $A^{(k)}$, as in the B–G approach, in the SOLA variant of B–G one aims at minimizing the integrated squared difference between each averaging kernel $A^{(k)}$ and its corresponding target kernel $T^{(k)}$. A minimization problem is then set up to directly compute each generalized-inverse vector $\mathbf{x}^{(k)}$, for each query point $\mathbf{r}^{(k)}$, which writes:

[††] In the SOLA variant of B–G, we wish the actual resolution, $A^{(k)}$, to be the closest possible to the specified *a priori* resolution, $T^{(k)}$.

$$\begin{cases} \underset{\mathbf{x}^{(k)} \in \mathbb{R}^N}{\arg\min}: & \underbrace{\int \left[A^{(k)}(\mathbf{r}) - T^{(k)}(\mathbf{r})\right]^2 d^3\mathbf{r}}_{\substack{\text{RESOLUTION MISFIT} \\ \text{(Favoring averaging- close to target-kernel}^{\dagger\dagger}\text{)}}} \quad + \quad \underbrace{[\eta^{(k)}]^2}_{\substack{\text{Trade} \\ \text{off}}} \quad \underbrace{\sigma^2_{\hat{m}^{(k)}}}_{\substack{\text{VARIANCE of } \hat{m}^{(k)} \\ \text{(Moderating uncertainty)}}} \\ \\ \text{subject to} & \underbrace{\int A^{(k)}(\mathbf{r}) d^3\mathbf{r} = 1\,.}_{\substack{\text{AVERAGING} \\ \text{CONDITION}}} \end{cases} \tag{3.2.13}$$

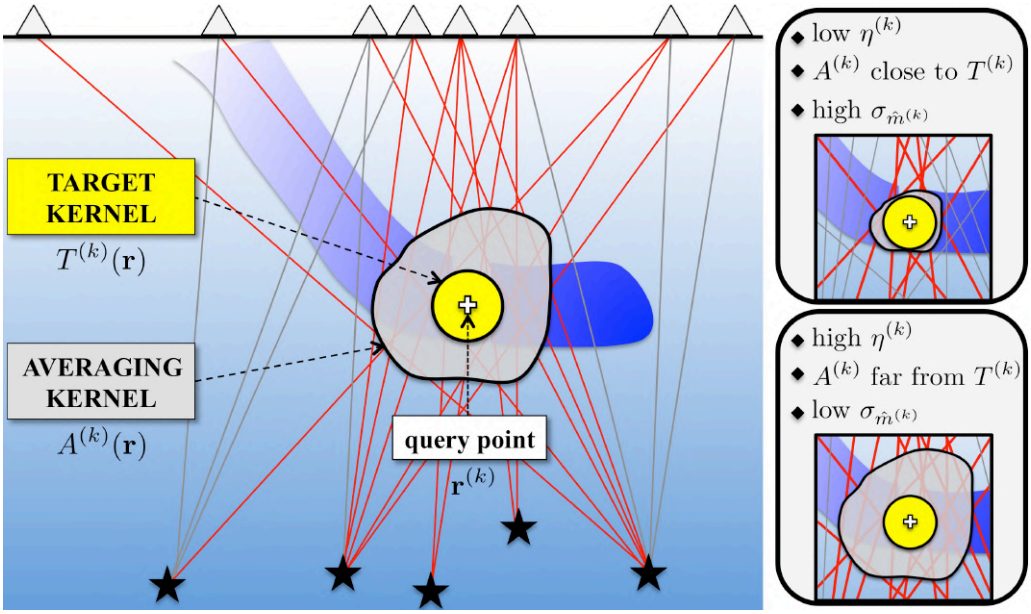


Figure 3.2.5: Illustration, for the SOLA–Backus–Gilbert inversion approach, of trade-off between poor/good resolution (i.e., poorly/optimally localized averaging kernel $A^{(k)}$ (grey shape), with respect to target kernel $T^{(k)}$ (yellow disk)), and low/high value of uncertainty $\sigma_{\hat{m}^{(k)}}$, respectively.

This leads to a compromise – via the trade-off parameter $\eta^{(k)}$ – between favoring the averaging kernel $A^{(k)}$ to be the closest possible to its associated target kernel $T^{(k)}$ (so that $A^{(k)}$ be *optimally localized*, with respect to $T^{(k)}$), while also moderating the uncertainty $\sigma_{\hat{m}^{(k)}}$, on the weighted-average estimate. That is, again, the unavoidable resolution/uncertainty trade-off; see Fig. 3.2.5. Note that the SOLA minimization problem is also constrained for averaging kernels to honor the averaging condition. Finally, (3.2.13) writes (see Text Box 3.4):

$$\begin{cases} \text{solve} \quad \mathbf{F}\,\mathbf{x}^{(k)} = \mathbf{u}^{(k)}, \\ \text{subject to} \quad \sum_{i=1}^{N} x_i^{(k)} \int K_i(\mathbf{r}) d^3\mathbf{r} = 1\,, \end{cases}$$
$$\text{with:} \begin{cases} F_{ii'} = \int K_i(\mathbf{r}) K_{i'}(\mathbf{r}) d^3\mathbf{r} + [\eta^{(k)} \sigma_{d_i}]^2 \delta_{ii'}^{\text{Kro}}\,, \; 1 \le i, i' \le N\,, \\ u_i^{(k)} = \int T^{(k)}(\mathbf{r}) K_i(\mathbf{r}) d^3\mathbf{r}\,, \; 1 \le i \le N\,. \end{cases} \tag{3.2.14}$$

The system (3.2.14) can be solved with the Lagrange-multipliers method (Pijpers & Thompson, 1992) or an LSQR-based approach (Zaroli, 2019).

Text Box 3.4: CONTINUOUS SOLA NORMAL EQUATIONS

Expanding out $A^{(k)}$ and $\sigma_{\hat{m}^{(k)}}$ in function of $\mathbf{x}^{(k)}$, the system (3.2.13) writes:

$$\begin{cases} \underset{\mathbf{x}^{(k)} \in \mathbb{R}^N}{\arg\min:} & \int \left[\sum_{i=1}^N x_i^{(k)} K_i(\mathbf{r}) - T^{(k)}(\mathbf{r}) \right]^2 d^3\mathbf{r} + [\eta^{(k)}]^2 \sum_{i=1}^N (x_i^{(k)})^2 (\sigma_{d_i})^2 \\ \text{subject to} & \sum_{i=1}^N x_i^{(k)} \int K_i(\mathbf{r}) d^3\mathbf{r} = 1 \,. \end{cases}$$

To obtain the normal equations corresponding to the 'unconstrained part' of the previous minimization problem, we do the following (i' is fixed):

$$\frac{\partial}{\partial x_{i'}^{(k)}} \left(\int \left[\sum_{i=1}^N x_i^{(k)} K_i(\mathbf{r}) - T^{(k)}(\mathbf{r}) \right]^2 d^3\mathbf{r} + [\eta^{(k)}]^2 \sum_{i=1}^N (x_i^{(k)})^2 (\sigma_{d_i})^2 \right) = 0 \,,$$

$$\Longrightarrow \quad \sum_{i=1}^N x_i^{(k)} \int K_i(\mathbf{r}) K_{i'}(\mathbf{r}) d^3\mathbf{r} + x_{i'}^{(k)} [\eta^{(k)} \sigma_{d_i}]^2 = \int T^{(k)}(\mathbf{r}) K_{i'}(\mathbf{r}) d^3\mathbf{r} \,, \quad 1 \le i' \le N \,,$$

which can be written in matrix form: $\mathbf{F}\,\mathbf{x}^{(k)} = \mathbf{u}^{(k)}$ – as in (3.2.14). ⊡

- **Computational considerations**. A major computational advantage of the SOLA variant is that the matrix $\mathbf{F}$ in (3.2.14) does *not* depend on the query-point location, $\mathbf{r}^{(k)}$. Therefore, the SOLA matrix $\mathbf{F}$ does not need to be re-computed for each newly considered query point, as opposed to the Backus–Gilbert matrix $\mathbf{F}^{(k)}$ in (3.2.10). In practice, this advantage is crucial, since it allows us to experiment, even with modest computational facilities, with the SOLA–Backus–Gilbert inversion philosophy for solving large-scale, linear seismic tomography problems. As illustrated in Fig. 3.2.6, to set up the continuous SOLA systems (3.2.14), one mainly has to compute (Zaroli, 2019):

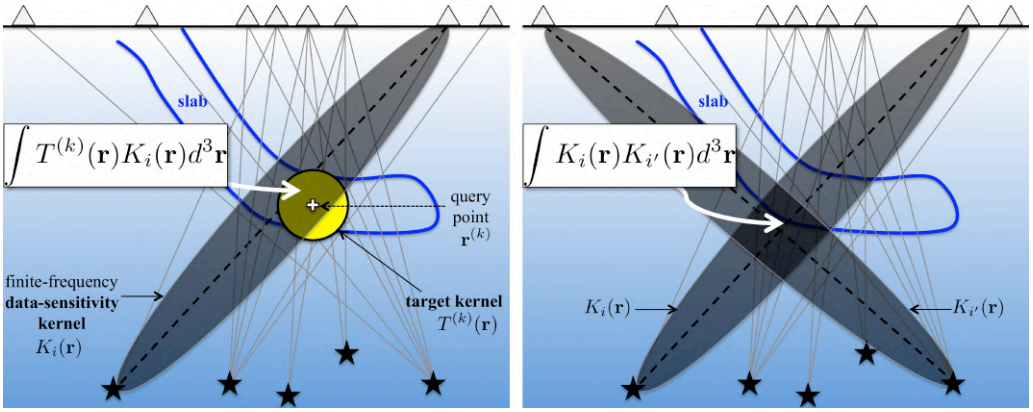


Figure 3.2.6: Visual illustration of (volumetric) integrals of the form (Left) $\int T^{(k)} K_i$ and (Right) $\int K_i K_{i'}$, to be computed to set up the continuous SOLA tomographic systems (3.2.14).

1. $P \times N$ (volumetric) integrals of the form $\int T^{(k)} K_i$, see Fig. 3.2.6(left), where P is the total number of query points, $T^{(k)}$ denotes a (3–D) target kernel, K_i is a (3–D) finite-frequency data sensitivity kernel, and N is the total number of data. Without loss of generality, I assume that there are much more data than query points, that is, $P/N \ll 1$.
2. $\sim N^2/2$ (volumetric) integrals of the form $\int K_i K_{i'}$, see Fig. 3.2.6(right), where $(K_i, K_{i'})$ is a pair of (3–D) finite-frequency data sensitivity kernels. Note that the computation of these integrals, $\int K_i K_{i'}$, is by far the most costly task involved in so-called *parameter-free*[†] *SOLA tomography*.

[†] Parameter-free means continuous.

- **SOLA target kernels and trade-off parameters**[††]. An important result of my research (Zaroli, 2016, 2019; Zaroli et al., 2017) has been to demonstrate, in the context of seismic tomography, a computationally efficient and versatile 'recipe' for designing appropriate shapes and sizes for target kernels, $T^{(k)}$, and selecting pertinent values for trade-off parameters, $\eta^{(k)}$ – so that it leads to *globally coherent* collections of local weighted-averages (although each local average is independently estimated). That recipe states the following:

  1. Use a simple form for target kernels, $T^{(k)}$, such as a disk for 2–D applications (isotropic *a priori* local resolving-length) or a spheroid for 3–D applications (isotropic, radial and lateral *a priori* local resolving-lengths).
  2. Use the data coverage (e.g., logarithmic scale of ray density) as a first-order proxy for the spatial variations of the *a priori* local resolving-length.
  3. Make an educated (physics-based) guess about the *a priori* resolving-length bounds (i.e., minimum and maximum radius of target kernels, $R^{\min}$ and $R^{\max}$, respectively).
  4. Interpolate (e.g., linearly) to determine the *a priori* resolving-length at a given query-point location, $\mathbf{r}^{(k)}$, and then use this value for the size (e.g., radius) of the associated target kernel, $T^{(k)}$.
  5. Choose a constant value for *all*[†††] the trade-off parameters, $\eta^{(k)} = \text{cst}$[††††]. The number of free hyper-parameters in the SOLA inversion scheme, hence the overall computational difficulty, is therefore greatly reduced (compared to having to select a potentially different value for each $\eta^{(k)}$).

[††] SOLA solutions, like with any inversion scheme, depend on tunable parameters, so that different choices would result in different models. Nevertheless, although one can not evade making specific choices for target kernels and trade-off parameters, it is important to realize that any particular solution represents *unbiased* averages, associated with different averaging volumes, that each can be quantitatively analyzed in conjunction with their respective model uncertainties.

[†††] It could also be an option to consider a different value for η for various regions.

[††††] In my view, using 'data-driven' target kernels $T^{(k)}$ allows a constant value for $\eta^{(k)}$ to lead to *globally coherent* collections of local averages.

Summary. A constant-valued trade-off parameter, $\eta^{(k)} = \text{cst}$, may lead to *globally coherent* (conventional) tomographic images when the size of target kernels, $T^{(k)}$, is set to spatially vary as ray density, that is, a proxy for the *a priori*, local model resolution – see Fig. 3.2.7.

As a remark, estimates of *a priori* local resolution are also crucial for tomographic studies that rely on data-driven, irregular tomographic grids (e.g., Nolet & Montelli, 2005; Zaroli et al., 2015). Finally, this practical recipe has proved to work well for various (both continuous and discrete) tomographic settings, from regional to global scale and with various seismological data (Zaroli, 2016, 2019; Zaroli et al., 2017; Latallerie, 2019; Ouattara, 2019; Dubois, 2020).

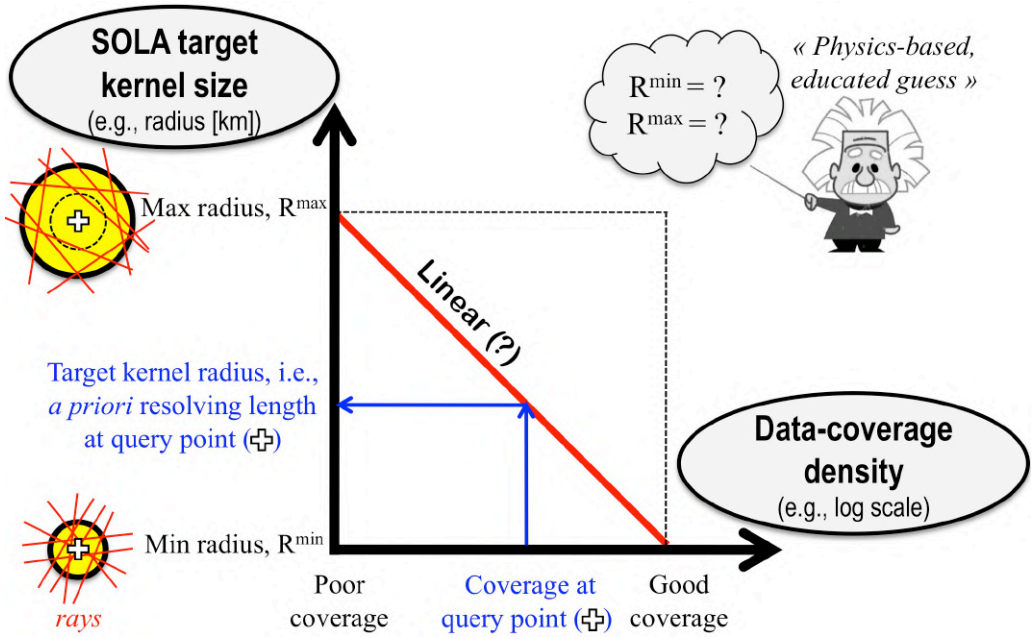


Figure 3.2.7: 'Data-driven' SOLA target kernels, for seismic tomography applications. Use the data coverage (e.g., logarithmic scale of ray density) as a first-order proxy for the spatial variations of the *a priori* local resolving-length. Make an educated (physics-based) guess about the *a priori* resolving-length bounds (i.e., minimum and maximum radius of target kernels, $R^{\min}$ and $R^{\max}$, resp.). Interpolate (e.g., linearly) to determine the *a priori* resolving-length at a given query-point location ($\mathbf{r}^{(k)}$), and use this value for the size (e.g., radius) of the associated target kernel ($T^{(k)}$).

- **Additional remarks**.

  1. The previously described continuous SOLA tomographic scheme is well suited for embarrassingly parallel[†] computations, since the problem can be easily separated into a number of independent tasks, for example[††] :
     - When computing all the P vectors $\mathbf{u}^{(k)} = (\int T^{(k)} K_i)_{1 \leq i \leq N}$;
     - When computing all the $\sim N^2/2$ integrals $\int K_i K_{i'}$;
     - When solving all the P systems (3.2.14).
  2. Since the number of integrals $\int K_i K_{i'}$ grows as a function of N^2, this may be a computational burden when facing large data sets (e.g., $N \gg 10^5$), especially if one aims at fully capturing the form of (e.g., finite-frequency body-wave) data-sensitivity kernels; see Sect. 3.2.3. Therefore, 'parameter-free' SOLA tomography may be better suited to handle moderate-size data sets, typically encountered at regional scales.
  3. To decrease the total computational cost, one may want[†††] to consider not too many query points, that is, not more than required to fit the spatial variations of the *a priori* local resolution.

[†] The same kind of parallelism applies to 'discrete' SOLA tomography.

[††] Recap: P and N are the number of query points and data, respectively.

[†††] This argument also holds for data-driven, irregular tomographic grids.

- **Example**. To illustrate how the SOLA–Backus–Gilbert approach can be exploited, in practice, I briefly present the results of a 'proof-of-concept' study on parameter-free SOLA tomography (Zaroli, 2019). That is, I have successfully inverted a moderate-size set of teleseismic S-wave time residuals, within a finite-frequency physical framework, with focus on imaging and appraising 3–D shear-wave velocity anomalies lying in the mantle below Southeast Asia.

  1. *Tomographic settings and numerical considerations.* The data[†] consist in $N = 27\,070$ teleseismic S and SS time residuals, measured by cross-correlation technique at 22 s central period (passband Gaussian filter). A selection criterion was applied to ensure that every data-sensitivity kernels, K_i, sample the Southeast-Asia region. The associated (4000) earthquakes and (250) receivers are shown in Fig.3.2.8(a). Data errors, σ_{d_i}, include earthquake-location and measurement-process errors. Each finite-frequency sensitivity kernel, K_i, is calculated at 22 s period on a grid[††] made of 50 km edge-length cubic cells, as illustrated in Fig.3.1.1. It takes $\sim$ 1 wk (CPU time) to compute in parallel, using 70 processors (Intel Xeon E5-4657L 2.40 GHz), all the $N^2/2$ integrals $\int K_i K_{i'}$ (with the aforementioned data-kernel discretization) – by far the most costly task.

[†] Frequency dependent crust and attenuation effects, as well as corrections for Earth's ellipticity and topography, were accounted for in the measurement process; see Zaroli et al. (2010).

[††] Each kernel K_i is discretized on a local Cartesian grid that consists in regularly spaced, 50 km edge-length, cubic cells spanning a rectangular parallelepiped region surrounding K_i. Such a kernel discretization allows one to capture all the characterizing sensitivity variations, while keeping low memory costs.

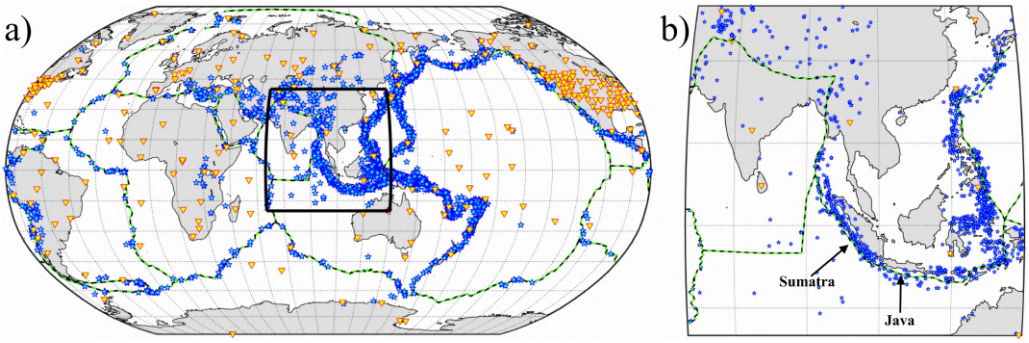
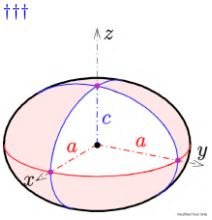


Figure 3.2.8: (a) Earthquakes (stars) and receivers (triangles) corresponding to the data set, and major tectonic plates (black-green dashed lines). (b) Zoom-in on 'Southeast Asia' (black frame). Modified from Zaroli (2019).

  2. *SOLA target kernels and query points.* Each target kernel, $T^{(k)}$, is chosen to be a spheroid[†††] centred on a query point, $\mathbf{r}^{(k)}$. The 'recipe' sketched in Fig. 3.2.7 is followed to specify the size of *a-priori-resolution-driven* target kernels. To limit the number of query points, their locations are determined such that they fit the spatial distribution of the *a priori* local resolving length. A total of $P = 4310$ query points are considered here, spanning Southeast Asia at 8 different depths, in the 350–1410 km range. Fig. 3.2.9 displays lateral views (i.e. circles) of some target kernels. Their lateral radius, driven by 'ray density', ranges from 200 to 1000 km and represents the *a priori*, local, *lateral* resolving length. Their radial radius

[†††]



gradually varies from 130 to 200 km in the 350–1410 km depth range, respectively, and represents the *a priori*, local, *radial* resolving length.

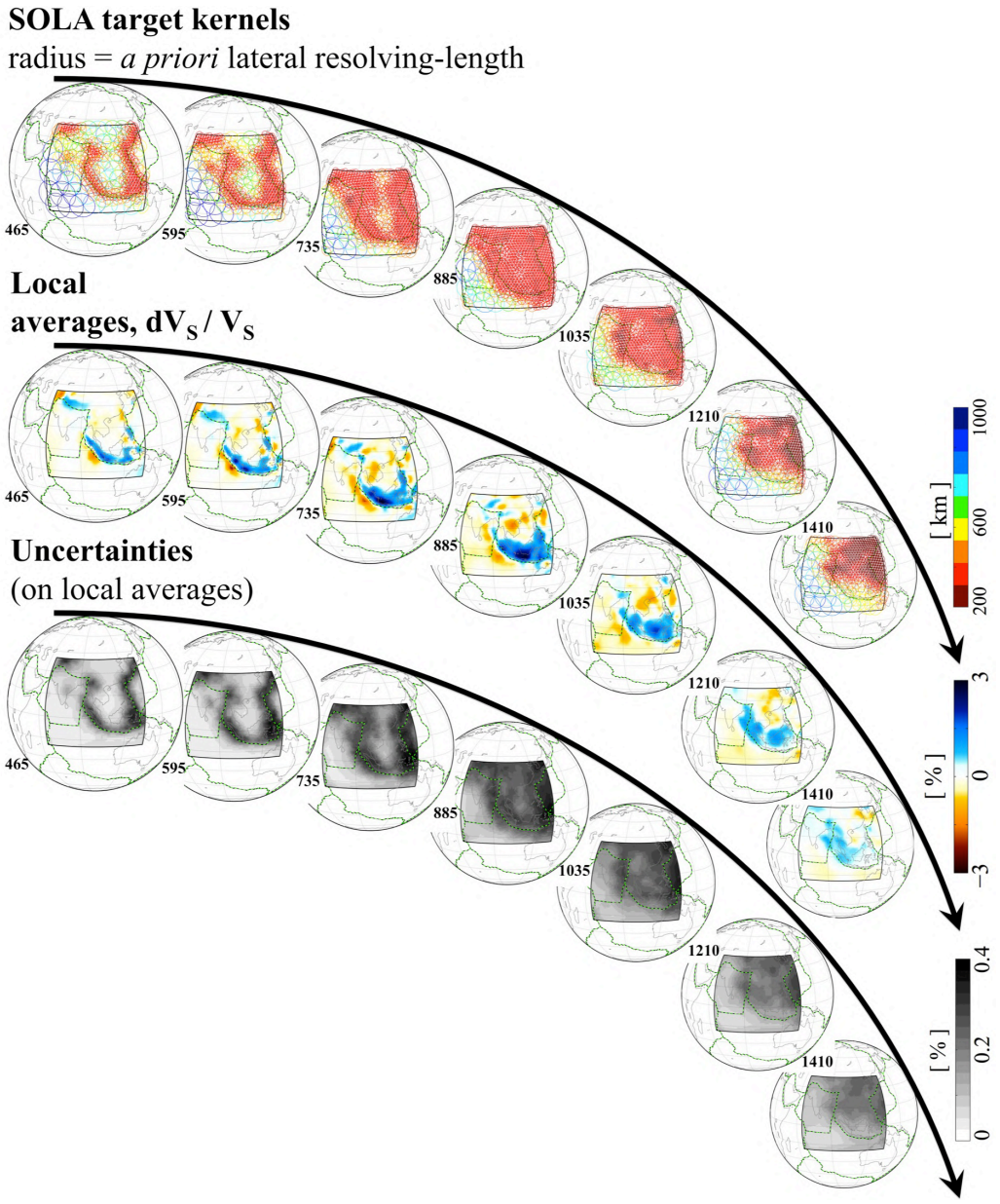


Figure 3.2.9: (First row) Drawn circles represent lateral, 2–D views of target kernels $T^{(k)}$ (here at 465, 595, 735, 885, 1035, 1210 and 1410 km depth). A query point $\mathbf{r}^{(k)}$ lies at the centre of each circle, whose radius is colour coded and ray-density driven. (Second row) 'Conventional' tomographic images (dV_S/V_S) built from the collection of local averages $\hat{m}^{(k)}$. (Third row) 'Conventional' images built from the uncertainties $\sigma_{\hat{m}^{(k)}}$ (on the averages). Modified from Zaroli (2019).

3. *SOLA inversion and tomographic results.* A constant value is chosen for the trade-off parameters, i.e. $\eta^{(k)} = \text{cst}$, after having tested a few different values. All the P systems (3.2.14) are solved with the LSQR algorithm (Paige & Saunders, 1982; Zaroli, 2019), in an embarrassingly parallel fashion (parallelization over the P systems). This leads to P generalized-inverse vectors, $\{\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(P)}\}$, from which can be inferred (see (3.2.8)) the weighted-average estimates, $\{\hat{m}^{(1)}, \cdots, \hat{m}^{(P)}\}$, uncertainties (on the averages), $\{\sigma_{\hat{m}^{(1)}}, \cdots, \sigma_{\hat{m}^{(P)}}\}$, and averaging kernels, $\{A^{(1)}, \cdots, A^{(P)}\}$. As previously discussed and sketched in Fig. 3.2.4, 'conventional' tomographic images can be built from such a set of local weighted-averages (or uncertainties) in the region of interest – through a linear interpolation process[†]. Finally, Fig. 3.2.9 displays[††] some of the resulting *local-average* and *uncertainty-on-local-average* 'conventional' images; so does Fig. 3.2.10, including three examples of 3–D views of averaging kernels.

[†] Here, I make use of Delaunay meshes and barycentric coordinates for laterally interpolating the local averages.

[††] As expected, the spatial variations of uncertainties are similar to those of the ray-density driven target kernels. In the regions where the size of target kernels is large (small), that is, the *a priori* local resolution is poor (good), the uncertainty is low (high), respectively. That is, the previously discussed trade-off between resolution and uncertainty.

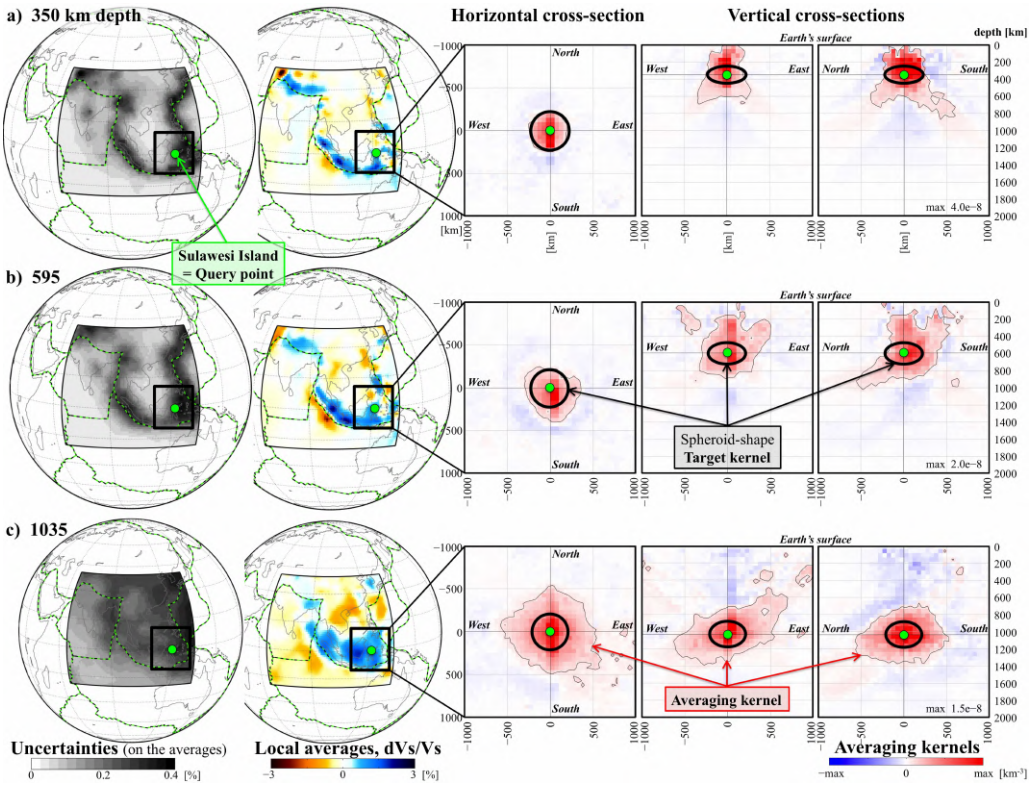


Figure 3.2.10: Cross-sections across 3 averaging kernels, associated to 3 query points (green dots) located at 350, 595 and 1035 km depth below the Sulawesi island. Conventional images of local averages and uncertainties (on the averages) are shown. Black circles (ellipses) represent horizontal (vertical) extents of 'spheroid' target kernels, respectively. Modified from Zaroli (2019).

4. *Asking a question and getting a quantitative answer (model appraisal).*
   - **Observation**. The lateral extent of the Sumatra–Java slabs (dV_S/V_S) appears to be smaller in upper mantle (Fig. 3.2.10(a)) and transition zone (Fig. 3.2.10(b)), compared to mid lower mantle (Fig. 3.2.10(c)).
   - **Question**. Is the apparent 'lateral flattening' of the Sumatra–Java slabs in mid lower mantle resolved?
   - **Resolving length**. To have a clue on the variations with depth of the lateral resolving length in this region, let consider the three averaging kernels ($A^{(k)}$) shown in Fig. 3.2.10 and estimate their lateral extent, that is, the local lateral resolving length. We can report that, below 'Sulawesi Island', the lateral resolving length is, at most, 200 km (300, 500[†]) at a depth of 350 km (595, 1035), respectively. In particular, Fig. 3.2.10(c) indicates[††] that the lateral extent of these slabs, around Sulawesi Island, at 1035 km depth is much larger (of order 2000 km) than the local lateral resolving length (of order 500 km).
   - **Local averages/uncertainties**. Moreover, local weighted-average estimates, $\hat{m}^{(k)}$, are 0.95, 0.92 and 1.03 per cent for the corresponding query points below Sulawesi Island at 350, 595 and 1035 km depth, while attached uncertainties (on the weighted-averages), $\sigma_{\hat{m}^{(k)}}$, are three times smaller, that is, 0.33, 0.30 and 0.29, respectively.
   - **Answer**. Thus, one can argue that the slab lateral flattening that takes place in mid lower mantle, at least below Sulawesi Island, is 'resolved' (given the data and their errors).

[†] Indeed, 500 km is the *lateral radius* of the averaging kernel whose the query point is located at 1035 km depth.

[††] Visually compare the 'slabs' inside the black-solid-line square, at 1035 km depth, with the corresponding averaging kernel (horizontal cross-section).

3.2.3 Discrete SOLA inversion

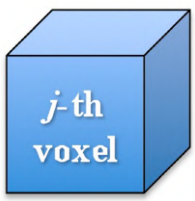
- **Model parameterization**. One may have to parameterize (i.e. discretize) the continuous 'true' model, $m(\mathbf{r})$, with a finite[†] set of M *model parameters*, $(m_j)_{1\leq j\leq M}$, through the use of some *basis functions*, $(h_j(\mathbf{r}))_{1\leq j\leq M}$, such that:

$$m(\mathbf{r}) \approx \sum_{j=1}^{M} m_j \, h_j(\mathbf{r}) \, . \tag{3.2.15}$$

[†] (i.e. $M < \infty$)

- **Local parameterization**. Having in mind the SOLA tomographic approach, let consider basis functions with *local support* (in space), as for example:

$$h_j(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in j\text{-th voxel} \\ 0 & \text{elsewhere} \, , \end{cases} \tag{3.2.16}$$



[††]

where the model space (e.g. Earth's mantle) would be filled in with M voxels[††].

For our 3–D[†] tomography example, let $\mathbb{V}_j$ denotes the *volume* of j-th voxel:

[†] For 2–D tomography, $\mathbb{V}_j$ would then be the *area* of j-th (2–D) cell.

$$\mathbb{V}_j = \int h_j(\mathbf{r}) d^3\mathbf{r} = \int_{\substack{\mathbf{r}\in j\text{-th} \\ \text{voxel}}} d^3\mathbf{r} \,. \tag{3.2.17}$$

For such a voxel-like parameterization, each model parameter m_j may be 'viewed' as the average value of the continuous true model within j-th voxel:

$$m_j \equiv \frac{\int m(\mathbf{r}) h_j(\mathbf{r}) d^3\mathbf{r}}{\int h_j(\mathbf{r}) d^3\mathbf{r}} = \frac{1}{\mathbb{V}_j} \int_{\substack{\mathbf{r}\in j\text{-th} \\ \text{voxel}}} m(\mathbf{r}) d^3\mathbf{r} \,. \tag{3.2.18}$$

Sometimes, one may want to consider voxels of *various size* ($\mathbb{V}_j$ may differ in function of index j), leading to *irregular* parameterizations[††], see Fig. 3.2.11.

[††] Irregular parameterizations usually are, in some sense, driven by the data (e.g., Sambridge et al., 1995; Nolet & Montelli, 2005; Zaroli et al., 2015).

- **Discrete tomographic problem**. The continuous, linear tomographic problem (3.1.1) can therefore be written in a discrete fashion as follows:

$$d_i = \underbrace{\sum_{j=1}^{M} G_{ij} \times \underbrace{m_j}_{\text{true parameters}}}_{\textbf{'discrete' forward problem}} + \underbrace{n_i}_{\text{'discrete' noise term}} \,, \; 1 \le i \le N \,, \tag{3.2.19}$$

where the elements G_{ij} (of the so-called sensitivity matrix $\mathbf{G}$) are as follows[†††]:

$$\begin{cases} G_{ij} = \underbrace{\int K_i(\mathbf{r}) h_j(\mathbf{r}) d^3\mathbf{r}}_{\substack{\text{sensitivity kernel } K_i \text{ projected} \\ \text{on model parameterization}}} = K_{ij} \times \mathbb{V}_j \\ K_{ij} = \frac{1}{\mathbb{V}_j} \int_{\substack{\mathbf{r}\in j\text{-th} \\ \text{voxel}}} K_i(\mathbf{r}) d^3\mathbf{r} = \text{average of } K_i \text{ in } j\text{-th voxel} \,. \end{cases} \tag{3.2.20}$$

[†††] Remarks:
1) Note that: $G_{ij} = K_{ij} \times \mathbb{V}_j$, where K_{ij} stands for the average of the *continuous* sensitivity kernel $K_i(\mathbf{r})$ in j-th voxel, with 'volume' $\mathbb{V}_j$. Here, $(K_{ij})_{1\le j\le M}$ is named a *discrete* sensitivity kernel.
2) For our running example (i.e., 3–D, finite frequency, $\frac{dV_S}{V_S}$ tomography), the physical units of d_i, n_i, $K_i(\mathbf{r})$, G_{ij}, $m(\mathbf{r})$, m_j, $\mathbb{V}_j$ and K_{ij} are: s, s, s.km^{-3}, s, percent, percent, km^3 and s.km^{-3}, respectively.

Fig. 3.2.12 shows a *continuous* finite-frequency *SS* sensitivity kernel $K_i(\mathbf{r})$ and corresponding *discrete* kernel $(K_{ij})_{1\le j\le M}$, for a voxel-like parameterization. Note that the 'discrete' noise term, i.e. n_i in (3.2.19), can be decomposed as:

$$\underbrace{n_i}_{\substack{\text{'discrete' noise} \\ \text{term in (3.2.19)}}} \longleftarrow \underbrace{n_i}_{\substack{\text{'continuous' noise} \\ \text{term in (3.1.1)}}} + \underbrace{\left(\int K_i(\mathbf{r}) m(\mathbf{r}) d^3\mathbf{r} - \sum_{j=1}^{M} G_{ij} m_j \right)}_{\substack{\text{Parametrization-related, data prediction error} \\ \text{('continuous' minus 'discrete' forward problems)}}} \,. \tag{3.2.21}$$

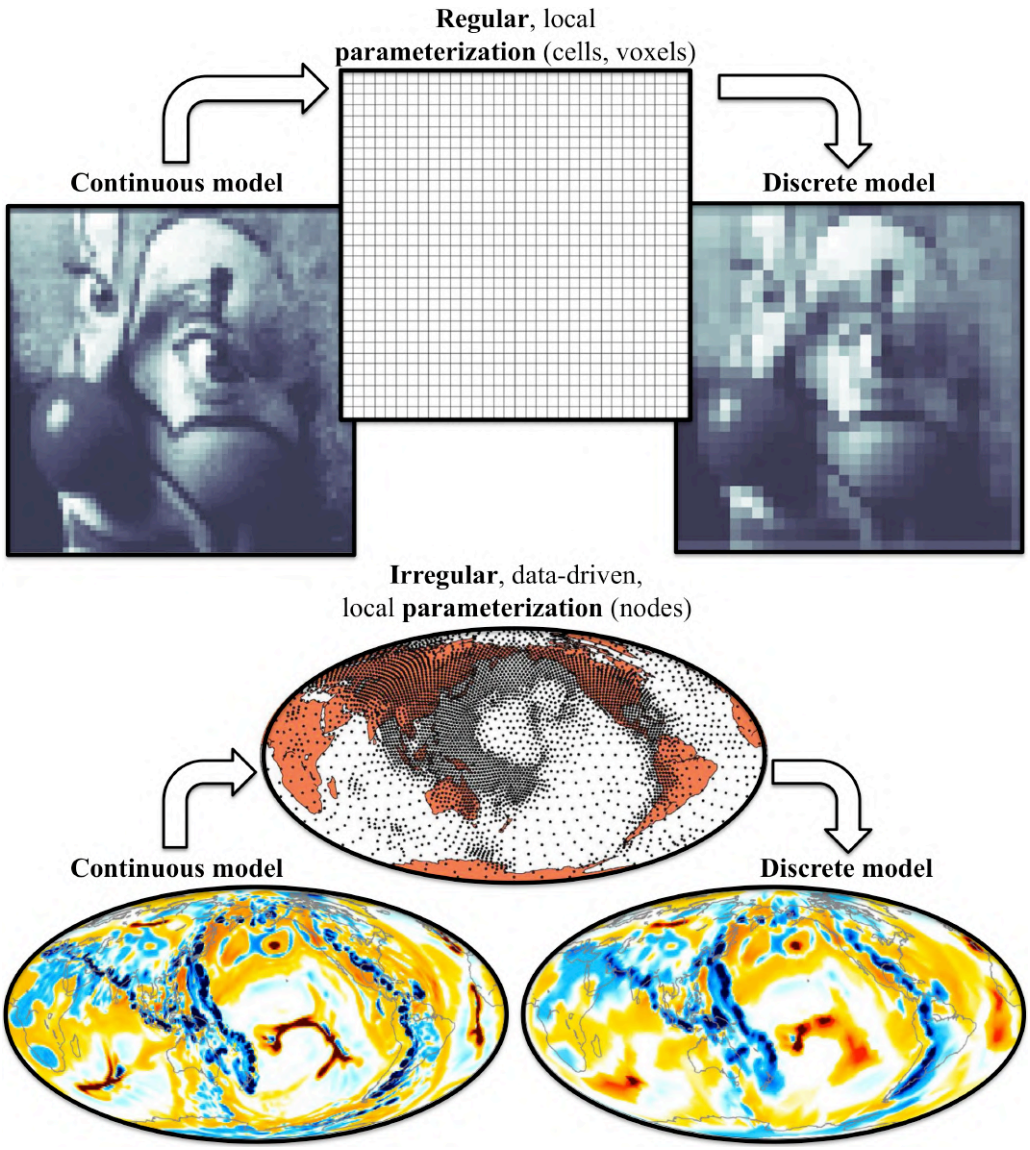


Figure 3.2.11: Parameterization of model space – I. (Top) Example of a *regular*, local model parameterization, that consists in 2–D square cells. The continuous 'true' model (clown picture), $m(\mathbf{r})$, is discretized, i.e. turned into a finite set of $M = 32 \times 32$ true-model parameters, $(m_j)_{1 \leq j \leq M}$. (Bottom) Example of an *irregular*, data-driven, local parameterization. Here, it consists in laterally parameterizing the Earth's mantle with a set of nodes (Delaunay mesh) whose spatial distribution is adapted to ray density (i.e. a proxy for *a priori* local resolution). See Zaroli (2010) and Zaroli et al. (2015) for details on such a nodes-based parametrization. The displayed dV_S/V_S 'continuous' model is N16-EB16 (885 km depth), that is, a *geodynamic* Mantle Circulation Model (Nerlich et al., 2016) – modified from Freissler et al. (2020). Note that, since one cannot know what the 'true' model actually looks like, one cannot ensure that a specific parameterization of model space is sufficiently accurate, to fully capture the form and amplitude of any of its structural features.

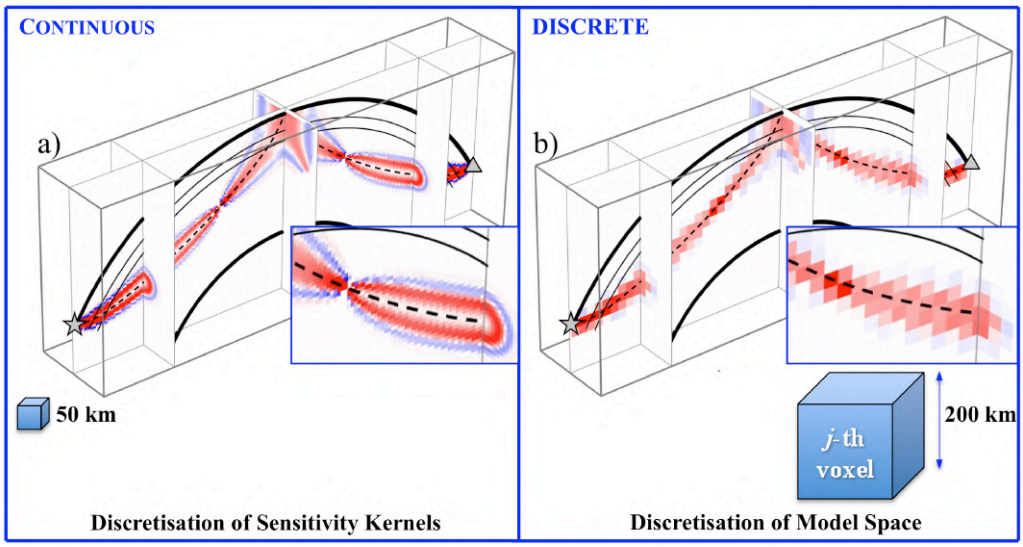


Figure 3.2.12: Parameterization of model space – II. (a) 3–D, finite-frequency, *SS*-phase time-residual (continuous) sensitivity kernel $K_i(\mathbf{r})$ (120° epicentral distance, 22 s central period of a passband Gaussian filter). It is discretized on a local Cartesian grid made of regularly spaced 50 km edge-length cubic cells, spanning a rectangular parallelepiped region (thick grey solid line). The Earth's surface and core–mantle boundary (transition zone) are depicted with thick (thin) black solid lines, respectively, and the earthquake and receiver with a star and triangle. (b) Same *SS* sensitivity kernel but after projection on a tomographic grid that consists in regularly spaced 200 km edge-length cubic cells (i.e. voxels); note that the projected sensitivity kernel (i.e. $(K_{ij})_{1\leq j\leq M}$, see (3.2.20)) looks like a 'fat' ray-theoretical kernel – that is, almost no sensitivity variation all around the geometrical ray path (black dashed line). Modified from Zaroli (2019).

Let assume that the additional, parametrization-related, data prediction error in (3.2.21) is also zero-mean and normally distributed, so that the 'discrete' data covariance matrix remains diagonal: $\mathbf{C}_d = \mathrm{diag}(\sigma_{d_i}^2)_{1\leq i\leq N}$, with elements:

$$\underbrace{(\sigma_{d_i})^2}_{\substack{\text{'discrete' data}\\ \text{variance}}} \longleftarrow \underbrace{(\sigma_{d_i})^2}_{\substack{\text{'continuous' data}\\ \text{variance}}} + \underbrace{\left(\sigma_{d_i}^{\text{para}}\right)^2}_{\substack{\text{variance due to the term, in (3.2.21), of}\\ \text{parametrization-related data prediction error}}} . \qquad (3.2.22)$$

In the course of Dubois (2020)'s PhD thesis, we have estimated $\sigma_{d_i}^{\text{para}}$ to be of order 0.5–1 s (1–1.5 s) for teleseismic *S* (*SS*, respectively) cross-correlation time residuals measured at 22 s central period – while using a data-driven, nodes-based, mantle parameterization (see Fig. 3.2.11) for the 'discrete' forward problem, as well as various input, Earth-like, mantle models (dV_S/V_S).

- **Continuous *versus* discrete inversion frameworks**. The 'continuous' SOLA–Backus–Gilbert inversion scheme is carried out in such a way that the

model space can retain its infinite dimensional nature[†], and a specific, somehow subjective, model parametrization need never be introduced (Sects. 3.2.1–2). In practice, of course, it is necessary to discretize the data-sensitivity kernels to perform numerical calculations. But there is a very important difference between discretizing sensitivity kernels, $K_i(\mathbf{r})$, and discretizing the model, $m(\mathbf{r})$; see Figs. 3.2.11–3.2.12. Sensitivity kernels are determined through some well-defined calculation, and one can ensure that a discretization is sufficient to fully capture their form; see Fig. 3.2.12(a). With the model, however, we can never know what it actually looks like, and so cannot verify that a given parametrization is sufficiently accurate (to fully capture the form and amplitude of any of its features); see Fig. 3.2.11. The 'continuous' SOLA–Backus–Gilbert framework enables numerically accurate sensitivity kernels to be *effectively* exploited in tomographic inversions. This may be viewed as a major benefit compared to 'discrete' tomographic methods, for which sensitivity kernels are often inaccurate[††], see Fig. 3.2.12(b), as they are projected on some model parametrization, see (3.2.20), prior to be exploited in the inversion – and these parameterizations are usually 'coarse' to limit the number of parameters (M) and keep computationally tractable the problems of model estimation and/or appraisal.

[†] It is this fact that really motivates the entire Backus–Gilbert approach, which seeks not to construct a particular model, but instead to determine *which* properties of the 'true' model can be estimated from the given data.

[††] Consequences would be that some, if not all, projected sensitivity kernels would become unsuitable for finite frequency imaging purposes; see Fig. 3.2.12(b).

- **SOLA tomography: continuous or discrete?** I aim to discuss when 'continuous' SOLA tomography should be preferentially used, rather than 'discrete' SOLA tomography, and vice versa. Let M^∞ be the minimum number of parameters required to parameterise the entire model space, so that every projected data-sensitivity kernels are accurate. In the context of teleseismic, finite-frequency, S-wave mantle tomography, it leads to $M^\infty \simeq 10^6$–10^7. If one aims at fully exploiting finite-frequency theory, but cannot handle discrete SOLA inversions with M^∞ parameters, then one should definitely use the 'continuous' SOLA approach. However, if the total number of data is too high, for example $N \gg 10^5$, it may not be tractable to compute the $\sim N^2/2$ elements of the matrix $\mathbf{F}$, see (3.2.14). Hence, one may have no choice but to move toward a 'discrete' SOLA approach with a total number of parameters $M \ll M^\infty$, and thus, as we shall see, simply have to project N sensitivity kernels on a given, coarse tomographic grid (to build the matrix $\mathbf{G}$, see (3.2.20)).

  Summary. When facing very large-scale data sets, and provided 'modest' computational facilities, one may have no choice but to apply the SOLA–Backus–Gilbert tomographic recipe[†††] in a *discrete* fashion.

[†††] See Sect. 3.2.2.

- **Discrete SOLA tomography**. In the following, I intend to emphasize on the main differences between discrete and continuous SOLA tomography, while assuming a local, voxel-like model parameterization. Note that all the key 'ingredients' of the recipe for continuous SOLA tomography, introduced earlier (e.g. tuning SOLA target kernels and trade-off parameters), remain the same.

  1. First of all, one explicitly seeks a k-th (model) parameter estimate, $\hat{m}_k$, which represents a weighted average over the 'true' parameters, $(m_j)_{1\leq j\leq M}$.

This 'discrete' averaging process[†] takes place through a discrete averaging kernel, $(A_j^{(k)})_{1\leq j\leq M}$, that we wish to be optimally localized around the k-th voxel, as illustrated in Fig. 3.2.13[††]. That leads to writing:

$$\hat{m}_k = \underbrace{\sum_{j=1}^{M} \underbrace{\left(\mathbb{V}_j A_j^{(k)}\right)}_{R_{kj}} m_j}_{\text{'DISCRETE' AVERAGING PROCESS}} \quad (+ \text{ propagated noise}), \tag{3.2.23}$$

[†] For 3–D, dV_S/V_S tomography, the physical units of $\hat{m}_k$, $\mathbb{V}_j$, $A_j^{(k)}$ and m_j are: percent, km^3, km^{-3} and percent, respectively. R_{kj} comes without any unit.

[††] Although model space obviously is 2–D in Fig. 3.2.13, the 'continuous' average is written as if it were 3–D, for consistency with main text (of this chapter).

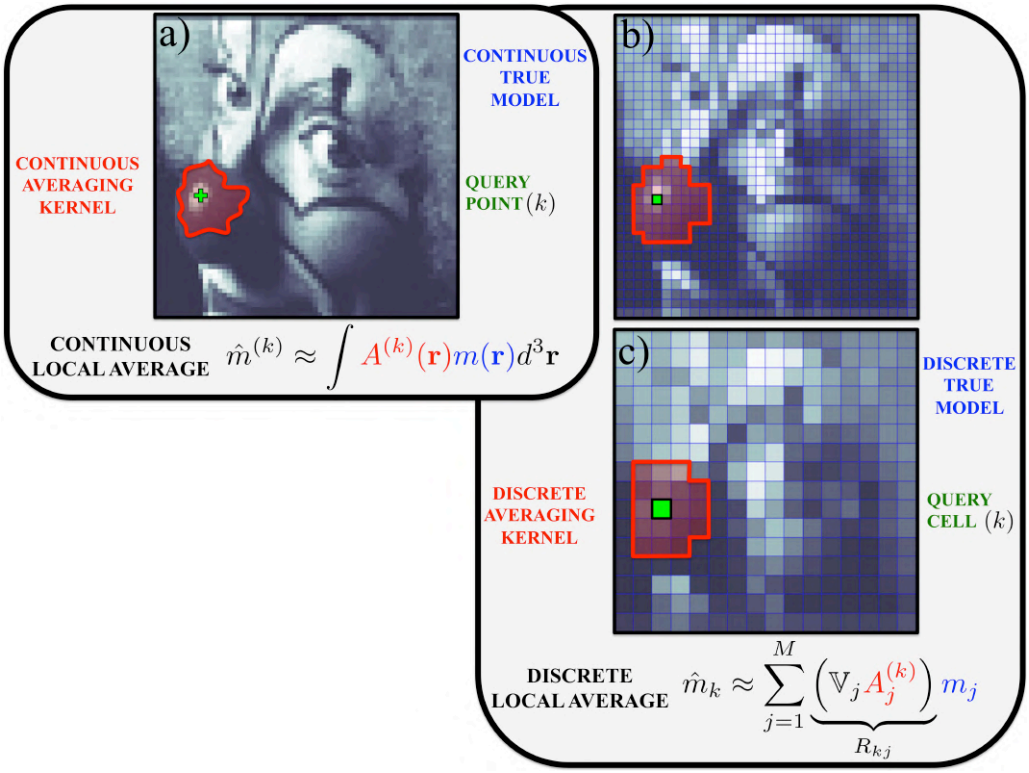


Figure 3.2.13: "*Resolution is no joke!*" (C. W. Harris, personal communication) — (a) The clown picture represents a continuous 'true' model, $m(\mathbf{r})$; the red contour-line is a continuous averaging kernel, $A^{(k)}(\mathbf{r})$, and the green cross the query-point location, $\mathbf{r}^{(k)}$. In the continuous SOLA (Backus–Gilbert) approach, one explicitly seeks a local (weighted) average of the form: $\hat{m}^{(k)} \approx \int A^{(k)}(\mathbf{r}) m(\mathbf{r}) d^3\mathbf{r}$. (b) Model $m(\mathbf{r})$ is turned (discretized) into $M = 32 \times 32$ (16×16 for (c)) true-model *parameters*, $(m_j)_{1\leq j\leq M}$; the red line represents a discrete averaging kernel, $(A_j^{(k)})_{1\leq j\leq M}$, and the green square the query cell (k-th parameter) centered on $\mathbf{r}^{(k)}$. In the discrete SOLA approach, one seeks a local average of the form: $\hat{m}_k \approx \sum_{j=1}^{M} R_{kj} m_j$, where $R_{kj} = \mathbb{V}_j A_j^{(k)}$.

where the components $(A_j^{(k)})_{1\leq j\leq M}$ have to be non-negative and to satisfy to the following *unbiased* (discrete) averaging condition:

$$\sum_{j=1}^{M} \mathbb{V}_j A_j^{(k)} = \sum_{j=1}^{M} R_{kj} = 1\,, \tag{3.2.24}$$

where, as we shall see, the elements $(R_{kj})_{1\leq j\leq M}$ can be viewed as the k-th row of a so-called 'resolution matrix', $\mathbf{R}$.

2. Since the 'discrete' tomographic problem is linear (that is, $d_i = \sum_j G_{ij} m_j + n_i = \sum_j (\mathbb{V}_j K_{ij}) m_j + n_i$), one can seek the k-th parameter estimate $\hat{m}_k$ as a linear combination of the data:

$$\hat{m}_k = \sum_{i=1}^{N} x_i^{(k)} d_i\,, \tag{3.2.25}$$

where $\mathbf{x}^{(k)} = (x_i^{(k)})_{1\leq i\leq N}$ are *some* generalized-inverse coefficients[†].

[†] The notation $\mathbf{x}^{(k)}$ is kept the same for both the 'continuous' and 'discrete' inversion schemes. However, it should be clear from the context to which $\mathbf{x}^{(k)}$ one refers to.

3. From combining the previous equations, one finds that the weighted-average parameter estimate $\hat{m}_k$, in (3.2.23), can be written out as:

$$\hat{m}_k = \underbrace{\sum_{j=1}^{M} \underbrace{\left(\mathbb{V}_j \underbrace{\left(\sum_{i=1}^{N} x_i^{(k)} K_{ij} \right)}_{A_j^{(k)}} \right)}_{R_{kj}} m_j}_{\text{'DISCRETE' AVERAGING PROCESS}} + \underbrace{\sum_{i=1}^{N} x_i^{(k)} n_i}_{\substack{\text{PROPAGATED NOISE} \\ \text{('discrete' noise in (3.2.19))}}} . \tag{3.2.26}$$

In the discrete SOLA approach, a *discrete averaging kernel*, $(A_j^{(k)})_{1\leq j\leq M}$, is then explicitly defined as a linear combination of N *discrete sensitivity kernels* $(K_{ij})_{1\leq j\leq M}$ (K_{ij} stands for the average of $K_i(\mathbf{r})$ in j-th voxel), whose weights are given by the generalized-inverse coefficients $(x_i^{(k)})_{1\leq i\leq N}$:

$$A_j^{(k)} = \sum_{i=1}^{N} x_i^{(k)} K_{ij} = \sum_{i=1}^{N} x_i^{(k)} \left(G_{ij} \mathbb{V}_j^{-1} \right) . \tag{3.2.27}$$

Fig. 3.2.14 shows a comparison between continuous / discrete averaging kernels, built from combining continuous / discrete sensitivity kernels.

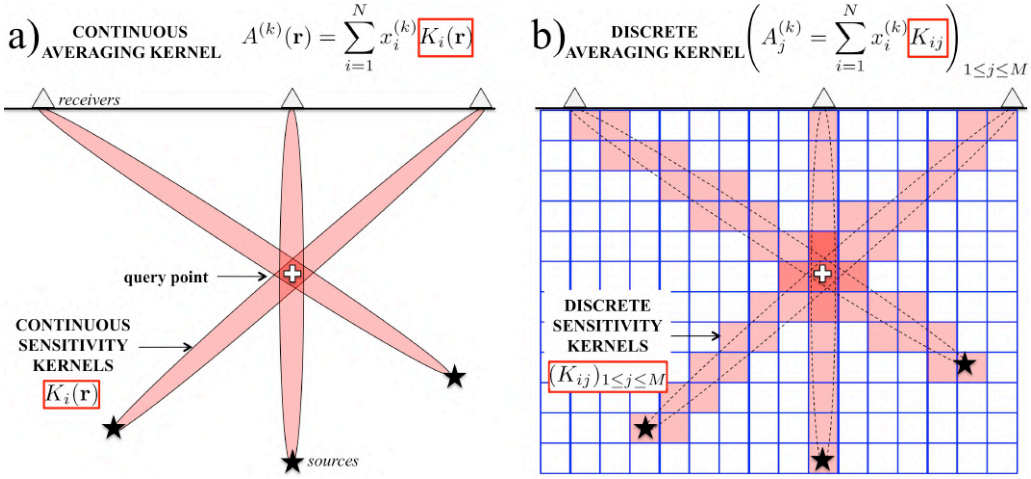


Figure 3.2.14: Illustration (similar to Fig. 3.2.2) showing that: (a) In the continuous SOLA–Backus–Gilbert approach, a continuous averaging kernel, $A^{(k)}(\mathbf{r})$, is explicitly defined as a linear combination of continuous sensitivity kernels, $K_i(\mathbf{r})$, whose weights are given by the generalized-inverse coefficients – that is, $(x_i^{(k)})_{1 \le i \le N}$, corresponding to the *continuous* SOLA minimization problem (3.2.13). (b) While in the discrete SOLA–Backus–Gilbert approach, a discrete averaging kernel, $(A_j^{(k)})_{1 \le j \le M}$, is defined as a combination of discrete sensitivity kernels, $(K_{ij})_{1 \le j \le M}$, whose weights are given by the generalized-inverse coefficients – that is, $(x_i^{(k)})_{1 \le i \le N}$, corresponding to the *discrete* SOLA minimization problem (3.2.33). Note that, here, the cartesian grid (blue lines) represents a local, regular, voxel-like parameterization of model space.

As a remark, the elements R_{kj} can be expressed as:

$$R_{kj} = \mathbb{V}_j A_j^{(k)} = \mathbb{V}_j \left(\sum_{i=1}^{N} x_i^{(k)} \underbrace{K_{ij}}_{\frac{G_{ij}}{\mathbb{V}_j}} \right) = \sum_{i=1}^{N} x_i^{(k)} G_{ij} \, . \tag{3.2.28}$$

Statistically, the weighted-average parameter estimate $\hat{m}_k$ writes:

$$\hat{m}_k \quad \equiv \quad \underbrace{\sum_{j=1}^{M} \underbrace{\left(\mathbb{V}_j A_j^{(k)} \right)}_{R_{kj}} m_j}_{\text{'DISCRETE' AVERAGING PROCESS}} \quad + \quad \underbrace{\mathcal{N}(0, \sigma^2_{\hat{m}_k})}_{\substack{\textbf{statistical propagation} \\ \textbf{of noise} \text{ (normal distribution)}}} \quad . \tag{3.2.29}$$

As a reminder, the standard deviation $\sigma_{\hat{m}_k}$ is referred to, throughout this work, as the *uncertainty* corresponding to the parameter estimate $\hat{m}_k$, and can be formally expressed as follows:

$$\sigma_{\hat{m}_k} = \sqrt{\sum_{i=1}^{N} \left(x_i^{(k)}\right)^2 (\sigma_{d_i})^2}\,, \tag{3.2.30}$$

where the data errors $(\sigma_{d_i})_{1\leq i\leq N}$ are those from (3.2.22). As a second reminder, the 'uncertainty' $\sigma_{\hat{m}_k}$ merely represents the amount of noise that may propagate into the 'weighted-average' parameter estimate $\hat{m}_k$.

4. Once a continuous SOLA target kernel $T^{(k)}(\mathbf{r})$ has been defined, one needs to turn it into a 'discrete' vector $(T_j^{(k)})_{1\leq j\leq M}$, where $T_j^{(k)}$ represents the average of $T^{(k)}(\mathbf{r})$ in j-th voxel (of the model parameterization), that is[†]:

$$T_j^{(k)} = \frac{\int T^{(k)}(\mathbf{r}) h_j(\mathbf{r}) d^3\mathbf{r}}{\int h_j(\mathbf{r}) d^3\mathbf{r}} = \frac{\int_{\mathbf{r}\in j\text{-th voxel}} T^{(k)}(\mathbf{r}) d^3\mathbf{r}}{\mathbb{V}_j} = \frac{\int_{\mathbf{r}\in j\text{-th voxel}\,\cap\,\mathbb{S}^{(k)}} d^3\mathbf{r}}{\mathbb{V}_j \times \mathrm{vol}\{\mathbb{S}^{(k)}\}}, \tag{3.2.31}$$

see Fig. 3.2.15. Since $T^{(k)}(\mathbf{r})$ was normalized such that its integral is one, one should verify that the discrete SOLA target kernel satisfies to:

$$\sum_{j=1}^{M} \mathbb{V}_j T_j^{(k)} = 1\,. \tag{3.2.32}$$

[†] A simple Riemman sum is often efficient enough to numerically compute $\int_{\mathbf{r}\in j\text{-th voxel}\,\cap\,\mathbb{S}^{(k)}} d^3\mathbf{r}$. Reminder: $\mathbb{S}^{(k)}$ is the spatial domain (in model space) over which the continuous target kernel $T^{(k)}(\mathbf{r})$ is defined, and $\mathrm{vol}\{\mathbb{S}^{(k)}\}$ stands for its actual volume; see (3.2.11). Here, unit of $T_j^{(k)}$ is km^{-3}.

5. In the 'discrete' SOLA approach, one aims at minimizing the integrated squared difference between each *discrete* averaging kernel $(A_j^{(k)})_{1\leq j\leq M}$ and its corresponding *discrete* target kernel $(T_j^{(k)})_{1\leq j\leq M}$. A minimization problem is set up to directly compute each generalized-inverse vector $\mathbf{x}^{(k)} = (x_i^{(k)})_{1\leq i\leq N}$, for each query, k-th, voxel-like[††], parameter, that is:

$$\begin{cases} \underset{\mathbf{x}^{(k)}\in\mathbb{R}^N}{\arg\min}: & \underbrace{\sum_{j=1}^{M} \mathbb{V}_j \left[A_j^{(k)} - T_j^{(k)}\right]^2}_{\substack{\text{RESOLUTION MISFIT}\\ \text{Favoring averaging- close to target-kernel}}} + \underbrace{[\eta^{(k)}]^2}_{\substack{\text{Trade}\\ \text{off}}} \underbrace{\sigma^2_{\hat{m}_k}}_{\substack{\text{VARIANCE of } \hat{m}_k\\ \text{Moderating uncertainty}}} \\ \text{subject to} & \underbrace{\sum_{j=1}^{M} \mathbb{V}_j A_j^{(k)} = 1\,.}_{\substack{\text{AVERAGING}\\ \text{CONDITION}}} \end{cases} \tag{3.2.33}$$

[††] In practice, one may want the k-th voxel to be 'centered' on the corresponding query-point location $\mathbf{r}^{(k)}$ (which was used to define the *continuous* target kernel); see Fig. 3.2.15(d).

The previous constrained minimization problem (3.2.33) can be written:

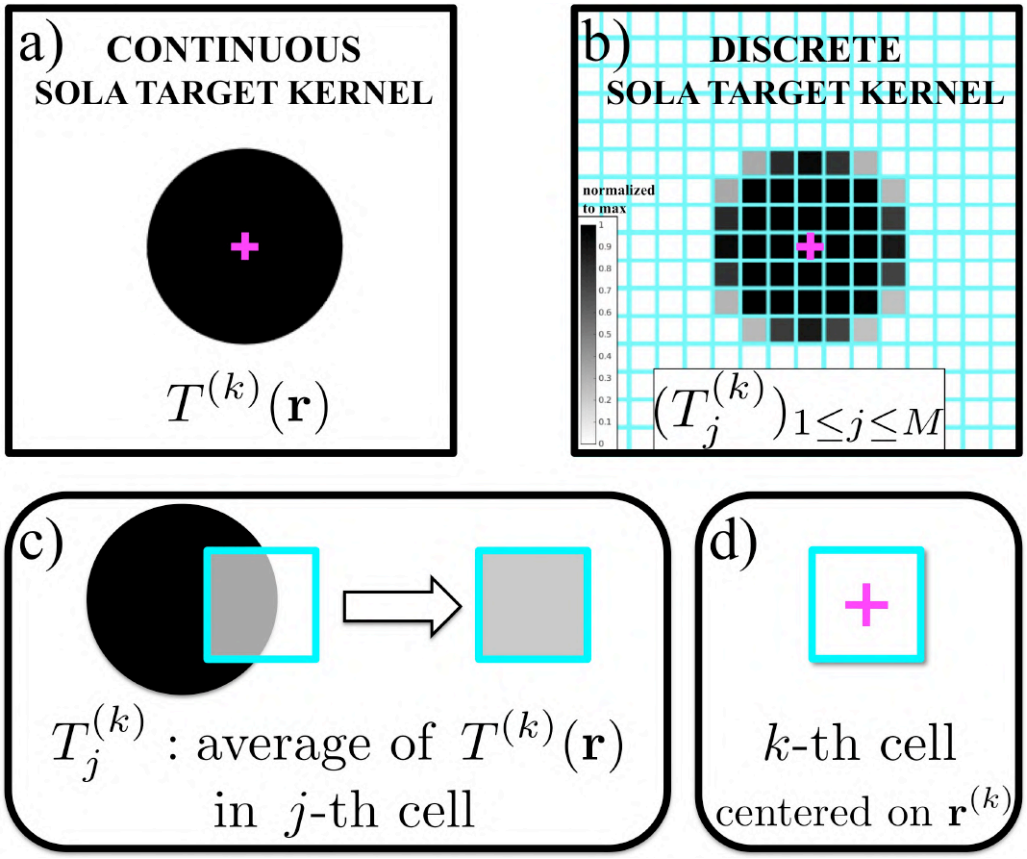


Figure 3.2.15: (a) The black disk represents a continuous SOLA target kernel, $T^{(k)}(\mathbf{r})$, centered on a query-point location, $\mathbf{r}^{(k)}$ (magenta cross). (b) The cyan grid represents a local parameterization of model space. Here, cell's area, $\mathbb{V}_j$, is identical for all $M = 16 \times 16$ cells. The discrete target kernel, $(T_j^{(k)})_{1\leq j\leq M}$, is displayed using a white-to-black color scale (normalized to maximum). As a reminder: $\int T^{(k)}(\mathbf{r}) = \sum_j \mathbb{V}_j T_j^{(k)} = 1$. (c) Illustration showing that $T_j^{(k)}$ is the average of $T^{(k)}(\mathbf{r})$ in the j-th cell; see (3.2.31). (d) In practice, one may want the k-th cell to be centered on $\mathbf{r}^{(k)}$.

$$
\left\{
\begin{array}{l}
\text{solve} \quad \begin{pmatrix} \hat{\mathbf{G}}^T \\ \eta^{(k)} \underset{1\leq i\leq N}{\text{diag}}\,(\sigma_{d_i}) \end{pmatrix} \mathbf{x}^{(k)} = \begin{pmatrix} \mathbf{t}^{(k)} \\ \mathbf{0}_N \end{pmatrix}, \\
\\
\text{subject to} \quad \sum_{i=1}^N x_i^{(k)} \left(\sum_{j=1}^M G_{ij} \right) = 1 \,,
\end{array}
\right. \tag{3.2.34}
$$

$$
\text{with:} \left\{
\begin{array}{lll}
\hat{\mathbf{G}} = \left(\hat{G}_{ij}\right)_{1\leq i,j\leq N,M} & , & \hat{G}_{ij} = \frac{G_{ij}}{\sqrt{\mathbb{V}_j}} = K_{ij}\sqrt{\mathbb{V}_j} \\
\\
\mathbf{t}^{(k)} = \left(t_j^{(k)}\right)_{1\leq j\leq M} & , & t_j^{(k)} = T_j^{(k)}\sqrt{\mathbb{V}_j}\,.
\end{array}
\right.
$$

Derivation from (3.2.33) to (3.2.34)[†] is detailed in Text Box 3.5; (3.2.34) can be numerically solved using an LSQR-based approach (Zaroli, 2016).

[†] $(\cdot)^T$ is the *transpose* operator.

Text Box 3.5: DISCRETE SOLA NORMAL EQUATIONS

Expanding out $A_j^{(k)}$ and $\sigma_{\hat{m}_k}$ in function of $\mathbf{x}^{(k)}$, the system (3.2.33) writes:

$$\begin{cases} \underset{\mathbf{x}^{(k)}\in\mathbb{R}^N}{\arg\min:} & \sum_{j=1}^{M} \mathbb{V}_j \left[\sum_{i=1}^{N} x_i^{(k)} \left(G_{ij}\mathbb{V}_j^{-1}\right) - T_j^{(k)}\right]^2 + [\eta^{(k)}]^2 \sum_{i=1}^{N} \left(x_i^{(k)}\right)^2 (\sigma_{d_i})^2 \\ \text{subject to} & \sum_{i=1}^{N} x_i^{(k)} \left(\sum_{j=1}^{M} G_{ij}\right) = 1\,. \end{cases}$$

To obtain the normal equations corresponding to the 'unconstrained part' of the previous minimization problem, we do the following (i' is fixed):

$$\frac{\partial}{\partial x_{i'}^{(k)}} \left(\sum_{j=1}^{M} \mathbb{V}_j \left[\sum_{i=1}^{N} x_i^{(k)} \left(G_{ij}\mathbb{V}_j^{-1}\right) - T_j^{(k)} \right]^2 + [\eta^{(k)}]^2 \sum_{i=1}^{N} \left(x_i^{(k)}\right)^2 (\sigma_{d_i})^2 \right) = 0\,,$$

$$\Longrightarrow \sum_{i=1}^{N} x_i^{(k)} \sum_{j=1}^{M} \underbrace{\left(\frac{G_{i'j}}{\sqrt{\mathbb{V}_j}}\right)}_{\hat{G}_{i'j}} \underbrace{\left(\frac{G_{ij}}{\sqrt{\mathbb{V}_j}}\right)}_{\hat{G}_{ij}} + x_{i'}^{(k)} [\eta^{(k)}]^2 (\sigma_{d_{i'}})^2 = \sum_{j=1}^{M} \underbrace{\left(\frac{G_{i'j}}{\sqrt{\mathbb{V}_j}}\right)}_{\hat{G}_{i'j}} \underbrace{\left(\sqrt{\mathbb{V}_j} T_j^{(k)}\right)}_{t_j^{(k)}}\,.$$

Considering that $1 \leq i' \leq N$, the previous equation can be expressed as:

$$\text{solve}: \quad \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^T + [\eta^{(k)}]^2 \underset{1\leq i\leq N}{\text{diag}} \left(\sigma_{d_i}^2\right)\right) \mathbf{x}^{(k)} = \hat{\mathbf{G}}\,\mathbf{t}^{(k)}\,,$$

which is equivalent (in a 'least-squares' sense) to the following system:

$$\text{solve}: \quad \begin{pmatrix} \hat{\mathbf{G}}^T \\ \eta^{(k)} \underset{1\leq i\leq N}{\text{diag}} \left(\sigma_{d_i}\right) \end{pmatrix} \mathbf{x}^{(k)} = \begin{pmatrix} \mathbf{t}^{(k)} \\ \mathbf{0}_N \end{pmatrix} \quad \text{– as in (3.2.34).} \ \square$$

6. In terms of computational cost, it is (much) cheaper to set up a 'discrete' SOLA system (3.2.34), compared to a 'continuous' system (3.2.14):
   (a) (3.2.14) mainly requires to compute $N^2/2$ volumetric integrals $\int K_i K_{i'}$, where $(K_i, K_{i'})$ is a pair of *continuous* sensitivity kernels;
   (b) (3.2.34) mainly requires to compute N *discrete* sensitivity kernels $(K_{ij})_{1\leq j\leq M}$, i.e. to project N kernels K_i on a model parameterization.
7. To get started with discrete SOLA tomography, all what is needed for is:
   (a) Computing[†] a sensitivity matrix, $\mathbf{G}$;
   (b) Computing[††] P discrete SOLA target kernels, $(T_j^{(k)})_{1\leq j\leq M}$.

   Therefore, since all the classical (discrete, linear, DLS-based) tomographic studies are routinely computing sensitivity matrices, see Sect. 3.3, it shall be 'straightforward' to move on toward using the presented *discrete* SOLA recipe (computing discrete SOLA target kernels is also a 'trivial' task).

[†] $\mathbf{G}$ is often sparse and can be computed and stored for large data sets.

[††] For some applications, it is possible to consider a *selection* of parameters ($P \leq M$), for which to solve (3.2.34).

- **'Toy' problem**. In the following, I aim to describe – in a pedagogical way – a *synthetic*, 2–D, 'discrete SOLA' tomographic experiment (Zaroli et al., 2017).

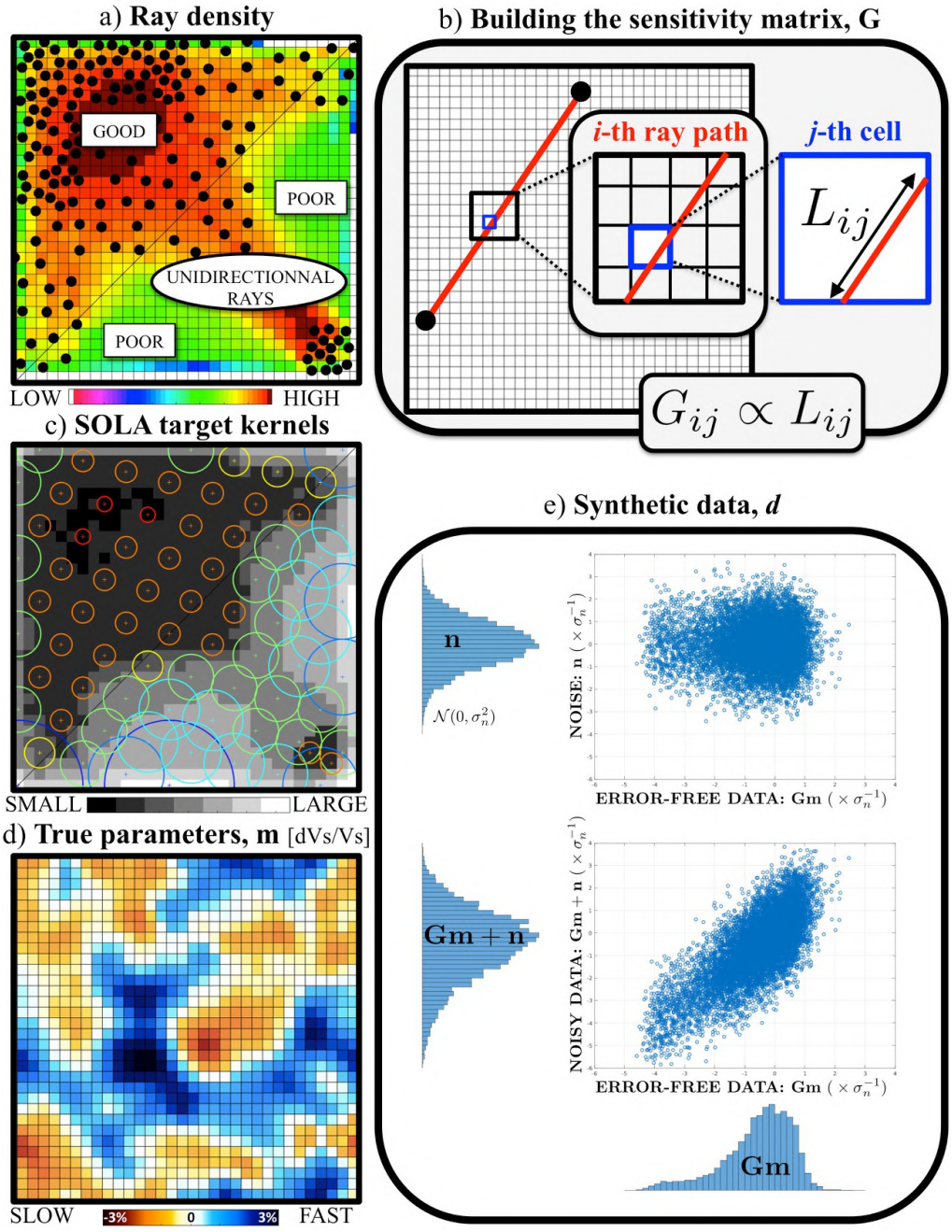


Figure 3.2.16: (a) Data coverage and parametrization of model space ($M = 32 \times 32$ cells). (b) G_{ij} is proportional to L_{ij} (ray theory). (c) Circular forms of some target kernels, $T^{(k)}$. (d) 'True' model, $\mathbf{m}$ (dV_S/V_S). (d) Synthetic data, $\mathbf{d} \leftarrow \mathbf{Gm} + \mathbf{n}$. Modified from Zaroli et al. (2017).

1. Tomographic settings are displayed in Fig. 3.2.16. Here the model parameters are shear wave velocity perturbations (dV_S/V_S), and the parametrization consists of $M = 1024$ square pixels (cells) of unit area each. In the framework of ray theory, the $N = 9778$ data[†] represent onset time residuals of direct S waves, whose the ray-paths are straight lines from one black dot to another. All pairs of black dots with an inter-distance larger than 8 pixel units are 'suitable' ray-paths. The data covariance matrix[††] is: $\mathbf{C}_d = \mathrm{diag}(\sigma_{d_i}^2)_{1\leq i\leq N} = \sigma_n^2 \mathbf{I}_N$. The SOLA inversion parameters are tuned according to the aforementioned *recipe*: 1) ray-density driven, circular-shape target kernels[†††], and 2) constant-valued trade-off parameter, η.
2. In the framework of discrete SOLA tomography, one can also compute in a *parallel* fashion (i.e., mutually independent) P generalized-inverse vectors $\mathbf{x}^{(k)}$, with $P \leq M$, as schematically illustrated in Fig. 3.2.17.

[†] For given true-model and noise vectors, *synthetic* data are computed as: $\mathbf{d} \leftarrow \mathbf{Gm} + \mathbf{n}$.

[††] σ_n stands for the standard deviation of the normally distributed, zero-mean distribution of noise, $\mathbf{n}$.

[†††] Each target kernel $T^{(k)}$ is centered on the corresponding query, k-th cell.

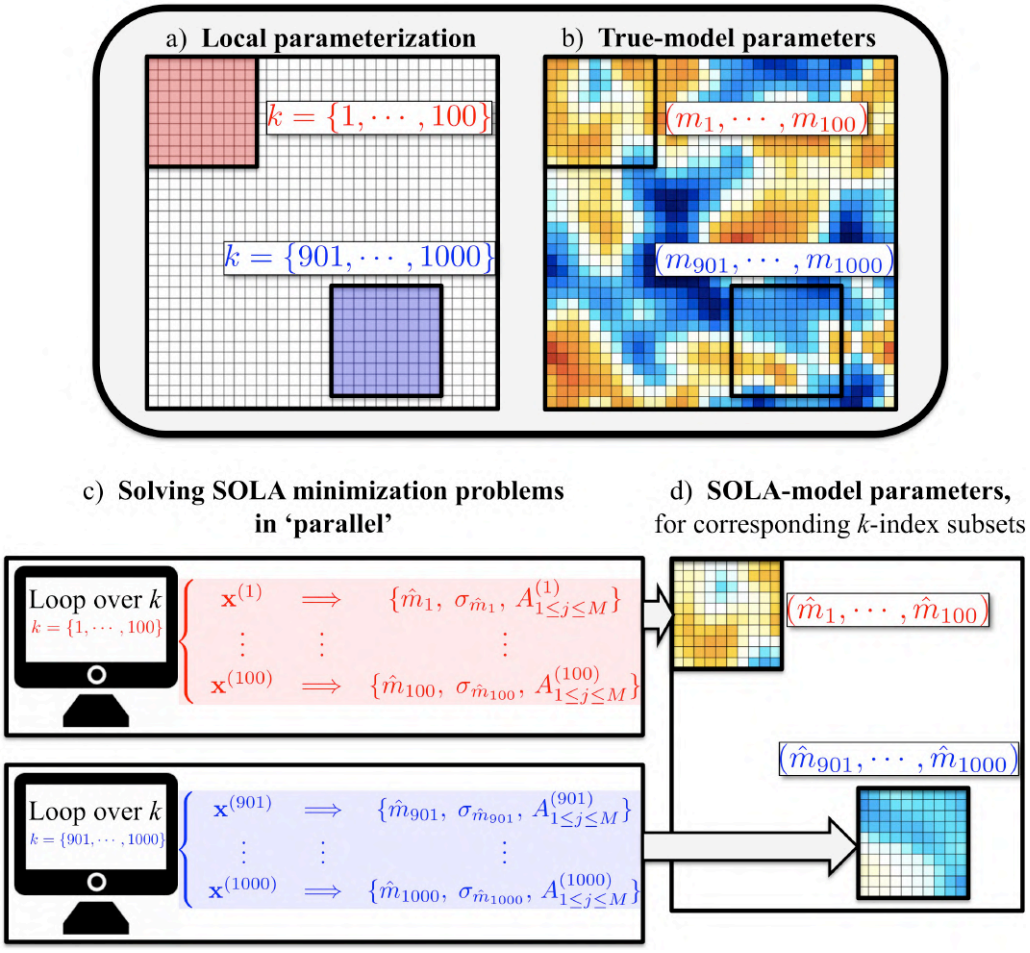


Figure 3.2.17: Multiple discrete-SOLA minimization problems (3.2.34) can be solved in *parallel*, by considering several – in fact, as many as the number of available 'computers' – subsets of cells.

3. Note that, in the discrete SOLA (Backus–Gilbert) inversion approach, there is always the possibility to focus on a restricted set of say $P \leq M$ query points, for which the generalized-inverse vectors $(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(P)})$ are effectively computed. In particular, one may want to consider *all* the M cells as 'query cells' – that is, $P = M$. In such a scenario, it will prove useful to introduce the following mathematical notations:

$\mathbf{d} = (d_i)_{1 \leq i \leq N}$	data vector
$\mathbf{m} = (m_j)_{1 \leq j \leq M}$	true-model vector
$\mathbf{G} = (G_{ij})_{1 \leq i,j \leq N,M}$	sensitivity matrix
$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{n}$	discrete tomographic problem
$\hat{\mathbf{m}} = (\hat{m}_j)_{1 \leq j \leq M}$	model-estimate vector
$\mathbf{G}^\dagger = \begin{pmatrix} x_1^{(1)} & \cdots & x_i^{(1)} & \cdots & x_N^{(1)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(k)} & \cdots & x_i^{(k)} & \cdots & x_N^{(k)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(M)} & \cdots & x_i^{(M)} & \cdots & x_N^{(M)} \end{pmatrix}$	generalized-inverse matrix
$\underbrace{\hat{\mathbf{m}}}_{\substack{\text{MODEL} \\ \text{ESTIMATE}}} = \mathbf{G}^\dagger \mathbf{d} = \underbrace{\mathbf{R}\mathbf{m}}_{\substack{\text{FILTERED} \\ \text{TRUE MODEL}}} + \underbrace{\mathbf{G}^\dagger \mathbf{n}}_{\substack{\text{PROPAGATED} \\ \text{NOISE}}}$	model estimate splits up into two terms
$\mathbf{R} = \mathbf{G}^\dagger \mathbf{G} = (R_{kj})_{1 \leq k,j \leq M}$	model-estimate 'resolution matrix'
$\mathbf{C}_{\hat{m}} = \mathbf{G}^\dagger \mathbf{C}_d (\mathbf{G}^\dagger)^T = \begin{pmatrix} \sigma^2_{\hat{m}_1} & \cdots & \text{off-diag.} \\ \vdots & \ddots & \vdots \\ \text{off-diag.} & \cdots & \sigma^2_{\hat{m}_M} \end{pmatrix}$	model-estimate covariance matrix
$\hat{\mathbf{d}} = \mathbf{G}\hat{\mathbf{m}}$	data vector 'predicted' by model estimate
$\chi^2_{\text{red}} = \begin{cases} \frac{1}{N} \left(\hat{\mathbf{d}} - \mathbf{d}\right)^T \mathbf{C}_d^{-1} \left(\hat{\mathbf{d}} - \mathbf{d}\right) \\ \frac{1}{N} \sum_{i=1}^N \left[\frac{\sum_{j=1}^M G_{ij} \hat{m}_j - d_i}{\sigma_{d_i}} \right]^2 \end{cases}$	*reduced-chi-square* measure of data misfit.

(3.2.35)

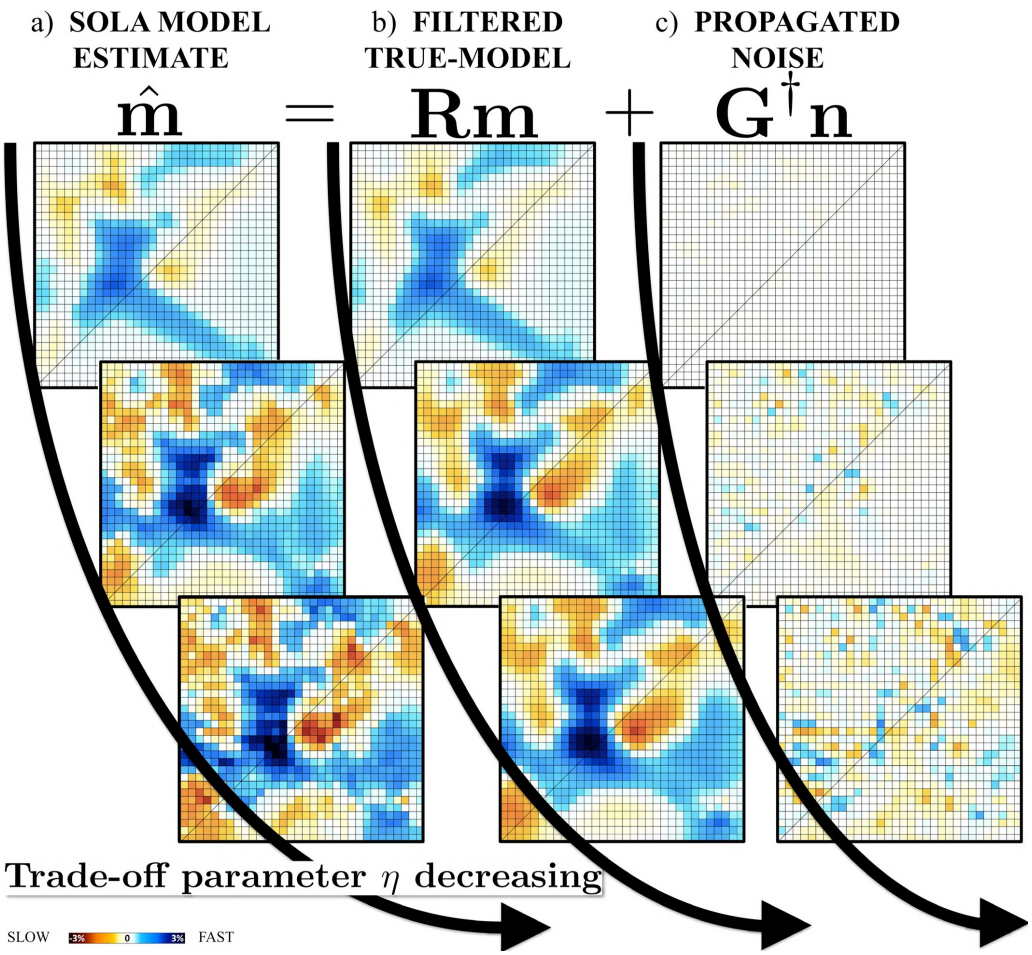


Figure 3.2.18: (a) Discrete-SOLA tomographic models, $\hat{\mathbf{m}}$, obtained for different values of trade-off parameter η (from top to bottom: η^{high}, η^{middle}, η^{low}, respectively, with $\eta^{\text{high}} = \eta^{\text{middle}} \times 6.0$, and $\eta^{\text{low}} = \eta^{\text{middle}} \times 0.3$). The model estimate, $\hat{\mathbf{m}}$, can be decomposed into the sum of two terms: (b) the filtered true-model, $\mathbf{Rm}$, and (c) the propagated noise, $\mathbf{G}^{\dagger}\mathbf{n}$. Note that the *reduced-chi-square* measure of data misfit, χ^2_{red}, is: 1.174, 1.008 and 1.006, for the discrete-SOLA model estimates, $\hat{\mathbf{m}}$, obtained for η^{high}, η^{middle} and η^{low}, respectively. As a remark, here, one sees that the lower η, the lower χ^2_{red}. Modified from Zaroli et al. (2017).

4. The term $\mathbf{Rm}$ is commonly referred to as the 'tomographically filtered' version of the true model $\mathbf{m}$ (e.g., Ritsema et al., 2007; Schuberth et al., 2009a; Koelemeijer et al., 2017; Freissler et al., 2020). This filtering provides valuable information on how heterogeneities are smeared and modified in amplitude given the available seismic data and underlying inversion strategy. The SOLA approach explicitly aims at building a resolution[†] matrix, $\mathbf{R}$, such that $\mathbf{Rm}$ represents an *unbiased* averaging process, see (3.2.24). Fig. 3.2.18 shows, for the 'toy' problem, the SOLA model estimate, $\hat{\mathbf{m}}$, see Fig. 3.2.18(a), which can be viewed as the sum of two terms:

[†] From the definition of $\mathbf{R}$, it is clear that the k-th row of $\mathbf{R}$ is related to the (discrete) averaging kernel $A^{(k)}$ as stated in (3.2.28). Averaging kernels are then also referred to, in the literature, as *resolving* kernels.

1) the filtered true-model, $\mathbf{Rm}$, see Fig. 3.2.18(b), and 2) the propagated noise, $\mathbf{G}^{\dagger}\mathbf{n}$, see Fig. 3.2.18(c). Results are displayed for three different values of trade-off parameter, η. As expected, from the SOLA minimization problem (3.2.33), the lower the trade-off parameter: 1) the closer the filtered true-model to the true model (indeed, the closer the averaging kernels to the target kernels, the better the local resolution), and 2) the more the amount of noise propagating into the model estimate.

5. The reduced-chi-square χ^2_{red} is an essential measure of the statistical goodness of data fit, for a given model estimate. In the hypothetical case that, for every datum, we have $(\hat{d}_i - d_i)^2/\sigma^2_{d_i} \approx 1$ (every datum is satisfied with a misfit of one standard deviation), then we expect that: $\chi^2_{\text{red}} \approx \left(\frac{1}{N} \times \sum_{i=1}^{N} 1\right) = 1$. Note that the way the (discrete) SOLA generalized-inverse matrix $\mathbf{G}^{\dagger}$ is computed, see (3.2.34), the SOLA model estimate $\hat{\mathbf{m}}$ could, *a priori*[†], not be expected to fit the data. However, when following the 'recipe' for tuning the SOLA target kernels and trade-off parameters, and inverting for *all* M parameters, I have demonstrated that one can build a 'globally coherent' SOLA model-estimate, $\hat{\mathbf{m}}$, that does fit the data – for the 'toy' problem, χ^2_{red} is close to one, see Fig. 3.2.18.

[†] Contrary to least-squares methods, which explicitly seek a set of model parameters such that it minimizes the data misfit, see Sect. 3.3.1.

6. Interpreting a tomographic image (model appraisal) is not an easy task. As sketched in Fig. 3.2.19, the conventional way to display a tomographic image may be somehow a bit confusing. For example, here, each 'pixel' of the image represents the value of a *weighted-average* parameter estimate, $\hat{m}_k$, with a corresponding averaging kernel, $A^{(k)}$, and uncertainty (on the weighted average), $\sigma_{\hat{m}_k}$. A major difficulty lies in the fact that: 1) the spatial extent of averaging kernels may change rapidly, from one pixel to another, and 2) the shape itself of each averaging kernel may not be well 'localized' – so that, in the end, it may be uneasy to infer *relevant*[††], average properties on the 'true' model. Therefore, when the data coverage is spatially irregular and/or anisotropic (e.g., unidirectional rays as in Fig. 3.2.16(a)), one may have no choice but to visually inspect every averaging kernels. Whenever there are too many $A^{(k)}$'s, an alternative to visualize the main properties of the underlying local resolution, may be to design several *input* 'true' models, $\mathbf{m}$, and see how they would look like through the tomographic 'lenses', that is, to analyze[†††] the *output* filtered models, $\mathbf{Rm}$. Examples are shown in Fig. 3.2.20(a,c), while considering different lenses (trade-off parameters, η) and various input 'true' models.

[††] In the sense: '*easy to apprehend for human brain*'.

[†††] Since analysing $\mathbf{Rm}$, for a given input model $\mathbf{m}$, obviously depends on the selected model $\mathbf{m}$, this cannot replace a careful analysis of $\mathbf{R}$ itself.

7. From the very formulation of the SOLA minimization problem, it is clear that, as the trade-off parameter η goes toward zero, the averaging kernel $A^{(k)}$ is expected to get closer to the target kernel $T^{(k)}$ (and the uncertainty $\sigma_{\hat{m}_k}$ to get higher). It may then be interesting to focus on the 'ideal' case for which one would have $A^{(k)} = T^{(k)}$, that is, to visualize, for a given *input* 'true' model, $\mathbf{m}$, the *output* true model 'filtered through the target kernels', $\mathbf{Tm}$, to be compared to the *output* true model 'filtered through the averaging kernels', $\mathbf{Rm}$, where the matrix $\mathbf{T}$ is defined as follows:

$$\mathbf{T} = (T_{kj})_{1\leq k,j\leq M} \ , \qquad T_{kj} = \mathbb{V}_j T_j^{(k)} \ , \tag{3.2.36}$$

with $(T_j^{(k)})_{1\leq j\leq M}$ the components of the discrete target kernel $T^{(k)}$, corresponding to the query k-th cell. Examples of the term **Tm** are provided in Fig. 3.2.20(b), while considering three different input 'true' models.

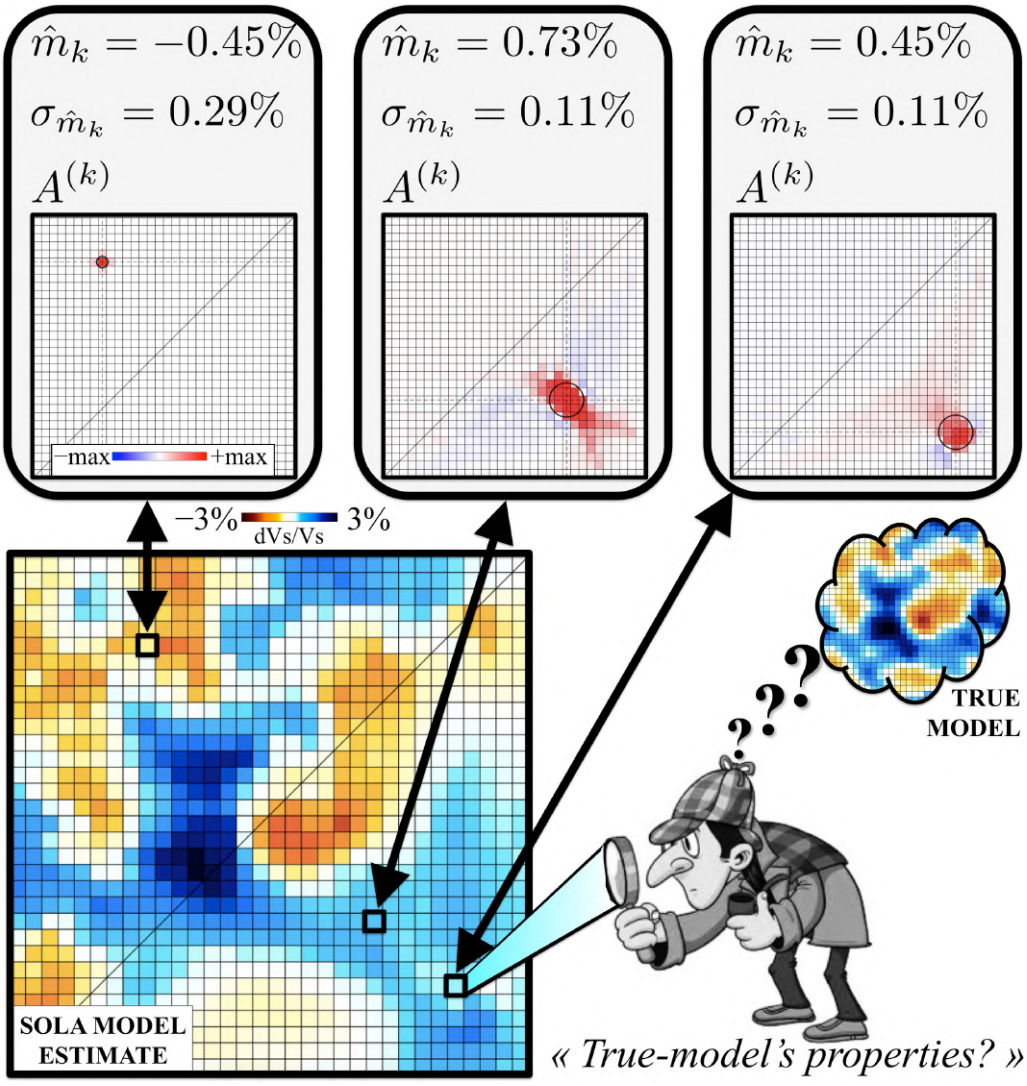


Figure 3.2.19: The SOLA approach provides a wealth of informations (weighted averages, $\hat{m}_k$, uncertainties, $\sigma_{\hat{m}_k}$, averaging kernels, $A^{(k)}$, that one needs to carefully analyze for making robust inferences on *some* true-model's properties. Model appraisal is not an easy task, in particular when the underlying local resolution (and uncertainty) differs from one pixel to another. One 'Sherlock Holmes' may then not be enough to appraise many features. Modified from Zaroli et al. (2017).

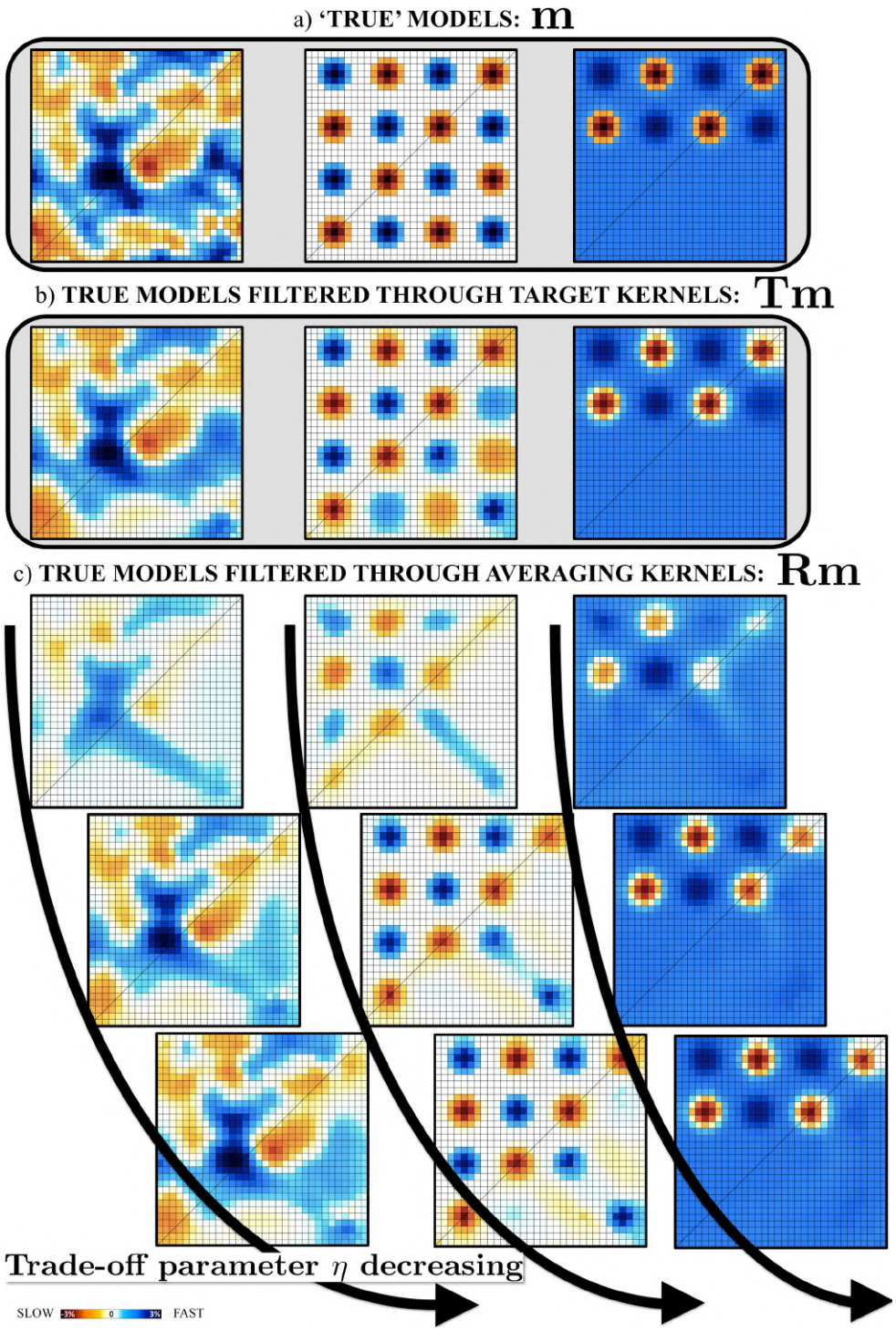


Figure 3.2.20: (a) Input 'true' models, **m**, filtered through (b) the target kernels, **Tm**, and (c) the averaging kernels, **Rm**, for η^{high}, η^{middle}, η^{low}, respectively. Modified from Zaroli et al. (2017).

8. The SOLA-model uncertainties, corresponding to local-average estimates[†], also play a key role in the appraisal of tomographic images. In the SOLA approach, the uncertainty $\sigma_{\hat{m}_k}$ is directly calculated from the generalized-inverse coefficients $\mathbf{x}^{(k)}$, see (3.2.30). Taking Fig. 3.2.21 as an illustrative example, I now aim at discussing several points about these uncertainties.

   (a) First, all the M uncertainties are displayed as 'conventional images' in Figs. 3.2.21(i) and (j), corresponding to high and low trade-off parameter (η) values, η^{high} and η^{low}, respectively. They are in agreement with the 'propagated-noise' images shown in Figs. 3.2.18(c, top) and (c, bottom), respectively. Clearly, some patterns in the uncertainty images can be identified and related to the previously chosen target kernels, see Fig. 3.2.16(b). However, it is important to realize that one cannot (always) 'guess' what uncertainty $\sigma_{\hat{m}_k}$ to expect for a given, query k-th cell, since it depends on several factors[††], such as:

      i. The size of target kernel $T^{(k)}$ (here, driven by ray density).

      ii. The value of trade-off parameter $\eta^{(k)}$ (here, $\eta^{(k)} = \eta$).

      iii. The spatial extent of averaging kernel $A^{(k)}$, which itself depends on $T^{(k)}$ and $\eta^{(k)}$, but also on the *actual* ray coverage. That is, when ray paths are locally anisotropic, ray density is a poor proxy for local resolution; $A^{(k)}$ is then expected to differ a lot from $T^{(k)}$.

      iv. The number of ray-paths which are effectively influenced by the spatial domain of $A^{(k)}$. Statistically, the more the number of rays (influenced by $A^{(k)}$), the more the amount of propagated noise, $(\mathbf{x}^{(k)})^T\mathbf{n}$, tends to cancel out[†††] – hence the lower $\sigma_{\hat{m}_k}$.

   (b) Second, without loss of generality, the variance of the estimate $\hat{m}_k$ writes: $\sigma^2_{\hat{m}_k} = \sum_{i=1}^{N}(x_i^{(k)})^2 \times \sigma_n^2$. Consequently, the higher $(x_i^{(k)})^2$, the higher the 'contribution' of i-th ray-path to the uncertainty $\sigma_{\hat{m}_k}$ (and, somehow[††††], to $A^{(k)}$ and $\hat{m}_k$). For example, Fig. 3.2.21(a) shows an averaging kernel, corresponding to some 'query' k-th cell and η^{high}; the attached uncertainty is 0.02%, see Fig. 3.2.21(i); in the following (main text), they are referred to as $A^{(k,\eta^{\text{high}})}$ and $\sigma_{\hat{m}_{k,\eta^{\text{high}}}}$. Fig. 3.2.21(b) shows the corresponding N ray-paths, blue-colored with a level of transparency inversely proportional, for the i-th ray-path, to the i-th, squared generalized-inverse coefficient. This allows one to identify *which* and *how many* rays are significantly contributing to $\sigma_{\hat{m}_{k,\eta^{\text{high}}}}$. Let's focus now on a lower trade-off parameter value, η^{low}. Figs. 3.2.21(c) and (j) show the corresponding averaging kernel and uncertainty (0.37%); in the following, they are referred to as $A^{(k,\eta^{\text{low}})}$ and $\sigma_{\hat{m}_{k,\eta^{\text{low}}}}$; Fig. 3.2.21(d) shows the corresponding ray-paths (blue-colored). As expected, from the trade-off between resolution (i.e. size of averaging kernel) and uncertainty, one sees that:

      i. The size (i.e. spatial extent) of $A^{(k,\eta^{\text{low}})}$ is smaller than for $A^{(k,\eta^{\text{high}})}$.

      ii. There are more rays (those with a low level of blue-color trans-

[†] Reminder: In the SOLA approach, the uncertainty $\sigma_{\hat{m}_k}$ cannot inform us on how much the (local-average) estimate $\hat{m}_k$ may differ from the 'true' parameter m_k, at least when the spatial variations of the 'true' model $\mathbf{m}$ are non smooth and/or the spatial extent of the averaging kernel $A^{(k)}$ is not strictly restricted to the k-th cell.

[††] Therefore, there is no choice but to 'compute' the uncertainties, through the computation of some generalized-inverse operator and the estimation of data errors.

[†††] Assuming zero-mean, normally distributed noise $\mathbf{n}$.

[††††] When computing $A^{(k)}$ and $\hat{m}_k$, the elements of $\mathbf{x}^{(k)}$ are *not* squared: their sign matters (unlike for $\sigma_{\hat{m}_k}$).

parency) significantly contributing to $\sigma_{\hat{m}_{k,\eta^{\text{high}}}}$ than to $\sigma_{\hat{m}_{k,\eta^{\text{low}}}}$.

iii. The uncertainty $\sigma_{\hat{m}_{k,\eta^{\text{low}}}}$ is higher than $\sigma_{\hat{m}_{k,\eta^{\text{high}}}}$, which is in agreement with the two previous points.

Summary. The lower the trade-off parameter η, the closer the averaging kernel $A^{(k)}$ to the target kernel $T^{(k)}$ (i.e. the better the local resolution), the fewer the number of 'ray-paths' (that is, data, data-sensitivity kernels, data errors) *significantly contributing* to the solution ($\hat{m}_k$, $A^{(k)}$, $\sigma_{\hat{m}_k}$), and the higher the uncertainty $\sigma_{\hat{m}_k}$.

(c) Third, Fig. 3.2.21(e) shows an averaging kernel for *another query cell*, i.e. the k'-th, and η^{high}; the uncertainty is 0.10%, see Fig. 3.2.21(i); here, they are referred to as $A^{(k',\eta^{\text{high}})}$ and $\sigma_{\hat{m}_{k',\eta^{\text{high}}}}$. Figs. 3.2.21(g) and (j) show, for η^{low}, the averaging kernel and uncertainty (0.32%), here referred to as $A^{(k',\eta^{\text{low}})}$ and $\sigma_{\hat{m}_{k',\eta^{\text{low}}}}$. Figs. 3.2.21(f) and (h) show the blue-colored ray-paths. As expected, one also sees that:

i. The spatial extent of $A^{(k',\eta^{\text{low}})}$ is slightly smaller than for $A^{(k',\eta^{\text{high}})}$.

ii. More rays are significantly contributing to $\sigma_{\hat{m}_{k',\eta^{\text{high}}}}$ than to $\sigma_{\hat{m}_{k',\eta^{\text{low}}}}$.

iii. The uncertainty $\sigma_{\hat{m}_{k',\eta^{\text{low}}}}$ is higher than $\sigma_{\hat{m}_{k',\eta^{\text{high}}}}$.

Furthermore, it may be interesting to compare the averaging kernels and uncertainties for the two previously considered query cells, i.e. the k- and k'-th. For the trade-off parameter η^{high}, one sees that:

i. The spatial extent of $A^{(k',\eta^{\text{high}})}$ is much smaller than for $A^{(k,\eta^{\text{high}})}$, that is, the local resolution is much better at the k'-th cell.

ii. The uncertainty $\sigma_{\hat{m}_{k',\eta^{\text{high}}}}$ (0.10%) is much higher than $\sigma_{\hat{m}_{k,\eta^{\text{high}}}}$ (0.02%), so there must be more rays significantly contributing to $\sigma_{\hat{m}_{k,\eta^{\text{high}}}}$ than to $\sigma_{\hat{m}_{k',\eta^{\text{high}}}}$ (though uneasy to see in Figs. 3.2.21(b,f)).

For a lower trade-off parameter, η^{low}, one sees that:

i. The spatial extent of $A^{(k',\eta^{\text{low}})}$ is smaller than for $A^{(k,\eta^{\text{low}})}$, that is, the local resolution is better at the k'-th cell.

ii. The uncertainty $\sigma_{\hat{m}_{k',\eta^{\text{low}}}}$ (0.32%) is now slightly 'lower' than $\sigma_{\hat{m}_{k,\eta^{\text{low}}}}$ (0.37%), thus indicating that there must be slightly 'less' rays significantly contributing to $\sigma_{\hat{m}_{k,\eta^{\text{low}}}}$, compared to $\sigma_{\hat{m}_{k',\eta^{\text{low}}}}$ (though uneasy to see in Figs. 3.2.21(d) and (h)).

Summary. When comparing two *far-apart* query cells, for example the k- and k'-th cells, if the local resolution is say better at the k'-th cell ($A^{(k')}$'s spatial extent smaller than $A^{(k)}$'s), then the uncertainty $\sigma_{\hat{m}_{k'}}$ can be higher *or* lower than $\sigma_{\hat{m}_k}$ – depending on the number of rays 'significantly-contributing' to each uncertainty.

Finally, in order to take uncertainties into account for appraising a model solution, there is no choice but to calculate them – thus requiring to compute a generalized-inverse operator (robust data errors are also needed).

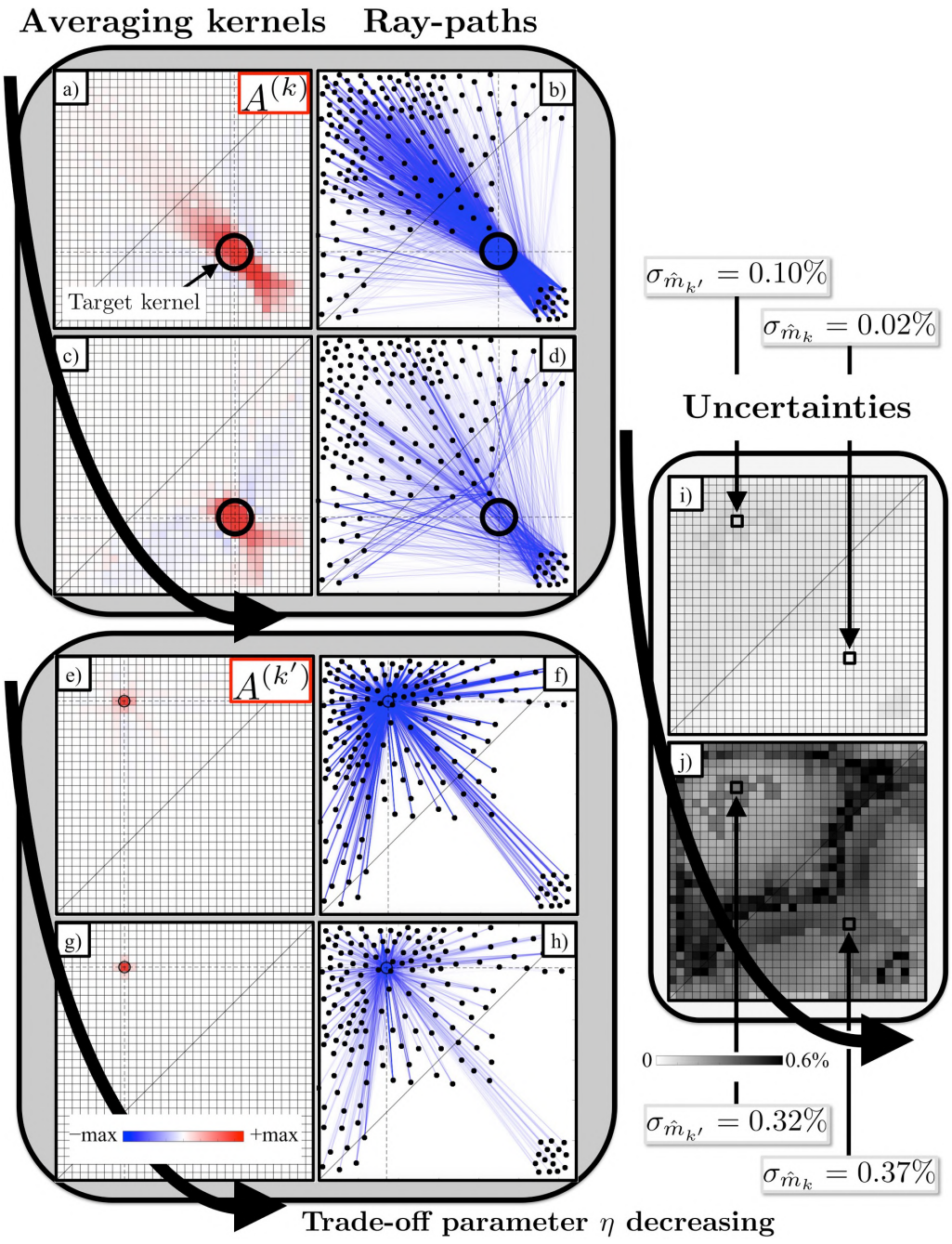


Figure 3.2.21: (a) Example of SOLA averaging kernel, $A^{(k)}$. (b) Corresponding ray-paths shown with blue-color transparency, inversely proportional, for the i-th ray path, to the i-th squared generalized-inverse coefficient, $(x_i^{(k)})^2$. This allows one to visualize the rays which contribute the most to the corresponding uncertainty $\sigma_{\hat{m}_k}$ (and, somehow, to the averaging kernel $A^{(k)}$ and the estimate $\hat{m}_k$). (c, d) Similar to (a, b), for lower η. (e–h) Similar to (a–d), for another 'query' cell (k'). (i) Uncertainties (corresponding to local averages). (j) Similar to (i), for lower η. Remark: $\eta = \eta^{\text{high}}$ for (a, b, e, f, i), and $\eta = \eta^{\text{low}}$ for (c, d, g, h, j). Modified from Zaroli et al. (2017).

- **Example**. To illustrate that the SOLA–Backus–Gilbert inversion approach can successfully be applied to large-scale discrete (linear) tomographic problems, in particular when dealing with 'real' data, I aim to present some results of a global-scale mantle tomography study (Zaroli, 2016).

  1. *Tomographic settings.* The data set consists in $N = 79\,765$, globally distributed, *S*- and *SS*-wave cross-correlation time-residuals, measured at 22 s central period (Zaroli et al., 2010). The data-sensitivity kernels, K_i, are calculated in a finite-frequency physical framework. An irregular, data-driven parameterization of the model space (Earth's mantle) is used. The mantle is radially divided into 18 spherical layers, see Fig. 3.2.22(a), which are spanned laterally with spherical triangles (Delaunay triangulations), see Fig. 3.2.22(b), whose the spatial distributions are optimized according to ray density. Let us denote as *nodes* all the vertices of every triangles, located (in depth) at the 'middle depth' of their corresponding layer, see Fig. 3.2.22(c). The tomographic grid[†] consists in $M = 38\,215$ model parameters (i.e., nodes). The physical parameter of interest is the isotropic 3–D variation of shear velocity ($dV_S/V_S = d \ln V_S$), in the mantle, with respect to radial V_S model iasp91 (Kennett & Engdahl, 1991). It is straightforward to compute the sensitivity matrix $\mathbf{G}$ (Zaroli, 2010).
  2. *SOLA inversions.* Each target kernel, $T^{(k)}$, is a spheroid centred on a 'query' node ($\mathbf{r}^{(k)}$) of the tomographic grid. Focus is on the 400–1710 km depth range, where our data coverage is the most relevant. The size of $T^{(k)}$ is specified according to the local size, both radially and laterally, of the data-driven parametrization, see Figs. 3.2.23(a,f). Target kernels are *discretized* (projected onto the parametrization), as required to set up the discrete SOLA problems (3.2.34). A constant value is selected for the trade-off parameter, $\eta^{(k)} = \text{cst}$, for every nodes within the transition zone (400–1710 km depth), and another value (slightly different) for those in the mid lower mantle (660–1710 km depth). Fig. 3.2.23(b) displays, as a 'conventional' image, the discrete-SOLA model estimates, $\hat{m}_k$, corresponding to all the nodes, see Fig. 3.2.23(a), in the 1310–1510 km depth layer. Fig. 3.2.23(c) displays all the uncertainties[††], $\sigma_{\hat{m}_k}$, corresponding to the local-average estimates, $\hat{m}_k$, for the same 1310–1510 km depth layer.
  3. *Asking a question and getting a quantitative answer (model appraisal).*
     - **Observation**. Let focus on the Farallon subduction system (positive velocity anomalies), taking place, in particular, in mid lower mantle, see Fig. 3.2.23(d) (zoom-in on the SOLA model at 1310–1410 km depth). In addition to the main subducted slab ('Farallon') beneath eastern North-America, one can identify two detached slab fragments located further to the West ('F1' and 'F2'). I have reported, in Zaroli (2016), that F2 appears to be well localized, offshore California, in the 1310–1510 km depth layer. Although F1 and F2 have not been so sharply imaged in previous global-scale *P*- and *S*-wave models (e.g., Van der Hilst et al., 1997; Montelli et al., 2004b; Ritsema et al., 2011),

[†] The model estimates, obtained for all the nodes in a layer, are linearly interpolated, using barycentric coordinates (see Zaroli (2010)), to build a tomographic image.

[††] Note the global pattern: low/high uncertainties in all the regions with low/high ray-path density, and thus with large/small size target kernels (and, consequently, large/small size resolving kernels), respectively. This means that, here, with the given ray-path coverage and chosen target kernels, the number of 'rays' *significantly contributing* to the uncertainty tends to be higher in the regions with poorer resolution.

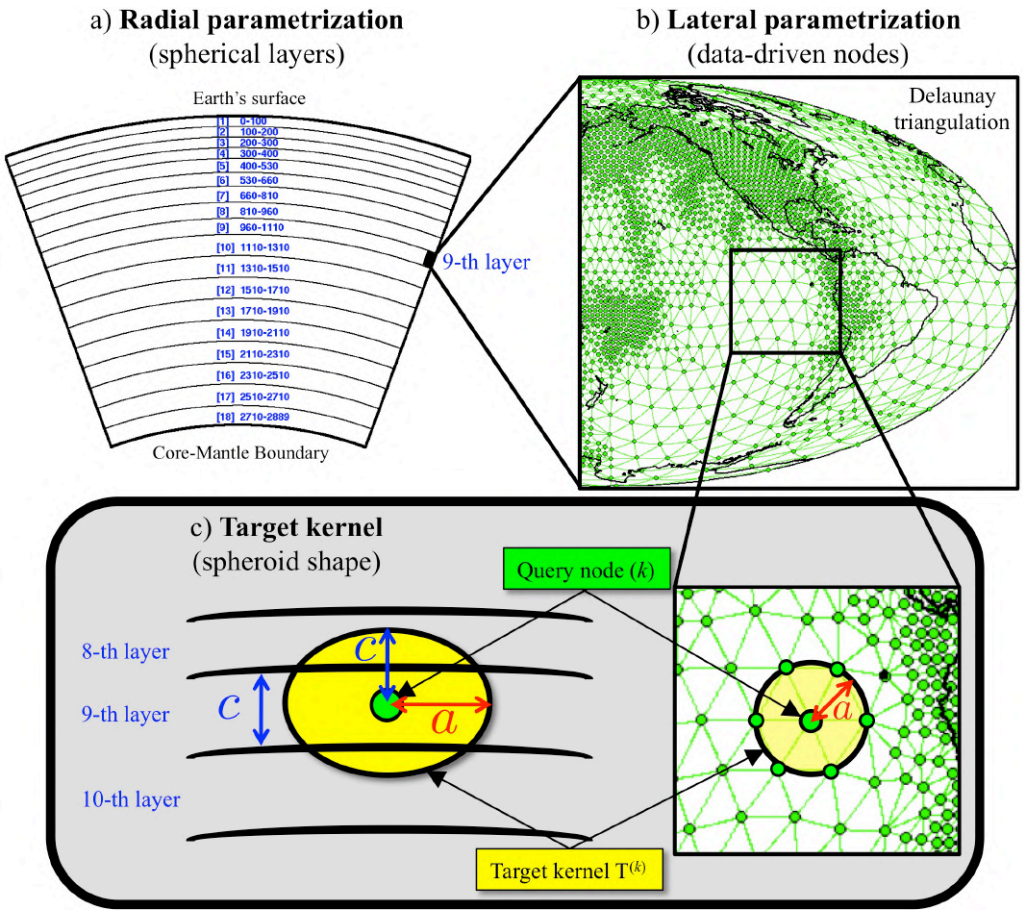


Figure 3.2.22: Irregular, data-driven parametrization of model space (whole mantle) – (a) Radial parametrization: eighteen spherical layers. (b) Lateral parametrization: ray-density driven Delaunay triangulation, obtained through an optimization process. Example: ninth layer (960–1110 km depth). The green dots represent the 'nodes' (model parameters), which are the vertices of every spherical triangles, located at the middle depth of the ninth layer. (c) Illustration of a spheroid-shape, SOLA target kernel, $T^{(k)}$, centred on a 'query' node (k). The lateral (radial) radius of $T^{(k)}$, denoted by a (c), represents the *a priori*, local, lateral (radial) resolving length, respectively. Here, a is the average distance between the query node (k) and its six direct neighbor nodes, and c is the width of the ninth layer. Modified from Zaroli et al. (2015) and Zaroli (2016).

recent regional-scale *P*-wave studies (Sigloch, 2011; Sigloch & Mihalynuk, 2013) have imaged both F1 and F2 and proposed that F1 could correspond to the deep root in lower mantle of the Cascadia subduction system, and F2 be related to some intra-oceanic subduction. However, in the aforementioned studies[†], mainly checker-board sensitivity tests were used to assess the resolution – coming with a number of potential drawbacks (e.g., Lévêque et al., 1993; Rawlinson & Spakman, 2016) – and there was no quantification of uncertainties.

[†] All these studies were based on a classical, damped least-squares (DLS) inversion approach.

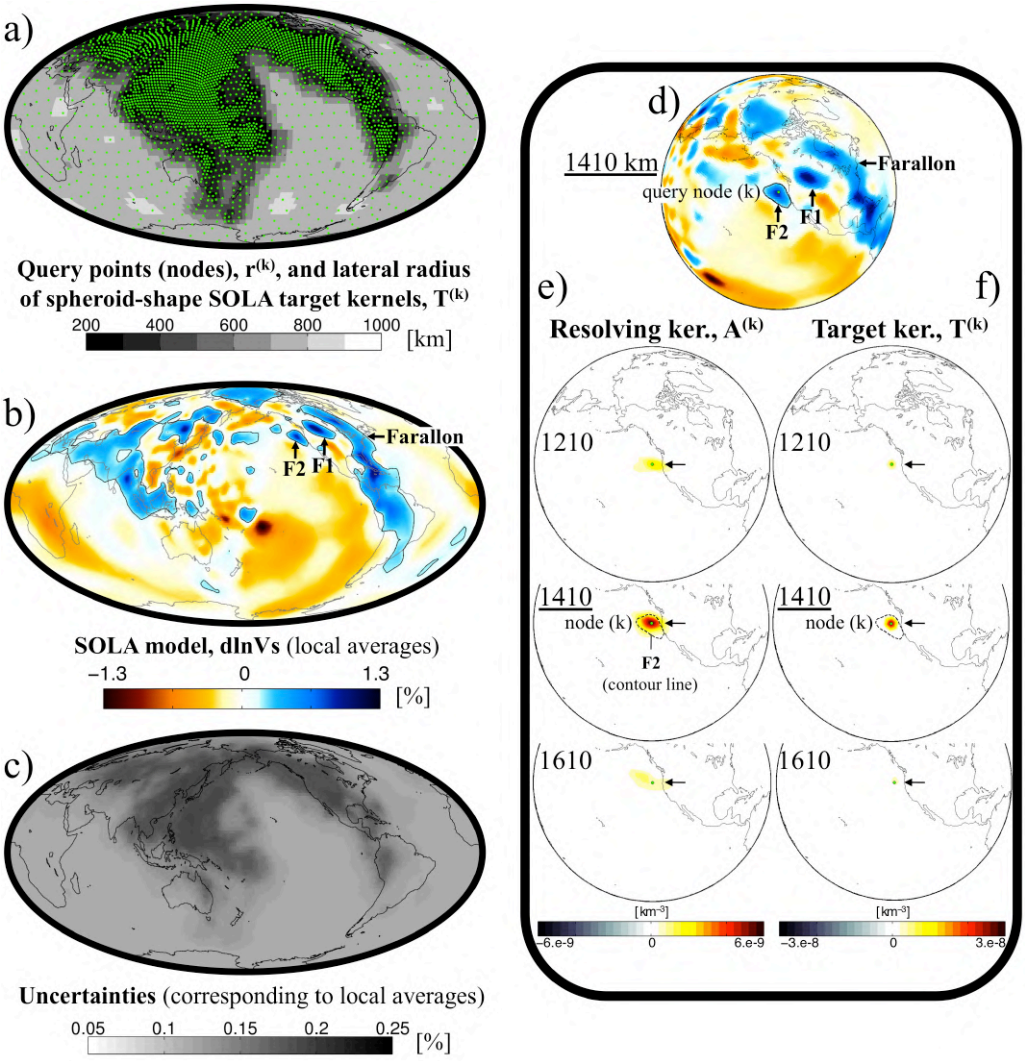


Figure 3.2.23: Global S-wave mantle tomography using *discrete* SOLA inversion — (a) Query nodes, $\mathbf{r}^{(k)}$, displayed as green dots in the 1310–1510 km depth layer, and lateral radius of spheroid-shape target kernels, $T^{(k)}$, that is, the *a priori*, local, lateral resolving length (black-to-white color scale). (b) 'Conventional' tomographic image showing the SOLA model estimates, $\hat{m}_k$ ($d \ln V_S$), in the 1310–1510 km depth layer. (c) 'Conventional' image showing the corresponding uncertainties, $\sigma_{\hat{m}_k}$, in the 1310–1510 km depth layer. (d) Zoom-in on the tomographic model around the feature F2. Remark: 1410 km (1210, 1610) denotes the 1310–1510 km depth layer (1110–1310, 1510–1710), respectively. (e) Constant-depth, 2–D sections showing the resolving kernel, $A^{(k)}$, obtained for the query node (k), centered on F2. (f) Corresponding target kernel, $T^{(k)}$. Modified from Zaroli (2016).

– **Question**. Is the feature F2 really resolved (at least in our model)?

– **Resolving length**. In Fig. 3.2.23(d), a zoom-in on the SOLA model is shown, at 1310–1510 km depth; F2 is contoured with a black solid

line, and the green dot represents the 'query' node (k) located at the center of F2. The corresponding target kernel $T^{(k)}$ is displayed in Fig. 3.2.23(f). Its form is that of a spheroid, centred on the query node (k), at 1410 km depth, with lateral and radial radii 200 km each (in fact, here, $T^{(k)}$ is a sphere). The resolving kernel $A^{(k)}$ is shown in Fig. 3.2.23(e), in the 1110–1710 km depth range. Note that $A^{(k)}$ is confined into a small enough spatial domain, near the query node (k), with respect to the size of F2. Thus, the local resolving length, both laterally and radially, is small enough to claim that F2 is resolvable[†].

– **Local average/uncertainty**. Moreover, the local weighted-average estimate, and its uncertainty, are: $\hat{m}_k = 0.85\,\%$ and $\sigma_{\hat{m}_k} = 0.18\,\%$, respectively – that is, $\sigma_{\hat{m}_k}$ is almost five times smaller than $\hat{m}_k$.

– **Answer**. Therefore, one can argue that the feature F2 is resolved (given the data and their errors).

[†] Although that resolution analysis was focused on a *single* query point (centered on F2), in general, it should be better to investigate the resolution in a *vicinity* of the feature of interest – when local resolution varies rapidly.

3.3 SOLA vs DLS tomography

Preamble. Without loss of generality, I shall consider a *discrete* linear tomographic problem, for which 'true' parameters are *perturbations* with respect to a reference model (in a probabilistic viewpoint, the *a priori* model would be zero). For more details on least-squares inverse theory, for example on damped least-squares (DLS) inversion – one of the most popular inversion approach in seismic tomography (in one form or another) – the reader is referred to: Menke (1989), Snieder & Trampert (1999), Tarantola (2005), Nolet (2008), Aster et al. (2012), Voronin & Zaroli (2018).

3.3.1 Discrete DLS inversion

- **DLS inversion: basic principles**. From a 'least-squares inversion' point of view, the model solution, $\hat{\mathbf{m}}$, corresponding to the linear, discrete tomographic problem, $\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{n}$, would be the one that gives the best fit to the data[††]:

$$\hat{\mathbf{m}} \;=\; \underset{\hat{\mathbf{m}} \in \mathbb{R}^M}{\text{argmin:}}\; \underbrace{(\mathbf{G}\hat{\mathbf{m}} - \mathbf{d})^T\, \mathbf{C}_d^{-1}\, (\mathbf{G}\hat{\mathbf{m}} - \mathbf{d})}_{\substack{\text{DATA} \\ \text{MISFIT}}} \tag{3.3.1}$$

where $\mathbf{C}_d$ is the data covariance matrix, usually assumed to be diagonal: $\mathbf{C}_d = \mathrm{diag}(\sigma_{d_i}^2)_{1 \le i \le N}$ (uncorrelated, zero-mean noise, $\mathbf{n}$). Owing to data errors and 'imperfect' data coverage, least-squares inverse problems (3.3.1) are, at least partly, 'underdetermined', causing the non-unicity of the model solution. To remove that non-uniqueness, one may seek a model that minimizes both the data misfit *and* the model 'complexity' (a process named *model regularization*), leading to the so-called *damped least-squares* (DLS) inversion approach, for which the minimisation problem can be formulated as follows:

[††] Remark: $\underset{\hat{\mathbf{m}} \in \mathbb{R}^M}{\text{argmin:}}\, \mathcal{F}(\hat{\mathbf{m}})$ means that one seeks a real-valued, M-dimensional, *vector*, $\hat{\mathbf{m}}$, minimizing $\mathcal{F}(\hat{\mathbf{m}})$.

$$\hat{\mathbf{m}} \;=\; \underset{\hat{\mathbf{m}}\in\mathbb{R}^M}{\text{argmin:}} \; \underbrace{(\mathbf{G}\hat{\mathbf{m}}-\mathbf{d})^T\,\mathbf{C}_d^{-1}\,(\mathbf{G}\hat{\mathbf{m}}-\mathbf{d})}_{\substack{\text{DATA}\\ \text{MISFIT}}} \;+\; \underbrace{\hat{\mathbf{m}}^T\mathbf{C}_m^{-1}\hat{\mathbf{m}}}_{\substack{\text{MODEL}\\ \text{REGULARIZATION}}} \quad . \tag{3.3.2}$$

The DLS solution can be written[†] in terms of its 'generalized inverse', $\mathbf{G}^\dagger$:

$$\begin{cases} \underbrace{\hat{\mathbf{m}}}_{\substack{\text{DLS model}\\ \text{estimate}}} \;=\; \underbrace{\left[(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G}+\mathbf{C}_m^{-1})^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}\right]}_{\substack{\text{DLS 'generalized inverse'}\\ \text{matrix, } \mathbf{G}^\dagger \text{ (see text box 3.6)}}} \times \mathbf{d} \;=\; \underbrace{\mathbf{R}\mathbf{m}}_{\substack{\text{filtered}\\ \text{true model}}} \;+\; \underbrace{\mathbf{G}^\dagger\mathbf{n}}_{\substack{\text{propagated}\\ \text{noise}}} \;, \\ \text{with} \quad \underbrace{\mathbf{R}=\mathbf{G}^\dagger\mathbf{G}}_{\substack{\text{model-estimate}\\ \text{resolution matrix}}} \;, \text{ and } \quad \underbrace{\mathbf{C}_{\hat{m}}=\mathbf{G}^\dagger\mathbf{C}_d(\mathbf{G}^\dagger)^T}_{\substack{\text{model-estimate}\\ \text{covariance matrix}}} . \end{cases} \tag{3.3.3}$$

[†] In this chapter some notations ($\hat{\mathbf{m}}$, $\mathbf{G}^\dagger$, $\mathbf{R}$, $\mathbf{C}_{\hat{m}}$) are kept the same for both the (discrete) SOLA and DLS inversion schemes. It should be clear from the context to which one refers to.

In a 'probabilistic' view, $\mathbf{C}_m$ should represent a *prior* model covariance matrix,

$$\mathbf{C}_m \;=\; \begin{pmatrix} \sigma^2_{m_1} & \cdots & \text{off-diag.} \\ \vdots & \ddots & \vdots \\ \text{off-diag.} & \cdots & \sigma^2_{m_M} \end{pmatrix} , \tag{3.3.4}$$

and $\hat{\mathbf{m}}$ a 'maximum likelihood' solution – though it may be difficult[††] to determine 'objective', physically-based, $\mathbf{C}_m$'s (i.e., *a priori* informations on the model itself). For example, there is not much *prior* knowledge on the characteristics (e.g., extreme values, spatial correlations) of seismic-velocity anomalies in the (deep) Earth. Recognizing this difficulty, the DLS approach used in many tomographic applications aims at mutually minimizing the data misfit and a 'damped' L_2-norm[†††] of the model (i.e., 'subjective' regularization):

$$\underbrace{\hat{\mathbf{m}}}_{\substack{\text{DLS model}\\ \text{estimate}}} \;=\; \underset{\hat{\mathbf{m}}\in\mathbb{R}^M}{\text{argmin:}} \; \underbrace{\sum_{i=1}^{N}\Big(\sum_{j=1}^{M} G_{ij}\hat{m}_j - d_i\Big)^2/\sigma^2_{d_i}}_{\text{DATA MISFIT } \left(=N\times\chi^2_{\text{red}}\right)} + \underbrace{\Theta^2}_{\substack{\text{trade-off via the}\\ \text{damping factor}}} \times \underbrace{||\hat{\mathbf{m}}||^2}_{\substack{\text{MODEL}\\ \text{NORM}}}, \tag{3.3.5}$$

[††] That is 'the' reason why the (SOLA) B–G philosophy carefully avoids using any *a priori* information on the model itself (see Sect. 3.1.2).

[†††] L_2-norm: $||\hat{\mathbf{m}}|| = \left(\sum_j \hat{m}_j^2\right)^{\frac{1}{2}}$. Other *ad hoc* model regularization constraints (L_1-norm, roughening operator, etc.), that is, other $\mathbf{C}_m$'s, might be preferred.

which corresponds to defining $\mathbf{C}_m$ through the simple (*ad hoc*) form[††††]:

$$\mathbf{C}_m \;=\; \Theta^{-2}\times\mathbf{I}_M \qquad \text{(zeroth-order Tikhonov regularization)}\,, \tag{3.3.6}$$

with Θ the so-called *damping factor* and $\mathbf{I}_M$ the identity matrix of order M.

[††††] With *irregular* parameterizations, one should rather use: $\mathbf{C}_m \propto \Theta^{-2} \times \text{diag}(1/\mathbb{V}_j)_{1\le j\le M}$ (Zaroli et al., 2015).

Note that, here, the DLS solution ($\hat{\mathbf{m}}$, $\mathbf{R}$, $\mathbf{C}_{\hat{m}}$, see (3.3.3)) clearly depends on the (*ad hoc*) damping factor Θ, through the term $\mathbf{C}_m$, see (3.3.6). As a remark, if $\mathbf{d}$ and $\mathbf{G}$ were scaled by the data errors, and if $\mathbf{C}_m = \Theta^{-2}\mathbf{I}_M$, then $\mathbf{C}_d = \mathbf{I}_N$ and the DLS generalized-inverse matrix writes: $\mathbf{G}^\dagger = (\mathbf{G}^T\mathbf{G} + \Theta^2\mathbf{I}_M)^{-1}\mathbf{G}^T$.

Text Box 3.6: DAMPED LEAST-SQUARES (DLS) GENERALIZED INVERSE

Let the quantity $\mathcal{F}(\hat{\mathbf{m}}) = (\mathbf{G}\hat{\mathbf{m}} - \mathbf{d})^T\,\mathbf{C}_d^{-1}\,(\mathbf{G}\hat{\mathbf{m}} - \mathbf{d}) + \hat{\mathbf{m}}^T\mathbf{C}_m^{-1}\hat{\mathbf{m}}$, that we wish to minimize, see (3.3.2). Here, we consider that the model parameters represent (e.g., velocity) *perturbations* with respect to a reference model (the *a priori* model would then be zero, in a probabilistic view), and do not consider any particular form for $\mathbf{C}_d$ or $\mathbf{C}_m$, except that the data covariance matrix $\mathbf{C}_d$ is symmetric. We want to show that $\frac{\partial\mathcal{F}(\hat{\mathbf{m}})}{\partial\hat{\mathbf{m}}} = \mathbf{0}_M$ leads to the following system: $\mathbf{G}^T\mathbf{C}_d^{-1}\,(\mathbf{G}\hat{\mathbf{m}} - \mathbf{d}) + \mathbf{C}_m^{-1}\hat{\mathbf{m}} = \mathbf{0}_M$, from which it is trivial to infer the general expression for the DLS generalized inverse matrix, $\mathbf{G}^\dagger$, that is: $\hat{\mathbf{m}} = \mathbf{G}^\dagger \times \mathbf{d} = \left[\left(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}\right)^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}\right] \times \mathbf{d}$, see (3.3.3).

Taking the derivative of $\mathcal{F}(\hat{\mathbf{m}})$ with respect to $\hat{m}_k$, and setting the result to zero, we have: $\frac{\partial\mathcal{F}(\hat{\mathbf{m}})}{\partial\hat{m}_k} = \frac{\partial\hat{\mathbf{r}}^T\mathbf{C}_d^{-1}\hat{\mathbf{r}}}{\partial\hat{m}_k} + \frac{\partial\hat{\mathbf{m}}^T\mathbf{C}_m^{-1}\hat{\mathbf{m}}}{\partial\hat{m}_k} = 0$, where the parameter index k is fixed, and $\hat{\mathbf{r}} = \mathbf{G}\hat{\mathbf{m}} - \mathbf{d}$ (data residual). Expanding out the partial-derivative terms, we obtain:

$$
\left\{
\begin{array}{rcl}
\frac{\partial\hat{\mathbf{r}}^T\mathbf{C}_d^{-1}\hat{\mathbf{r}}}{\partial\hat{m}_k} &=& \frac{\partial}{\partial\hat{m}_k}\left(\sum_{i=1}^N \hat{r}_i \sum_{p=1}^N \left(C_d^{-1}\right)_{ip} \hat{r}_p\right) \\
&=& \sum_{i=1}^N \left(\underbrace{\frac{\partial\hat{r}_i}{\partial\hat{m}_k}}_{G_{ik}} \sum_{p=1}^N \left(C_d^{-1}\right)_{ip} \hat{r}_p + \hat{r}_i \sum_{p=1}^N \left(C_d^{-1}\right)_{ip} \underbrace{\frac{\partial\hat{r}_p}{\partial\hat{m}_k}}_{G_{kp}} \right) \\
&=& \sum_{i=1}^N G_{ik} \sum_{p=1}^N \left(C_d^{-1}\right)_{ip} \left(\sum_{j=1}^M G_{pj}\hat{m}_j - d_p\right) \\
&& + \sum_{p=1}^N G_{pk} \sum_{i=1}^N \underbrace{\left(C_d^{-1}\right)_{ip}}_{\left(C_d^{-1}\right)_{pi}} \left(\sum_{j=1}^M G_{ij}\hat{m}_j - d_i\right) \\
&=& 2 \times \sum_{i=1}^N G_{ik} \sum_{p=1}^N \left(C_d^{-1}\right)_{ip} \left(\sum_{j=1}^M G_{pj}\hat{m}_j - d_p\right) \\
\frac{\partial\hat{\mathbf{m}}^T\mathbf{C}_m^{-1}\hat{\mathbf{m}}}{\partial\hat{m}_k} &=& \frac{\partial}{\partial\hat{m}_k}\left(\sum_{j=1}^M \hat{m}_j \sum_{l=1}^M \left(C_m^{-1}\right)_{jl} \hat{m}_l\right) \\
&=& \frac{\partial}{\partial\hat{m}_k}\left(\hat{m}_k \sum_{l=1}^M \left(C_m^{-1}\right)_{kl} \hat{m}_l + \sum_{\substack{j=1 \\ j\neq k}}^M \hat{m}_j \sum_{l=1}^M \left(C_m^{-1}\right)_{jl} \hat{m}_l \right) \\
&=& \sum_{l=1}^M \left(C_m^{-1}\right)_{kl} \hat{m}_l + \left(C_m^{-1}\right)_{kk} \hat{m}_k + \sum_{\substack{j=1 \\ j\neq k}}^M \left(C_m^{-1}\right)_{jk} \hat{m}_j \\
&=& 2 \times \sum_{l=1}^M \left(C_m^{-1}\right)_{kl} \hat{m}_l
\end{array}
\right.
$$

$$
\Longrightarrow \sum_{i=1}^N G_{ik} \sum_{p=1}^N \left(C_d^{-1}\right)_{ip} \left(\sum_{j=1}^M G_{pj}\hat{m}_j - d_p\right) + \sum_{l=1}^M \left(C_m^{-1}\right)_{kl} \hat{m}_l = 0\,,\quad 1 \leq k \leq M\,,
$$

which can be written in the matrix form: $\mathbf{G}^T\mathbf{C}_d^{-1}\,(\mathbf{G}\hat{\mathbf{m}} - \mathbf{d}) + \mathbf{C}_m^{-1}\hat{\mathbf{m}} = \mathbf{0}_M$. ⊡

- **DLS trade-off curve**. One way to select an appropriate Θ consists in analyzing the trade-off curve between the reduced chi-square measure of the data misfit, $\chi^2_{\text{red}} = \chi^2_{\text{red}}(\mathbf{d}, \hat{\mathbf{m}}(\Theta))$, and the L_2-norm of the model, $||\hat{\mathbf{m}}||^2 = ||\hat{\mathbf{m}}(\Theta)||^2$, see

Fig. 3.3.1. This way, it is possible to deal with the problem of non-uniqueness[†] (of the least-squares solution) and construct 'reasonable' models. It allows one to seek the most data-fitting *and* the less complex model (Occam's razor principle) – here, the model complexity is simply measured by the L_2-norm.

[†] In contrast to the regularization in DLS through 'damping' that acts directly on the model values, the problem of non-uniqueness is in the SOLA approach dealt with by explicitly computing averaging kernels under the *unbiased averaging condition* and the constraint on the model variance controlled by the trade-off parameter.

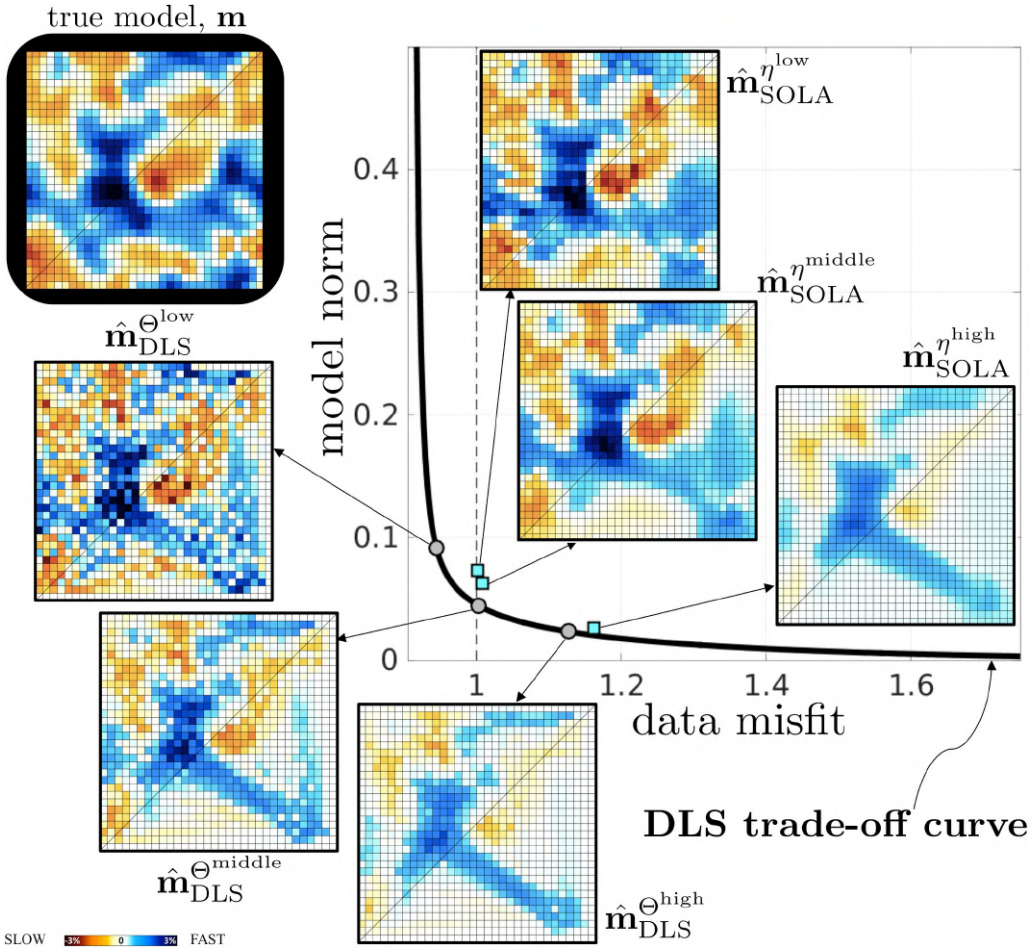


Figure 3.3.1: DLS trade-off curve, $||\hat{\mathbf{m}}(\Theta)||^2$ ('model norm') *versus* $\chi^2_{\text{red}}(\mathbf{d}, \hat{\mathbf{m}}(\Theta))$ ('data misfit'), corresponding to the 'toy' problem, see Fig. 3.2.16. Three DLS (SOLA) models are displayed for decreasing values of Θ (η), that is: $\hat{\mathbf{m}}^{\Theta^{\text{high}}}_{\text{DLS}}$, $\hat{\mathbf{m}}^{\Theta^{\text{middle}}}_{\text{DLS}}$ and $\hat{\mathbf{m}}^{\Theta^{\text{low}}}_{\text{DLS}}$ ($\hat{\mathbf{m}}^{\eta^{\text{high}}}_{\text{SOLA}}$, $\hat{\mathbf{m}}^{\eta^{\text{middle}}}_{\text{SOLA}}$ and $\hat{\mathbf{m}}^{\eta^{\text{low}}}_{\text{SOLA}}$ – see Fig. 3.2.18), respectively. Model $\hat{\mathbf{m}}^{\Theta^{\text{middle}}}_{\text{DLS}}$ was defined such that $\chi^2_{\text{red}}(\mathbf{d}, \hat{\mathbf{m}}^{\Theta^{\text{middle}}}_{\text{DLS}}) \to 1$. Remark: $\Theta^{\text{high}} = \Theta^{\text{middle}} \times 1.9$ and $\Theta^{\text{low}} = \Theta^{\text{middle}} \times 0.4$; $\chi^2_{\text{red}}(\mathbf{d}, \hat{\mathbf{m}}^{\Theta^{\text{high}}}_{\text{DLS}}) = 1.126$, $\chi^2_{\text{red}}(\mathbf{d}, \hat{\mathbf{m}}^{\Theta^{\text{middle}}}_{\text{DLS}}) = 1.002$ and $\chi^2_{\text{red}}(\mathbf{d}, \hat{\mathbf{m}}^{\Theta^{\text{low}}}_{\text{DLS}}) = 0.944$. Note that: the two SOLA models $\hat{\mathbf{m}}^{\eta^{\text{low}}}_{\text{SOLA}}$ and $\hat{\mathbf{m}}^{\eta^{\text{middle}}}_{\text{SOLA}}$ do fit the data as well as the DLS model $\hat{\mathbf{m}}^{\Theta^{\text{middle}}}_{\text{DLS}}$, and are *visually closer* to the 'true' model, $\mathbf{m}$, in particular where the data coverage is poor (lower-right triangle). Modified from Zaroli et al. (2017).

Note that, the lower Θ, the lower $\chi^2_{\text{red}}(\mathbf{d}, \hat{\mathbf{m}}(\Theta))$ and the higher $||\hat{\mathbf{m}}(\Theta)||^2$. In general, one may also expect that, the lower Θ, the better the model resolution

and the higher the model uncertainties (merely describing the propagation of noise into the model) – though, as we shall see, this view may sometimes be too simplistic and even erroneous[†]. If the data errors (σ_{d_i}) were well known, one should select, from a statistical point of view, the value for Θ corresponding to $\chi^2_{\text{red}} = 1$. Otherwise, a common practice consists in trying to identify the 'corner' of the 'L-shape' trade-off curve, and select the corresponding Θ value. It may be useful to analyze a couple of model solutions, for lower and higher Θ values, to better apprehend the range of 'acceptable' solutions, see Fig. 3.3.1.

[†] That is, the DLS solution (hence the resolution and uncertainties) can be locally 'biased'.

- **Averaging bias effect**. Figure 3.3.2 shows the three DLS models ($\hat{\mathbf{m}}^{\Theta^{\text{high}}}_{\text{DLS}}$, $\hat{\mathbf{m}}^{\Theta^{\text{middle}}}_{\text{DLS}}$, $\hat{\mathbf{m}}^{\Theta^{\text{low}}}_{\text{DLS}}$), marked on the DLS trade-off curve displayed in Fig. 3.3.1.

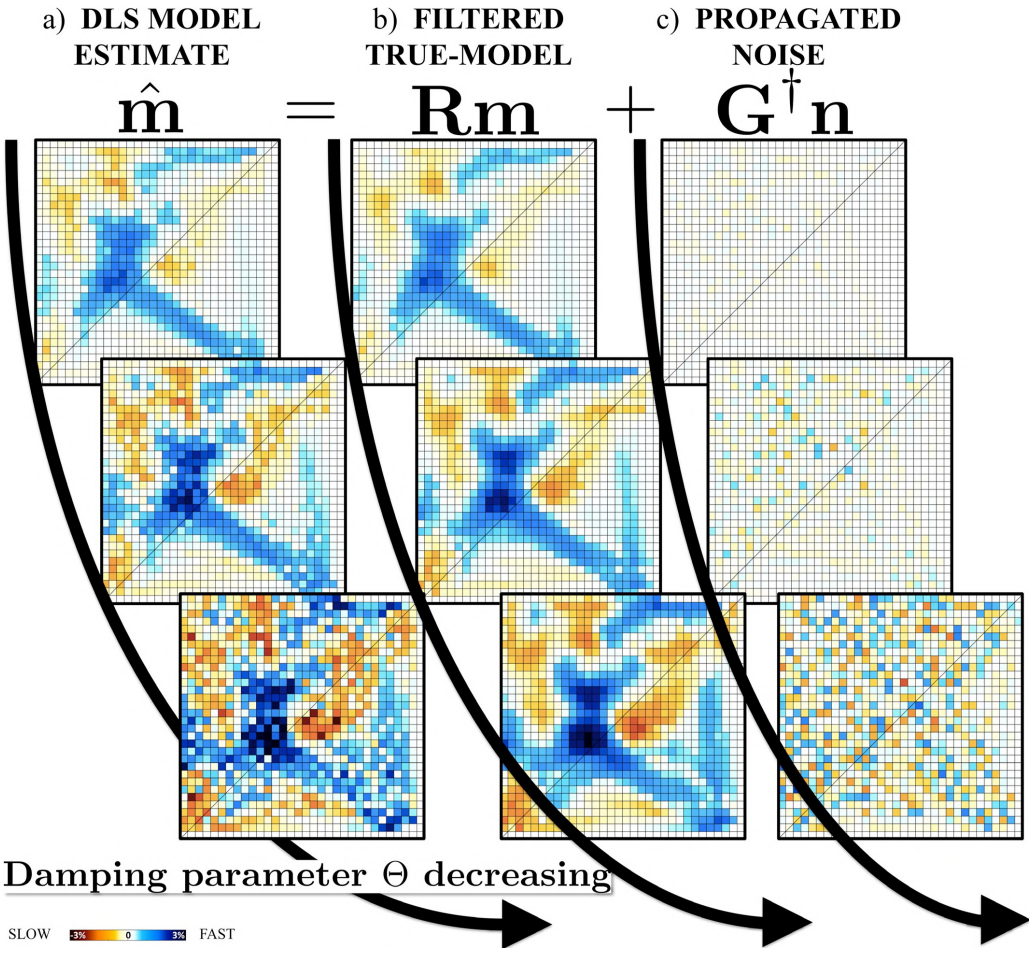


Figure 3.3.2: (a) DLS models, $\hat{\mathbf{m}}$, obtained for different values of damping parameter Θ (from top to bottom: Θ^{high}, Θ^{middle}, Θ^{low}, respectively, with $\Theta^{\text{high}} = \Theta^{\text{middle}} \times 1.9$, and $\Theta^{\text{low}} = \Theta^{\text{middle}} \times 0.4$), corresponding to the 'toy' problem, see Fig. 3.2.16. The model estimate, $\hat{\mathbf{m}}$, is decomposed into two terms: (b) the filtered true-model, $\mathbf{Rm}$, and (c) the propagated noise, $\mathbf{G}^{\dagger}\mathbf{n}$. These three models, $\hat{\mathbf{m}}$, are marked on the DLS trade-off curve shown in Fig. 3.3.1. Modified from Zaroli et al. (2017).

These DLS models, $\hat{\mathbf{m}}$, are decomposed into the sum of two contributions: the filtered true-model, $\mathbf{Rm}$, and the propagated noise, $\mathbf{G}^{\dagger}\mathbf{n}$. Note that, in the regions where the ray coverage is good (upper left triangle), the SOLA and DLS filtered true-models look similar, see Figs. 3.2.18(b) and 3.3.1(b), respectively, indicating that their resolution, $\mathbf{R}$, is locally similar. However, in the regions of poor ray coverage (lower right triangle), the SOLA and DLS filtered true-models often differ – more or less, depending on the true model, $\mathbf{m}$, and on the actual SOLA and DLS resolving kernels (i.e. the rows of $\mathbf{R}$).

For a robust interpretation of DLS or SOLA tomographic images (that is, for a robust model appraisal), we wish that $\mathbf{Rm}$ represents a weighted-averaging process over the true-model parameters, $\mathbf{m}$. The k-th row of the resolution matrix, $\mathbf{R}$, here referred to as $\mathbf{R}_{k.} = (R_{kj})_{1\leq j\leq M}$, represents[†] the resolving (averaging) kernel, linearly relating $\hat{m}_k$ to the true parameters ($\hat{m}_k \approx \sum_{j=1}^{M} R_{kj} m_j$). Let us define the *averaging bias* quantity, U_k, related to $\hat{m}_k$, as follows:

$$U_k = \sum_{j=1}^{M} R_{kj} \,. \tag{3.3.7}$$

[†] *Stricto senso*, it is $A^{(k)}$ which represents the resolving (averaging) kernel. However, $\mathbf{R}_{k.}$ and $A^{(k)}$ are closely related, see (3.2.28).

Therefore, for every index k, we wish that $\mathbf{R}_{k.}$ be: 1) spatially localized (near the location of $\hat{m}_k$), 2) non-negative, and 3) such that $U_k = 1$.

Contrary to the SOLA approach[††], the DLS inversion scheme may lead to model estimates that are locally *biased* toward either lower (if $U_k < 1$) or higher (if $U_k > 1$) values, especially in the regions of poor data illumination (Zaroli et al., 2017). *Remark*: Since uneven data coverage has been, and will remain, a serious issue in seismic tomography, from local to global scale, such averaging bias effect should systematically be investigated in DLS models.

[††] In the SOLA approach, specifying a target resolving kernel $T^{(k)}$ favors $\mathbf{R}_{k.}$ to be well localized and non-negative, and from the very formulation of the minimization problem, U_k is guaranteed to be one, see (3.2.24).

Fig. 3.3.3 aims at illustrating this point: two input 'true' models, $\mathbf{m}$, are filtered through the DLS resolving (averaging) kernels[†††], that is, $\mathbf{Rm}$. Fig. 3.3.3(c) displays the DLS *averaging bias* quantity, U_k, corresponding to the damping factor Θ^{middle}. Clearly, some apparent structural features, in the DLS filtered models ($\mathbf{Rm}$), can be directly related to the averaging bias effect. For example, in the lower right corner (of the model space) the bias effect consists in locally increasing, up to 80%, the amplitudes, thus artificially giving rise to the darker blue anomaly – that Sherlock Holmes is staring at, in Fig. 3.3.3(b).

In those poorly covered regions, the input anomalies also tend to be more *stretched* (smeared and elongated) in the DLS filtered models, see Fig. 3.3.3(b), than in the SOLA ones, see Fig. 3.2.20(c). Indeed, the shape of DLS resolving kernels is often more stretched, compared to SOLA kernels, see Fig. 3.3.4.

[†††] The two terminologies, 'resolving kernel' and 'averaging kernel', are alternatively used in the literature and this manuscript. However, when a (DLS) resolving kernel is *biased* ($U_k \neq 1$), it would be better to not refer to it as an averaging kernel (since it does not describe an averaging process).

Summary. In regions of poor data illumination, DLS tomographic models may suffer from an averaging bias effect (and some enhanced stretching effect) – so that the 'true-model' features may be better retrieved in SOLA models.

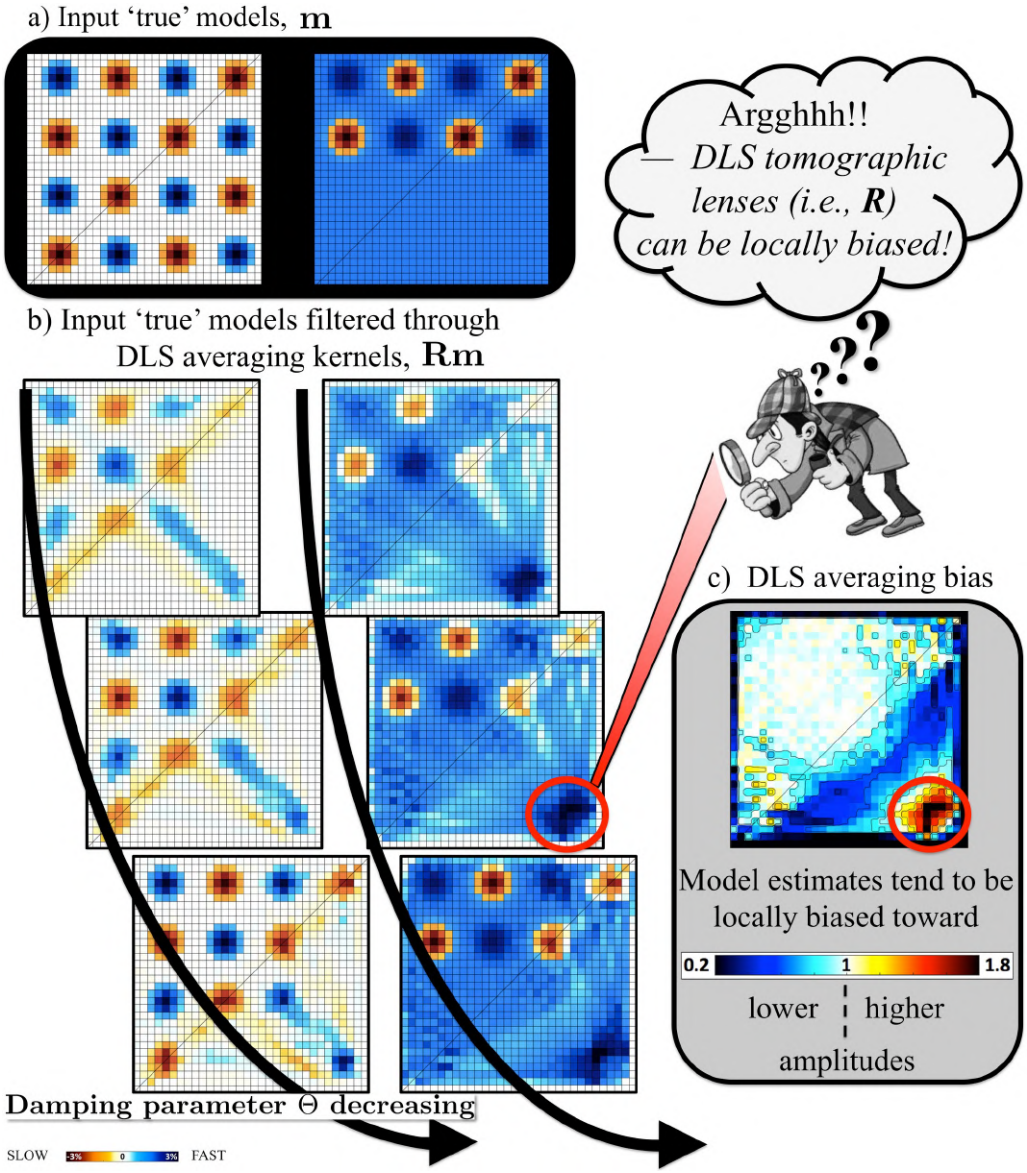


Figure 3.3.3: (a) Input 'true' models, **m**, filtered through (b) the DLS averaging (resolving) kernels, **Rm**, for decreasing values of the damping factor Θ (from top to bottom: Θ^{high}, Θ^{middle}, Θ^{low}, respectively). (c) Averaging bias quantity, $U_k = \sum_{j=1}^{M} R_{kj}$, corresponding to Θ^{middle}. Some apparent structural features in DLS filtered models (**Rm**) can be directly related to the averaging bias effect, such as in the lower right corner of the model space (poor data illumination) where the bias effect is to locally increase up to 80% ($U_k = 1.8$) the amplitudes, thus artificially giving rise to the darker blue anomaly, that Sherlock Holmes is staring at. Modified from Zaroli et al. (2017).

- **Resolving kernels**. DLS resolving kernels can sometimes be misleading, in terms of model interpretation (appraisal). For example, Fig. 3.3.4(e) shows

a DLS resolving kernel, $\mathbf{R}_{k'.}$, located in the lower-right corner region (dense, anisotropic ray coverage); index k' refers to the considered pixel. Its spatial extent is closely restricted around the k'-th pixel, and its value at the k'-th pixel location is relatively[†] high ($R_{k'k'} = 0.46$). At first glance, one could be tempted to conclude that the value of $\hat{m}_{k'}$ represents a true-model averaging process taking place over a small-size area around the k'-th pixel ('good' resolution), with a moderate uncertainty ($\sigma_{\hat{m}_{k'}} = 0.23\%$), on that local average.

However, since the averaging bias quantity $U_{k'}$ is larger than one ($U_{k'} = 1.8$), see Fig. 3.3.3(c), the value of $\hat{m}_{k'}$ actually represents a *biased* local average (biased toward 80% higher amplitude). Such an averaging bias makes it difficult to interpret apparent features in DLS models (near the k'-th pixel, in this example). The appraisal of DLS models would even be more difficult if U_k were not quantified. I urge the reader to do so in the case of DLS tomography (that is, to calculate[††] and carefully analyze the averaging bias quantity).

As a final remark, Fig. 3.3.4(j) shows that the k'-th SOLA averaging kernel[†††] expands over a broad-size region. That means that the SOLA resolution is quite 'poor' near the k'-th pixel, as expected from the local, very anisotropic, ray-path distribution, see Fig. 3.2.16(a).

[†] If $R_{k'k'}$ was equal to one (and $\mathbf{R}_{k'.}$ be zero elsewhere), the resolution corresponding to the k'-th pixel would then be 'perfect'.

[††] If possible in terms of computational cost, see toward the end of Sect. 3.3.1.

[†††] The k'-th SOLA averaging kernel is guaranteed to be unbiased, see (3.2.24), that is: $U_{k'} = 1$.

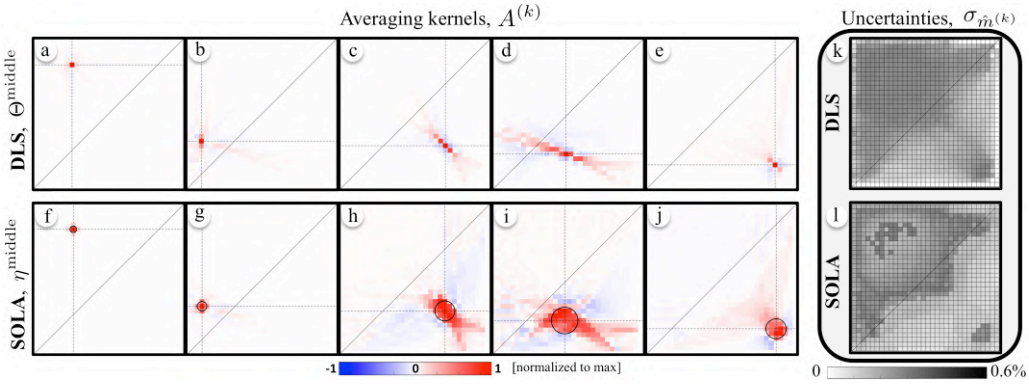


Figure 3.3.4: (a–e) Five examples of DLS resolving (averaging) kernels, $A^{(k)}$, for the damping factor Θ^{middle}. As a remark, $A^{(k)}$ is almost equivalent to $\mathbf{R}_{k.}$, that is, the k-th row of $\mathbf{R}$. (k) DLS model uncertainties, $(\sigma_{\hat{m}_k})_{1\leq k\leq M}$, for Θ^{middle}. (f–j) Five examples of SOLA averaging kernels, for the trade-off parameter η^{middle}. The black circles represent the SOLA target kernels, $T^{(k)}$. (l) SOLA model uncertainties for η^{middle}. As a reminder, these uncertainties are merely describing the statistical propagation of data noise into the model estimates. Modified from Zaroli et al. (2017).

- **SOLA *vs* DLS models**. Fig. 3.3.1[††††] shows that, with the aforementioned recipe for tuning target kernels and trade-off parameters, one can easily obtain a (discrete) SOLA model which does fit the data as well as the 'best' DLS model (the model for which $\chi^2_{\text{red}} = 1$), *and* whose the apparent features seem to be 'visually closer', when compared to that 'best' DLS model, to those from the 'true' model – in particular in the regions with poor ray-path illumination.

[††††] See text emphasized in red in figure's caption.

- **Global mantle tomography: uncovering the bias effect**. We aim at quantifying the averaging bias effect in whole-mantle, 'real-data-based', DLS tomographic models (Zaroli et al., 2017). When considering a large number of data and model parameters, computing the DLS model resolution matrix, required to evaluate the DLS bias effect, can become computationally challenging[†]. Thus, in most global-scale DLS tomographic studies, the resolution matrix has often been ignored, or at best its diagonal elements have been approximated (e.g., Trampert et al., 2013; Rawlinson & Spakman, 2016).

[†] See toward the end of Sect. 3.3.1.

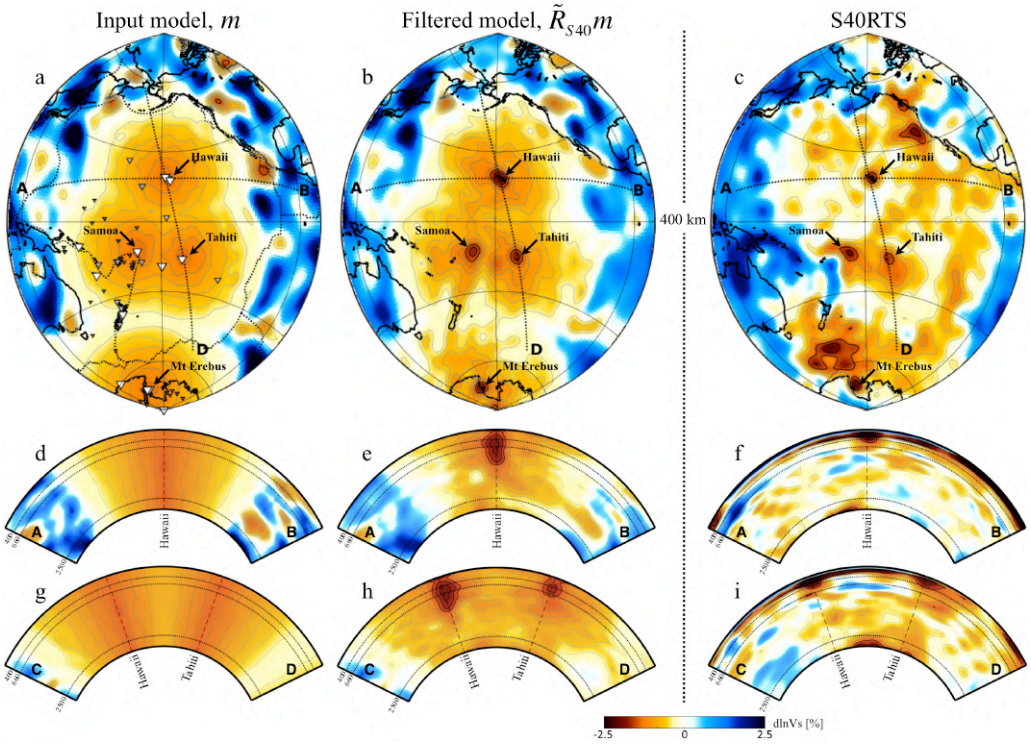


Figure 3.3.5: Quantitative illustration of the local averaging bias effect in model S40RTS. (Left) Input model, $\mathbf{m}$; (Middle) Filtered model, $\tilde{\mathbf{R}}_{S40}\mathbf{m}$; (Right) Model S40RTS. (a) Large-size-white (medium-size-gray, small-size-black) triangles denote all receivers located within the Pacific or Antarctica regions for which there is a large (moderate, small) number of recorded body-wave arrival times in S40RTS data set, respectively. Modified from Zaroli et al. (2017).

We consider model S40RTS, one of the few models for which the resolution matrix, here referred to as $\tilde{\mathbf{R}}_{S40}$, was fully calculated (Ritsema et al., 2011). S40RTS consists of isotropic, 3–D shear-wave velocity variations in the whole mantle, resulting from a joint DLS inversion of surface-wave, body-wave and normal-mode data. In particular, we aim to investigate how much and where S40RTS may be locally biased toward *higher* amplitudes. Such averaging bias effects may be expected to happen in mantle regions below isolated receivers

(e.g., located on oceanic islands), where body-wave ray paths should predominantly be unidirectional (vertical). Indeed, such anisotropic ray coverages are similar to that encountered in the previous 'toy' problem (for example, the bias effect occurring in the lower-right corner; see Figs. 3.2.16(a) and 3.3.3(b)).

Let us consider the input model $\mathbf{m}$ shown in Fig. 3.3.5(left), which consists of 3–D shear-wave velocity perturbations, spatially distributed like a Gaussian random field within the whole mantle, except in the Pacific and Antarctica regions, where four very long-wavelength low-velocity anomalies are superimposed (much longer than the local resolving lengths), centered in Hawaii, Tahiti, Samoa and Mt Erebus. These broad-extent, slow features are radially extending throughout the entire mantle, and laterally varying from dlnVs $\simeq -1.3\%$ at their center to zero in a linear fashion over 40 degrees great-circle distance, see Fig. 3.3.5(d). Some isolated (groups of) receivers, for which there are many recorded arrival times in the S40RTS data set, appear to be located at these four locations – see the large-size-white triangles in Fig. 3.3.5(a).

The filtered model $\tilde{\mathbf{R}}_{\text{S40}}\mathbf{m}$ is shown in Fig. 3.3.5 (middle). Quite remarkably, note the four *darker*-red anomalies showing up at all four locations, for instance at 400 km depth, see Fig. 3.3.5(b). Amplitudes have been locally enhanced, up to 50% higher values (dlnVs $\simeq -2\%$), thus creating some artificial, 'hotspot-like' features (!), see Figs. 3.3.5(e,h). Below Hawaii (Tahiti), this averaging bias effect is significant from the near-surface down to 1200 (700) km depth – amplitudes are increased by 10–50% (10–25%), respectively.

Model S40RTS is shown in Fig. 3.3.5(right) to be compared[†] with the filtered model $\tilde{\mathbf{R}}_{\text{S40}}\mathbf{m}$. Four low velocity features are visible in S40RTS, for instance at 400 km depth, below Hawaii, Tahiti, Samoa and Mt Erebus, see Fig. 3.3.5(c) – often referred to as 'hotspots' (e.g., Montelli et al., 2006). The apparent Hawaii and Tahiti hotspots, see Figs. 3.3.5(f,i), do overlap with the mantle regions where the previously identified averaging bias effect is significant, see Figs. 3.3.5(e,h), which then raises the following question:

- What is the actual imprint of the bias effect on these apparent hotspot features? Recall that the bias signature also depends on what the (input) true model is *and* how it spatially relates to the resolving kernels[††].
- Could this bias imprint be misleading in terms of physical interpretations? Could the apparent low shear-wave velocity structures, interpreted as the Hawaii and Tahiti hotspots, to some extent be questioned?

Although there are other geophysical, geochemical or geological evidences favoring such hotspot features in the mantle (e.g., Courtillot et al., 2003), these results should at least be a 'quantitative reminder' that:

Summary. The appraisal of DLS tomographic images should definitely be more quantitative, to avoid physical misinterpretations (if any); in particular, a thorough quantification of potential *averaging bias effects* is crucial.

[†] One should bear in mind that S40RTS is the sum of the filtered (unknown) true-mantle model and the propagated (unknown) data noise.

[††] Artificial structures that may potentially appear in DLS filtered models, $\mathbf{Rm}$, owing to some averaging bias effect, clearly depend on $\mathbf{R}$ *and* $\mathbf{m}$, see Fig. 3.3.3.

- **Computational efficiency**. We aim to compare the theoretical, computational efficiency of the (discrete) SOLA and DLS inversion methods, when it comes to calculate the full generalized inverse matrix – required to infer the model estimate *and* the resolution and uncertainty (model appraisal).

  - As previously mentioned, the DLS generalized-inverse matrix , $\mathbf{G}^\dagger$, can be expressed as: $\mathbf{G}^\dagger = (\mathbf{G}^T\mathbf{G} + \Theta^2\mathbf{I}_M)^{-1}\mathbf{G}^T$. It can be calculated from the eigen decomposition of $\mathbf{G}^T\mathbf{G}$. However, since $\mathbf{G}^T\mathbf{G}$ is much less sparse than $\mathbf{G}$, it may be too large to be diagonalized or even to fit in computer memory for large-scale applications. Thus, the generalized inverse has not been (could not be) computed in most DLS tomographic experiments.
  - It is extremely parallel to compute the SOLA generalized inverse, $\mathbf{G}^\dagger$, since each row is obtained independently from the others (see Sect. 3.2.3 and Fig. 3.2.17). Furthermore, as detailed in Zaroli (2016), computing the k-th row of $\mathbf{G}^\dagger$ only involves one LSQR inversion of a unique (i.e., independent from index k) matrix $\mathbf{Q}^{(\eta)}$, of size $(M+1)\times(N-1)$, that is almost as sparse as $\mathbf{G}$ – so that $\mathbf{Q}^{(\eta)}$ can easily fit in computer memory, and its sparsity still be exploited by the LSQR algorithm.

  Summary. Depending on the computational facilities (e.g., number of processors), the SOLA approach may be (much) more efficient than the traditional DLS inversion scheme, for computing a full generalized-inverse matrix.

  Remark. Another, computationally cheaper, way to estimate the DLS model solution, $\hat{\mathbf{m}}$, is to directly solve the DLS minimization problem (3.3.5), for example using the LSQR algorithm. When dealing with too 'large' problems and 'limited' computational facilities, this strategy saves considerable computational efforts, but does not allow one to have direct access to the DLS model resolution and uncertainty – since the generalized inverse is not computed. Note that such a shortcut, toward estimating only the model solution, $\hat{\mathbf{m}}$, does not exist in the SOLA (Backus–Gilbert) approach, since one is forced to directly solve for the generalized inverse to infer, all at once, the model solution *and* the corresponding (crucial) resolution and uncertainty informations.

- **Example**. I aim to briefly present a visual comparison of 'real-data' based, DLS *vs* (discrete) SOLA models – corresponding to the aforementioned, global-scale, finite-frequency S-wave tomographic study (Zaroli, 2016). Let DLS-Z16 and SOLA-Z16 denote the DLS and SOLA models, respectively; note that only the inversion approach differs. DLS-Z16 and SOLA-Z16 are shown in Figs. 3.3.6(a, b) within the transition zone, at 530–660 km depth. Though the models look similar, significative differences can be spotted. As an example, the reader may focus on the complex system of subducted slabs (positive anomalies) related to the subduction regions of New-Hebrides-Fiji, Tonga-Kermadec and Hikurangi. DLS-Z16 and SOLA-Z16 are shown in Figs. 3.3.6(c, d) at 1310–1510 km depth, focussing on the Farallon subduction system. In

general, one sees that tomographic features appear to be 'smoother or better focalized' and with 'higher amplitudes' in SOLA-Z16 (compared to DLS-Z16).

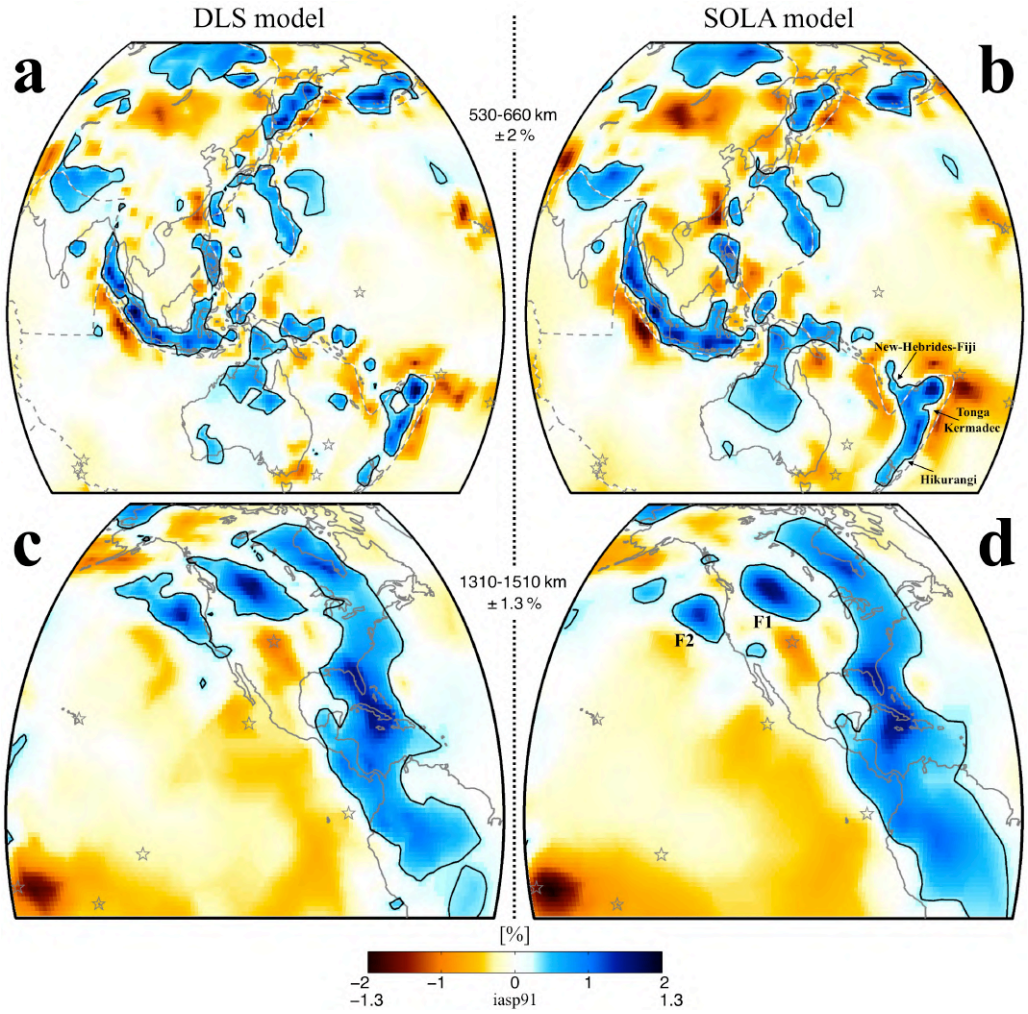


Figure 3.3.6: Comparison of two, global-scale, dV_S/dV_S tomographic models: (a, c) DLS-Z16 (DLS inversion) and (b, d) SOLA-Z16 ('discrete' SOLA inversion). Focus is set on (a, b) several active subduction zones, at 530–660 km depth, and on (c, d) the ancient and complex Farallon subduction system, taking place at 1310–1510 km depth. Modified from Zaroli (2016).

3.3.2 Take-home messages

To conclude, Fig. 3.3.7 summarizes some important 'take-home' messages about the SOLA–Backus–Gilbert inversion scheme, that I have recently introduced and adapted to solve linear(ized) seismic tomography problems. In particular, SOLA is compared to the classical DLS approach – one of the most popular inversion method used in seismic tomography studies over the past decades (in one form or another).

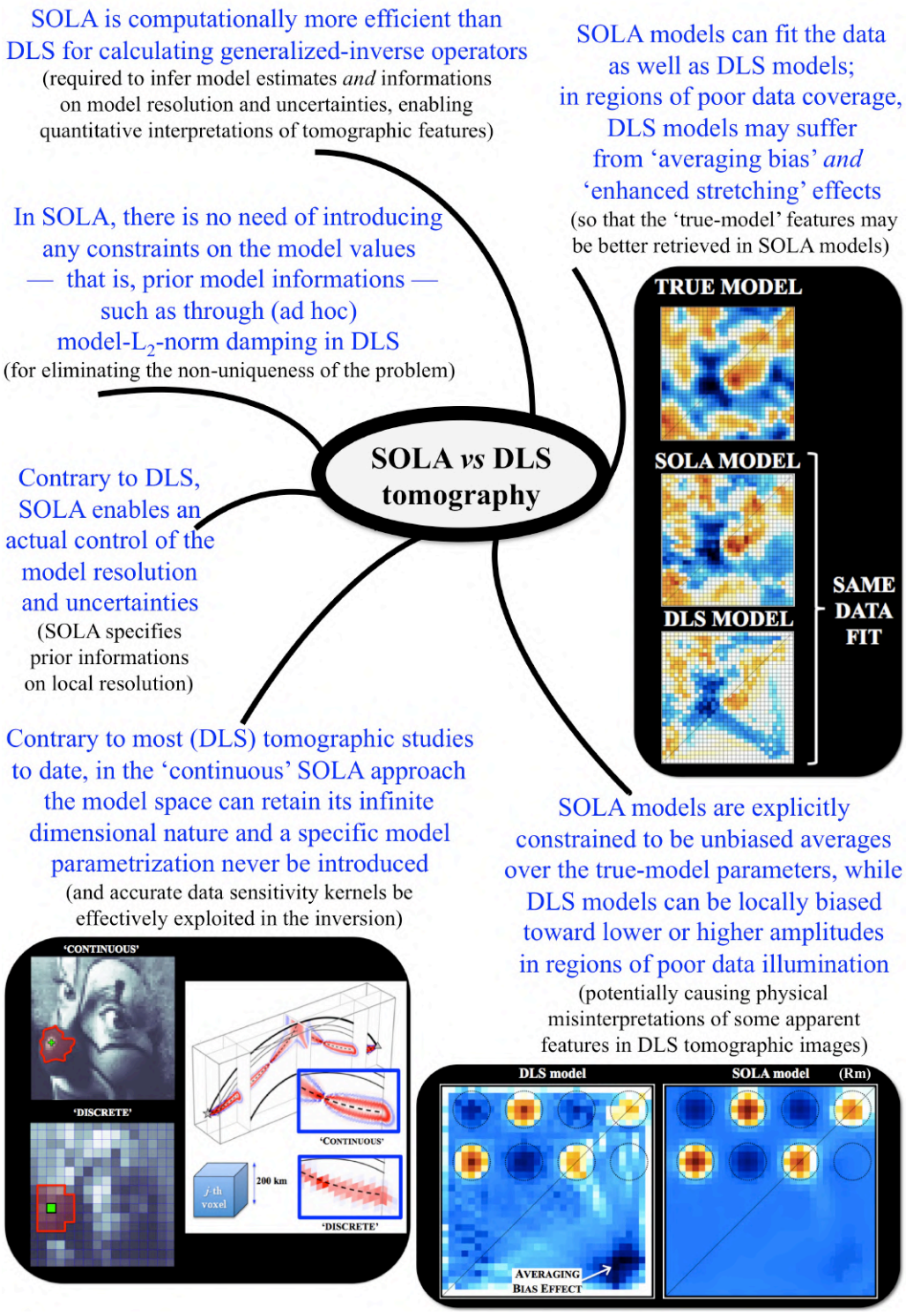


Figure 3.3.7: Key points of SOLA–Backus–Gilbert vs Damped Least-Squares inversion methods, for solving linear tomographic problems. Modified from Zaroli (2016, 2019) & Zaroli et al. (2017).

Afterword

In this chapter, we have seen what both *discrete* and *continuous* SOLA–Backus–Gilbert tomography is about (Zaroli, 2016, 2019), and why it should be preferred to the classical (discrete) DLS approach (Zaroli et al., 2017). In particular, we have explained why the *continuous* SOLA inversion approach may be expected to be more accurate, but also more costly (computationally), than the *discrete* SOLA formalism. Last, but not least, we have emphasized that all what is needed for to get started with *discrete* SOLA tomography is the same sensitivity matrix ($\mathbf{G}$) as required to perform DLS tomography, and an estimation of *prior* local model resolution (through the target kernels, $T^{(k)}$). Consequently, it should be 'straightforward' for everyone who is somehow experienced with DLS tomography to move on toward (at least) *discrete* SOLA tomography – and thus be among the first tomographers to be able to start experimenting with the half-century-old inversion philosophy of Backus & Gilbert !