Supporting Information for "Toward Seeing the Earth's Interior Through Unbiased Tomographic Lenses"

Christophe Zaroli[1], Paula Koelemeijer[2], Sophie Lambotte[1]

[1]Institut de Physique du Globe de Strasbourg, UMR 7516, Université de Strasbourg, EOST/CNRS, France
[2]Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, United Kingdom

Further Details on the SOLA Method

We aim to provide further details on the SOLA (Subtractive Optimally Localized Averages) method, recently introduced and adapted to large-scale, linear and discrete tomographic problems by *Zaroli* [2016]. In the SOLA approach, both the model estimation and its appraisal are computed all at once; that is, the generalized inverse matrix $\hat{\mathbf{G}}^{\dagger}$ is directly evaluated (see Table 1). When considering a local and 'orthonormal' model parametrization, as for the toy problem, the system to be solved for the k-th row of the generalized inverse, $\hat{\mathbf{G}}^{\dagger}_{k.} = (\hat{G}^{\dagger}_{ki})_{1\leq i\leq N}$, is as follows:

$$\left(\mathbf{G}\mathbf{G}^T + \eta_k^2 \mathbf{I}_N\right) \hat{\mathbf{G}}^{\dagger}_{k.} = \mathbf{G}\mathbf{t}^{(k)}\,, \quad \text{subject to} \quad \sum_{j=1}^{M}\sum_{i=1}^{N} \hat{G}^{\dagger}_{ki} G_{ij} = 1\,, \tag{1}$$

where η_k and $\mathbf{t}^{(k)} = (T_j^{(k)})_{1\leq j\leq M}$ are the k-th trade-off parameter and target resolving-kernel vector, respectively; k is the index of considered model parameter. Here all M target kernels are imposed to be unimodular: $\{\sum_j T_j^{(k)} = 1,\ \forall k\}$. For an example for some local and 'irregular' parametrization, the reader is referred to *Zaroli* [2016].

As mentioned in Sect. 2.4.2, and similarly to *Zaroli* [2016], we report that 'globally coherent' tomographic images (see Fig. 2(third column)) can be obtained when using: 1) Target kernels whose size is tuned to the spatially irregular data coverage; *and* 2) Constant-valued trade-off parameters, $\{\eta_k = \eta,\ \forall k\}$. Note that 2) prevents us from having to compute all M trade-off curves (resolution misfit *versus* model variance), to select every η_k based on the 'shape' of such trade-off curves – hence significantly increasing the total computational cost. In practice, η may roughly be determined from analyzing a few trade-off curves, for randomly chosen pixels.

Let us now define the following quantities, using same notations as *Zaroli* [2016]:

$$\begin{cases} \mathbf{x}^{(k)} &= \left(x_i^{(k)}\right)_{1\leq i\leq N}\,, & x_i^{(k)} = \hat{G}^{\dagger}_{ki} \\ \hat{\mathbf{x}}^{(k)} &= \left(x_i^{(k)}\right)_{2\leq i\leq N} & \\ \mathbf{c} &= (c_i)_{1\leq i\leq N}\,, & c_i = \sum_{j=1}^{M} G_{ij} \\ \hat{\mathbf{c}} &= (c_i/c_1)_{2\leq i\leq N} & \\ \mathbf{e}_1 &= (\delta_{i1})_{1\leq i\leq N} & \\ \mathbf{B} &= \begin{pmatrix} -\hat{\mathbf{c}}^T \\ \mathbf{I}_{N-1} \end{pmatrix} & \\ \mathbf{Q}^{(\eta)} &= \begin{pmatrix} \mathbf{G}^T\mathbf{B} \\ -\eta\hat{\mathbf{c}}^T \end{pmatrix} & \\ \mathbf{y}^{(k,\eta)} &= \begin{pmatrix} \mathbf{t}^{(k)} - c_1^{-1}\mathbf{G}^T\mathbf{e}_1 \\ -c_1^{-1}\eta \end{pmatrix}\,, & \end{cases} \tag{2}$$

Corresponding author: Christophe Zaroli, c.zaroli@unistra.fr

where c_1 is assumed to be non-zero and δ denotes the Kronecker symbol. Therefore, solving the system (1) consists in solving for $\hat{\mathbf{x}}^{(k)}$ the following set of normal equations:

$$\begin{pmatrix} \mathbf{Q}^{(\eta)} \\ \eta \mathbf{I}_{N-1} \end{pmatrix} \hat{\mathbf{x}}^{(k)} = \begin{pmatrix} \mathbf{y}^{(k,\eta)} \\ \mathbf{0}_{N-1} \end{pmatrix} , \tag{3}$$

using for instance the LSQR algorithm [*Paige and Saunders*, 1982], and then to infer the final solution $\mathbf{x}^{(k)}$, i.e., the k-th row of the generalized inverse, from $\hat{\mathbf{x}}^{(k)}$:

$$\mathbf{x}^{(k)} = \mathbf{B}\hat{\mathbf{x}}^{(k)} + c_1^{-1}\mathbf{e}_1 . \tag{4}$$

In terms of numerical considerations, since the rows of the SOLA generalized inverse matrix can be computed in *parallel* on P processors, computing all M rows takes $t \times M/P$ CPU-time, where t is the average CPU-time to numerically solve (3) – provided that some constant-valued η holds. A crucial point is that the matrix $\mathbf{Q}^{(\eta)}$, of size $(M+1) \times (N-1)$, does *not* depend on index k, so that there is no need to recompute it M times (as in the original Backus–Gilbert approach, see *Zaroli* [2016]). Although the vector $\mathbf{y}^{(k,\eta)}$ needs to be recomputed for every index k (i.e., M times), that is computationally cheap. As a remark, only the last row of $\mathbf{Q}^{(\eta)}$ and last element of $\mathbf{y}^{(k,\eta)}$ depend on η, so that $\mathbf{Q}^{(\eta)}$ and $\mathbf{y}^{(k,\eta)}$ may easily be reconstructed for different η values. Furthermore, *Zaroli* [2016] shows that a simple mathematical trick (related to re-ordering the data such that the *first* row of $\mathbf{G}$ is the sparsest one) allows the matrix $\mathbf{Q}^{(\eta)}$ to be almost as sparse as the sensitivity matrix $\mathbf{G}$, which is very useful in terms of storage, efficiency of the LSQR algorithm, and memory footprint.

Finally, one possible artifact that SOLA could be prone to arises when resolving kernels appear to be very oscillatory (i.e., significantly negative). That is, even if they are constrained to be such that $\{\sum_j \hat{R}_{kj} = 1,\ \forall k\}$, they cannot anymore be considered as truly 'averaging' kernels. Thus, in this special case, SOLA models may not represent unbiased averages over the true-model parameters. To avoid such artifacts, following *Pijpers and Thompson* [1994], we progressively enlarge the size of target resolving-kernels and solve for the new minimization problems (keeping η unchanged) until: $\{\sum_j (\hat{R}_{kj} - T_j^{(k)})^2 < \xi^{\hat{R}},\ \forall k\}$. This allows us to ensure that all M resolving kernels are mostly well localized (close to the target kernels) and non-negative (non-oscillating). In practice, only a small fraction of the target kernels may have to be enlarged, depending on how robust is the *a priori* local model resolution used to generate the first generation of target kernels (first step in Sect. 2.4.2). Note that the target kernels may also be enlarged to keep the model uncertainties limited (trade-off between resolution misfit and model variance), that is: $\{\sigma_{\hat{m}_k} < \xi^{\sigma_{\hat{m}}},\ \forall k\}$. For the toy problem, we use: $(\xi^{\hat{R}} = 0.03,\ \xi^{\sigma_{\hat{m}}} = 0.3\%)$; these numerical values are chosen somewhat arbitrarily, tailored to our experiment.

Variability of DLS and SOLA Solutions

Figures S1–S4 illustrate the variability of damped least-squares (DLS) and SOLA solutions (in terms of filtered true models, resolving kernels and propagated noise), when changing some of the tunable inversion parameters – that is, the damping Θ and trade-off parameter η (keeping unchanged the target resolving-kernels, for simplicity reasons), respectively.

As a remark, when the data errors are poorly known, Θ is often somewhat subjectively chosen based on the trade-off curve between data misfit and model norm. Note that the DLS averaging bias effect is expected to vary as a function of Θ; it should vanish for $\Theta = 0$. Although there is also some subjectivity in the SOLA approach, when choosing the target kernels and trade-off parameters, different choices would simply

result in different, *unbiased* model estimates and appraisals, thus corresponding to different, *robust* model interpretations.

References

Paige, C., and M. Saunders (1982), LSQR: an algorithm for sparse, linear equations and sparse least squares, *A.C.M. Trans. Math. Softw.*, *8*, 43–71.

Pijpers, F., and M. Thompson (1994), The SOLA method for helioseismic inversion, *Astron. Astrophys.*, *281*, 231–240.

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

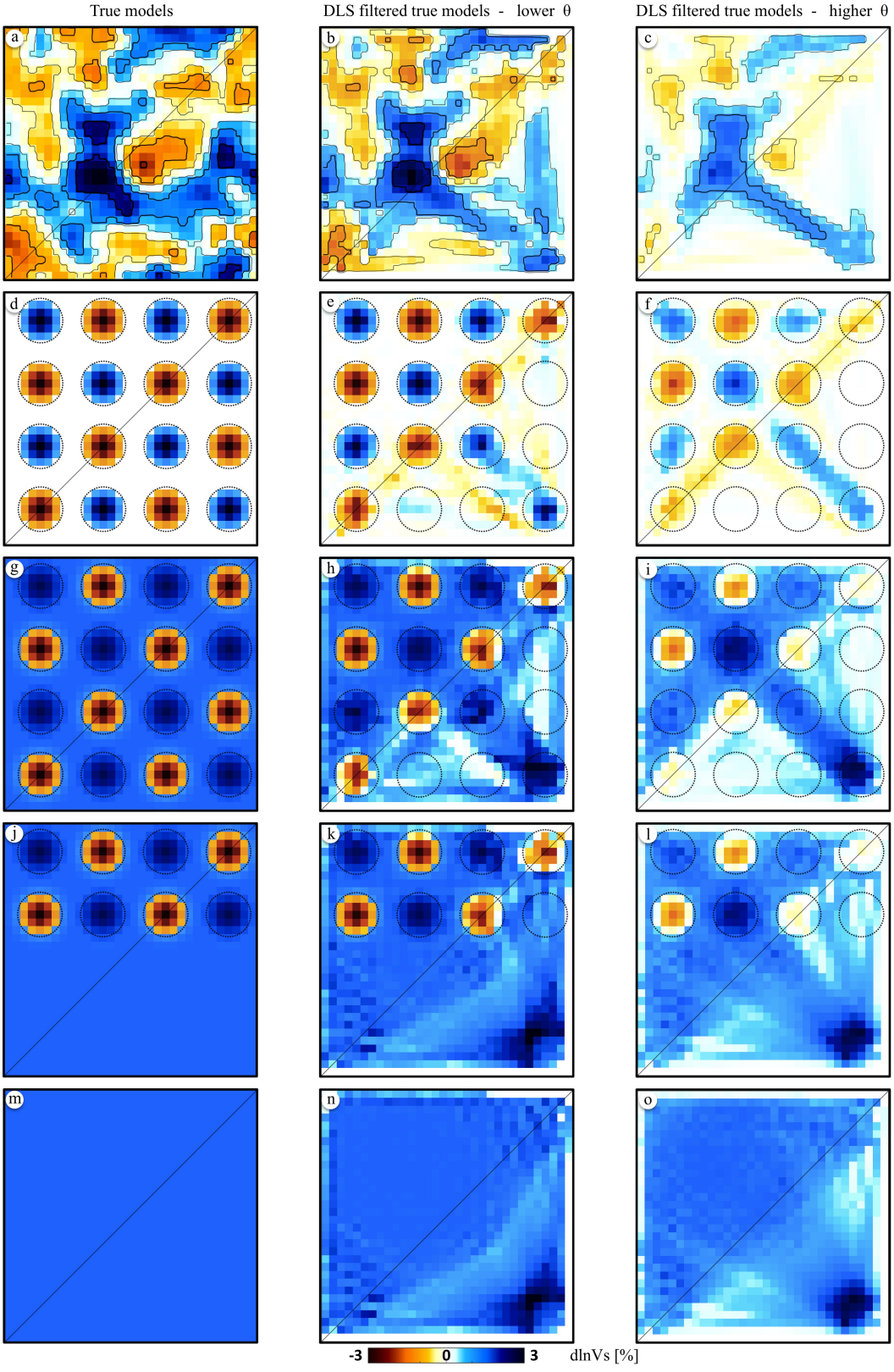

Figure 1. Variability of DLS solutions – (First column) True models; (Second and Third columns) Filtered true models for some lower ($\Theta^{low} = 0.4 \times \Theta$) and higher ($\Theta^{high} = 1.9 \times \Theta$) damping value than used (Θ) in Fig. 2(second column), respectively.

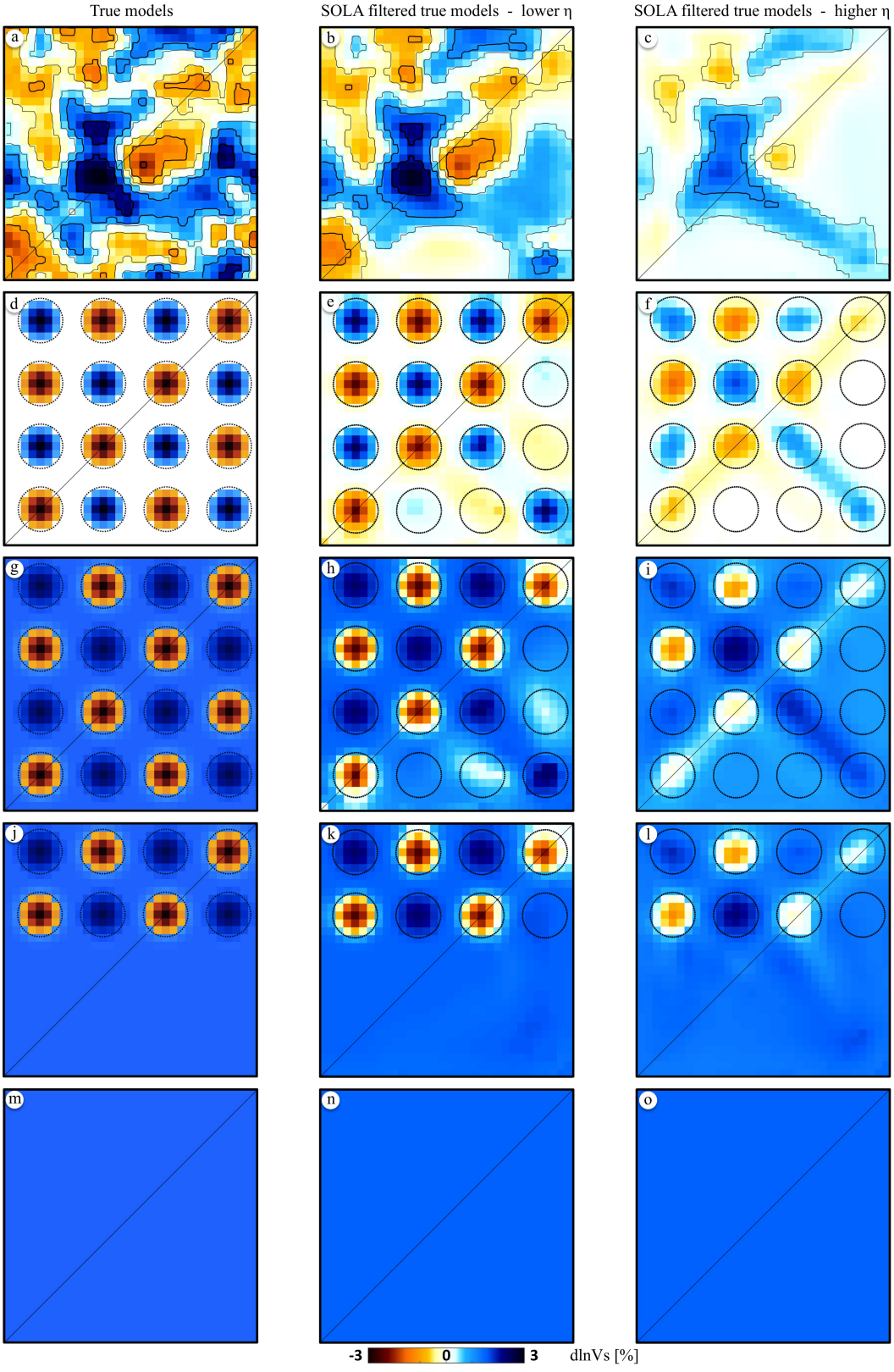

Figure 2. Variability of SOLA solutions – (First column) True models; (Second and Third columns) Filtered true models for some lower ($\eta^{\text{low}} = 0.3 \times \eta$) and higher ($\eta^{\text{high}} = 6.0 \times \eta$) trade-off parameter value than used (η) in Fig. 2(third column), respectively.

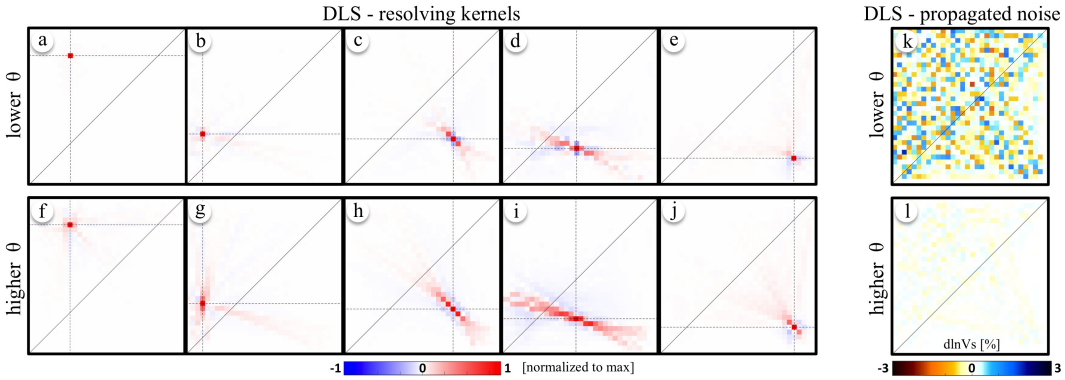

Figure 3. Variability of DLS solutions – (a–e, f–j) Resolving kernels for Θ^{low} and Θ^{high}, respectively; (k, l) Propagated data errors for Θ^{low} and Θ^{high}, respectively.

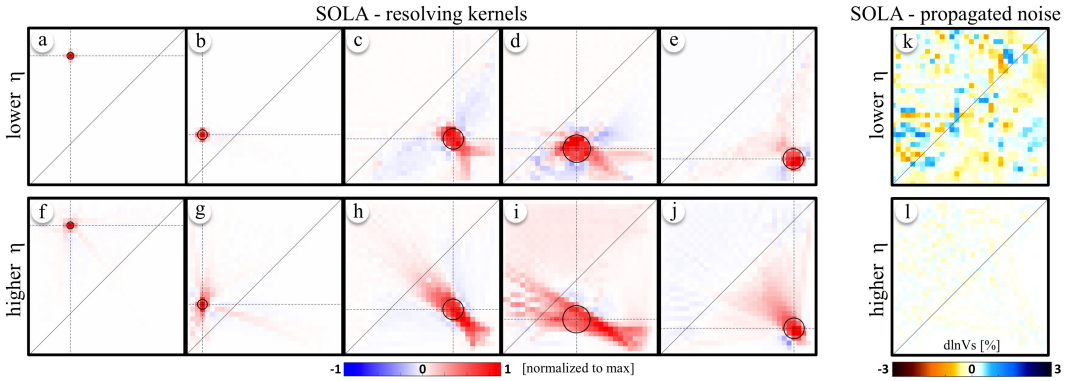

Figure 4. Variability of SOLA solutions – (a–e, f–j) Resolving kernels for η^{low} and η^{high}, respectively; (k, l) Propagated data errors for η^{low} and η^{high}, respectively. All target resolving-kernels are kept unchanged as in Fig. 1(b).