Christophe Zaroli
EOST
5 rue René Descartes
67084 Strasbourg cedex
E-mail : c.zaroli@unistra.fr

August 29, 2013

Authors Comments : research article se-2013-23

"An objective rationale for the choice of regularisation parameter with application to global multiple-frequency S-wave tomography"

C. Zaroli, M. Sambridge, J.-J. Lévêque, E. Debayle, and G. Nolet

A point by point response to the two reviewers comments is given below. We believe the revised manuscript has benefitted from all the comments received, and hope it is now suitable for publication in *Solid Earth*.

---

1 Reviewer 1 (*Masayuki Obayashi*)

1.1 *Comment*/Reply

In this paper, the authors present an objective rationale for the choice of a priori damping parameter used in global multiple-frequency tomography. I understand the individual argument to get the optimal range of damping in sections 3.1 and 3.2.

OK.

1.2 *Comment*/Reply

However it sounds complicated for me. I do not know why they did not adopt the simple cross validation rather than their criteria. I mean testing how well a dataset sampled randomly from all dataset is fitted by the model obtained from the remainder of the dataset. If there is any reason to avoid such simple method I would like to hear it.

Cross validation (CV) or generalized cross validation (GCV) is a well studied subject in statistics, for which there are many variants, particularly in areas such as regression problems. It is certainly simple, however numerous studies in statistics have pointed out the limitations of this generalised approach, e.g. Rivals & Personnaz (1999). These include that the tendency for resulting solutions is to favour regularisation over data fitting (i.e. producing overly smooth models) and that for model selection problems CV performs poorly compared to standard statistical tests. In addition there is no guarantee that a well defined minimum in the CV cost function will exist, nor is there any objective rationale for choosing the number of datasets to divide the data into, in a k-fold cross validation. CV will therefore not always provide satisfactory answers, nor is it clear how more effectively implement it in any particular problem (CV would likely involve prohibitive computational effort to fully implement on our very large dataset).

In this study, we present an objective rationale for the determination of under-damped models (or "dominated by noise"), whose philosophy can be viewed as analogous to CV (cf. section 3.2.1, where we say that : "*The philosophy of this approach is somewhat analogous to the CV method, in that we aim at estimating the fit of a model to a data set that is similar but not identical to the data that were used to derive the model.*"). However, the similarity/analogy of our approach with CV stops there. A crucial step in CV consists in randomly partitioning the original data set into k equal size subsets. Thus, the fundamental difference with our approach comes from the fact that we exploit the multiple-frequency character of the data set, by splitting it into five single-frequency data subsets ($\mathbf{d}_{10}$, $\mathbf{d}_{15}$, $\mathbf{d}_{22}$, $\mathbf{d}_{34}$, $\mathbf{d}_{51}$) with similar sensitivity to the mantle. Hence our data partitioning is based on physics, rather than randomness, which we argue is more meaningful in this case. In particular, by partitioning the data by frequency we exploit the lack of expected coherency of the data in different frequency bands, which is the key (cf. section 3.2).

The presented rationale offers a problem specific alternative to CV. In order to clarify this point, we added the following sentences in the revised manuscript (beginning of section 3.2.1) : *"The philosophy of this approach is somewhat analogous to the cross-validation (CV) method, in that we aim at estimating the fit of a model ($\boldsymbol{SB}_{34}^{\lambda'}$) to a data set ($\boldsymbol{d}_{10}$) that is similar but not identical to the data ($\boldsymbol{d}_{34}$) that were used to derive the model. However, our data partitioning approach markedly differs from the randomly subdivision of data inherent in CV. Partition into frequency bands, based on similar sensitivities to mantle structure, seems to make more sense in the context of our multi-frequency data set."*

Reference : Rivals, I., Personnaz, L., 1999. *On Cross Validation for Model Selection*, Neural Computation, Vol. 11, No. 4, Pages 863-870.

1.3 *Comment*/Reply

And I suppose that the damping value close to their preferred one may be given by the misfit function of a single-band data subset with respect to multi-band model that obtained from all dataset except the single-band data subset. For instance, in case of 10 s data misfit, the model is constructed from the data except 10 s data and the damping value will be given by reversal point in the trade-off curve between $\|\mathbf{MB}^{\lambda}_{15,22,34,51}\|_2^2$ *and* $\chi^2_{red}(\mathbf{MB}^{\lambda}_{15,22,34,51}, \mathbf{d}_{10})$.

This comment is very interesting. However, it seems difficult to intuitively anticipate if doing this would lead to identify a larger or smaller range of under-damped multi-band models. In addition, fully testing this hypothesis would require too much computational efforts at the moment, and in any case it could not change the overall validity of the presented rationale for the choice of damping. Therefore, we postpone such calculations for future work.

1.4 *Comment*/Reply

If that is the case, I think they do not need the criterion in section 3.3 that I feel subjective as the authors also mention.

As mentioned in section 3.3, one cannot avoid some subjectivity concerning the choice of *one* particular solution among the optimal range of models. Hence, we do not present a criterion for *objectively* selecting *one* model among the permitted solutions. The goal of section 3.3 is more to present an efficient way to estimate if this remaining degree of subjectivity is reduced enough in terms of model interpretation (cf. beginning of section 3.3).

1.5 *Comment*/Reply

And if the damping values obtained above strongly depend on what period data are used for misfit calculation, it may reflect the noise level of each period.

Yes, we agree. As explained at the end of section 3.2.2, we only show our coherency analysis applied to single-band models derived from the 34 s data ($\mathbf{SB}_{34}^{\lambda'}$) and their fit to the 10 s data. However, though not shown here, we also computed their fit to the 15, 22, and 51 s data subsets. As expected, we found that the corresponding L-curve reversals were reached for smaller values of damping (λ') than if considering the 10 s data fit. We think that it reflects the noise level of each period, but also the degree of similarity of the corresponding sensitivity kernels.

In order to clarify this point, we added in the revised manuscript the following explanations (end of section 3.2.2) : *"We analysed single-band models $\boldsymbol{SB}_{34}^{\lambda'}$ (or $\boldsymbol{SB}_{22}^{\lambda'}$) with their fit to the 10, 15, 22 (or 34), and 51 s data subsets, respectively. Corresponding L-curve reversals were reached for different values of damping (λ'), which mainly reflects, at each period, the data noise level and the sensitivity kernel. We observed that the analysis of models derived from the 34 s data, and their fit to the 10 s data, lead to identify the narrowest range of acceptable multi-band models."*

1.6 *Comment*/Reply

I do not have a problem with their statement about their tomographic result, because this paper does not address to interpretation of the obtained detail structure.

OK.

2 Reviewer 2 (*Kazunori Yoshizawa*)

2.1 *Comment*/Reply

This is a very interesting paper that treats one of the fundamental issues in seismic tomography; pursuing an objective way for the determination of reasonable damping parameter (and/or *its range). The authors' attempt for finding a reasonable model in the case of multiple-frequency shear wave tomography is intriguing, and the similar idea can be useful for other types of tomographic problems working with multiple frequency models that are expected to have some correlations among them ; e.g., surface wave tomography based on multiple-frequency phase/group speed models. This paper is well written and I am convinced of the way that the authors have proposed for the determination of the exclusive range of both over-damped and under-damped models (or the range for models with "poor data exploitation" and that "dominated by noise").*

OK.

2.2 *Comment*/Reply

The final determination for the best compromised model from the limited range of the trade-off curve seem to be somewhat subjective, as the authors have mentioned, so that the method proposed here may better be represented as "quasi-objective".

The method proposed here leads to a "quasi-objective" model solution in our case (cf. "*Comment*/Reply §1.4"). However, the presented rationale, for determining the optimal range of models (i.e. neither dominated by noise nor with poor data exploitation), is fully objective.

2.3 *Comment*/Reply

Though it may not be fully objective since the final model is selected with an arbitrary criterion based on the correlations of models derived from slightly different damping (over 98 % of correlation is achieved within the optimal range of tradeoff curve, in this case), such slight subjectivity is unavoidable in inverse problems and the authors' criteria seems to be practically useful for the automatic determination of the final model. I have only a few minor comments as summarized below, and I recommend the paper be published in Solid Earth with some minor corrections.

OK.

2.4 *Comment*/Reply

Page 852-854, Section 3.2 : I understand the reason why the authors have chosen 34 s as the reference period (as explained in P854, L8-16). Still, it will be worth adding statements on how the results could be affected if, for example, they choose 22 s as the reference, or use d_{15} rather than d_{10} in Fig 3. It should be briefly explained in the main text or in an appendix, or in a form of electronic supplement. (It seems to me that the differences in the wavelength of different period, e.g., 22 and 34 s can affect the scale-length of resolvable heterogeneity, which may have some influences on both axes of the tradeoff curve in Fig 3.)

In order to clarify this point, we added some explanations in the main text of the revised manuscript. See "*Comment*/Reply §1.5".

2.5 *Comment*/Reply

Page 857 line 20-29, explanations for Fig7 : Explanations and interpretation for model differences (Fig.7) in section 4 are too brief. I think PRI-S05 is derived from the same finite-frequency (FF) technique as used in this paper, but the other two (S40RTS and TX2007) are based on ray theory (RT). I believe that this is one of the reasons why FF models (this study and PRI-S05) in Fig 7 tend to show apparently larger velocity perturbations than RT models (S40RTS, TX2007), since the effects of diffractive healing on travel times (which tend to underestimate the velocity perturbations in the framework of ray theory) are naturally taken into account through the banana-doughnut kernels used in the FF models. Such additional explanations are likely to be useful for readers, since the differences in velocity perturbations among these models can also be caused by the subjective choice of damping.

In order to clarify this point, we added in the revised manuscript the following explanations (end of section 4) : *"S40RTS results from Rayleigh wave phase velocity, teleseismic shear-wave traveltime and normal mode splitting function measurements, and is parameterised by spherical harmonics up to degree 40 and by 21 vertical spline functions. PRI-S05 and* $\boldsymbol{MB}^{pref}$ *result from cross-correlation shear-wave delay-times measured in one or several frequency band(s), respectively. They are irregularly parameterised according to data coverage. TX2007 results from a joint inversion of seismic data (teleseismic shear-wave traveltimes) and geodynamic constraints (global free-air gravity, tectonic plate divergence, excess ellipticity of crust-mantle boundary, dynamic surface topography), and is parameterised with an irregular grid. S40RTS and TX2007 are based on RT, while PRI-S05 and* $\boldsymbol{MB}^{pref}$ *rely on BDT. [...] PRI-S05 and* $\boldsymbol{MB}^{pref}$ *tend to show slightly larger amplitudes, in mid lower-mantle, than S40RTS and TX2007. These model differences may partly be explained by the theoretical frameworks (e.g. BDT can compensate for wavefront-healing effect but RT cannot), or by the different nature of the data sets. However, [...] "*