# Response to Referee 1's comments concerning manuscript se-2014-82

P. Tong, D. Zhao, D. H. Yang, X. Yang, J. Chen and Q. Liu

October 22, 2014

We have very carefully revised our manuscript 'Wave-equation based traveltime seismic tomography - Part 1: Method' [MS No. se-2014-82], taking into account ALL of the comments and suggestions by Dr. Yang Luo, Referee 1. We are grateful for the comments and suggestions which have been extremely helpful in further improving the paper. Below, we have provided detailed point-by-point replies to the comments, highlighted in blue colour. Meanwhile, our revisions are also highlighted in blue in the updated manuscript.

We look forward to hearing from you (Dr. Yang Luo) on the revised manuscript. Thank you very much.

## Responses to Referee 1

*Review report*

*In this manuscript, authors discuss about traveltime seismic tomography based on wavefield modelling and adjoint method. Compared to traditional traveltime tomography, the new technique, WETST, enables seismologists to examine seismograms in greater details allowing wiggle-to-wiggle comparisons at appropriate frequency bands, which in turn provides a lot more information on Earth's interior. Because WETST is similar and closely related to Reverse Time Migration and Full Waveform Inversion in oil industry, there are opportunities for both community to develop fast, by sharing experience with each other. However, due to its demanding computational requirements, WETST is not feasible at all until very recently, after advances in hardware dramatically improve our computational capability.*

*This manuscript is a great summary on the theory of adjoint method, with a practical step-by-step workflow. The main new contributions are two parts. First, authors propose a new algorithm to measure traveltime differences automatically, in additional to the method using envelop short-term long-term ratio. The new algorithm measures an initial traveltime using ray tracing, and then updates it subsequently in each iteration via cross-correlation between synthetic seismograms from two models. The advantage of this algorithm is obvious: the tedious measurements are done once and for all at the beginning of the inversion, with corrections easily obtained with high accuracy iteratively. Second, authors describe a 2D-3D approach to further reduce the computational burden. In short, the inverse problem is established in 3D, but numerical simulations are constrained to a 2D vertical plane that passes through both source and receiver. Effectively, only 2D wave propagation is needed, which is much faster than 3D propagation. Both proposals are theoretically correct and are good approximations we may adapt, when efficiency is a bigger issue.*

**Reply**: Thank you very much for your kind comments.

***Comment 1****: On the new traveltime measurement: the initial traveltime calculated via ray tracing seems critical. If it has some nontrivial discrepancies, all subsequent model updates will be systematically biased. Further, while cross-correlation measurements can be frequency-dependent, ray-based measurements are asymptotic high frequency solutions. The new algorithm may not work for dispersive scenarios.*

**Reply**: Thanks for the critical comments. We have added some discussion on the data processing and selection for the wave-equation based traveltime seismic tomography (WETST) in the last paragraph of Section 2 and the second paragraph of Section 7. We suggest manually pick $T^{obs}$ on observed data that have been bandpass filtered to the frequency range consistent with the frequency spectrum of synthetic seismograms. That is to say, the traveltime residuals $T^{obs} - T^{syn}$ used by WETST are finite-frequency traveltime residuals. Admittedly, ray-based measurements are asymptotic high frequency solutions, and may not work for dispersive scenarios. To reduce the influence of ray-based measurements on the calculation of $T^{syn}$, we suggest choosing 1D horizontally layered velocity model in which we conduct ray tracing and then calculate the theoretical traveltime of the specific phase therein. The dispersive effects are not significant and can be ignored in this simple layered homogeneous model. The traveltime of seismic waves in complex (iteratively updated, heterogeneous) velocity models can then be obtained by adding on the cross-correlation traveltime measurement, which naturally takes into account the finite-frequency and dispersive effects.

***Comment 2****: On the 2D-3D approach: is there a guideline on how to choose the inversion grid and the simulation grid? It is not intuitive as it might look. Also, it may be good to have some toy synthetic example to illustrate this approach provides acceptable inversion result, i.e., the 3D effects are not severe.*

**Reply**: Thanks for the good question. To determine the size of the inversion grid, we should first consider the theoretical resolving ability of the inversion method. For the discussed wave-equation based traveltime seismic tomography, its resolving ability is at the scale of $\sqrt{\lambda L}$ ($\lambda$ is the wavelength of the propagating seismic waves and $L$ is the travelling distance). Therefore, the grid spacing of the inversion grid should be related to $\sqrt{\lambda L}$. However, in real applications, the resolution of the inversion results also relies on the amount and distribution of seismic data. An even dense coverage of seismic data greatly helps improve the resolution. Checkerboard resolution test is a good approach to measure the resolving ability of an inverse algorithm with given seismic data and inversion grid. We did not include synthetic test results of WETST in this manuscript (Paper I), but in our manuscript se-2014-83 (Paper II, regarding applications), we use the checkerboard resolution tests to judge whether WETST can well resolve subsurface heterogeneities at the scale of the inversion grid size with the selected seismic data. But due to the demanding computational cost, we did not test the lower limit of the inversion grid size in this study. This should be investigated in future studies. The detailed procedure for searching the optimal damping parameter and the analysis of data and model variance are also presented in Paper II.

***Comment 3****: On the 2D-3D approach: is there some rough estimation on how much faster an inversion will be using the 2D-3D approach, compared to the 3D-3D approach? It might depend on the problem itself. For example, although 3D simulations run much slower, only one set of simulations is required for each earthquake, whereas in the 2D-3D approach, multiple simulations are needed for one earthquake, depending on the number of stations. Worst case, we need one set of simulations for each source-receiver pair, which may not be a small number. Or alternatively, as an approximation, azimuthal bins can be formed for each earthquake and the computational cost is proportional to the number of azimuthal bins. But based on the new traveltime measurement algorithm, it seems to me that the actual case is even worse - authors may have ran simulations for each measured phase of each source-receiver pair. I'm not against this route (which is purely what I guess authors have used, but my guess might be completely wrong), but because of these trade-offs, it is not straightforward to me that the 2D-3D approach is less computationally intensive. Its better if authors can provide some estimated numbers to approve that statement. But even if the 2D-3D approach is not significantly faster than the 3D-3D approach, there are benefits from it, which I will mention later.*

**Reply**: Thank you very much for this question and suggestion. Yes, how much faster is the 2D-3D approach compared to the 3D-3D approach depends on the problem itself. Here we would like to give a rough comparison. For the chosen earthquakes and stations in paper II, if we use the 3D version of the high-order central difference method as presented in Appendix A of this manuscript, the computation time for simulating the propagation of seismic waves generated by one earthquake is averagely about 500 times that of the 2D method. In the chosen data set, each earthquake provides on average 38 *P*-arrivals and 23 *S*-arrivals. It can then be estimated that generating individual traveltime kernel based on 2D forward modelling method is about 12-20 times faster than generating an event kernel based on the 3D method. Besides that, an inversion algorithm using event kernels (such as the non-linear conjugate-gradient method) generally requires more iterations than that based on individual kernels, as individual kernels provide the Hessian matrix for the inverse problem and allow the use of efficient solvers such as LSQR. For example, Tape et al. (2009) iteratively updated 16 times of the crustal structure of southern California from a 3D heterogeneous starting model. Zhu et al. (2012) iteratively updated the structure of the European upper mantle for 30 times. In the application manuscript se-2014-83, we have obtained a reasonably good final model in three iterations. Based on these rough comparisons, we may cautiously conclude that the 2D-3D approach is about 50 times faster than the 3D-3D approach for regional seismic tomography study. In the revised Paper II, we have added a rough comparison between the 2D-3D and 3D-3D tomography methods for the specific study (please see the third paragraph of Section 7 of Paper I and the last paragraph of Section 5 of Paper II for more discussions). An additional advantage of the 2D-3D approach is that the 2D forward modelling uses much smaller storage space, and allows for modelling and inversions of high-frequency seismic data that may not be modelled by 3D forward modelling technique on current standard cluster.

***Comment 4****: In terms of the inversion method, it may be good to include the L-BFGS algorithm, which is easy to implement and proved to be superior. In cases where we do have luxury resources, i.e., individual traveltime sensitivity kernels, the so-called LSQR solver in the manuscript is the ultimate solution, which is essentially Gauss-Newton. Conjugate gradient, L-BFGS and other variants are all approximations to the Newton or Gauss-Newton solution.*

**Reply**: Thanks for the comments and suggestion. Yes, we have learned that the L-BFGS method is particularly well suited for optimization problems with a large number of variables. We have given a brief description of this method in the last paragraph of Section 5 in the revised manuscript. We believe that Dr. Yang Luo's PhD thesis (Princeton University, 2012) is a good summary of different inversion methods, including the L-BFGS algorithm and a new square root variable metric (SRVM) algorithm adapted from adjoint tomography. Therefore, for the sake of compactness, we cite this thesis and the paper "Updating quasi-Newton matrices with limited storage" (Nocedal 1980) for the details on the L-BFGS method.

***Comment 5****: Having individual traveltime sensitivity kernels not only enables us to use Gauss-Newton method to solve the inverse problem, but also helps to address post-inversion resolution analysis. As long as we have the Jacobian matrix, it is straightforward to construct the resolution matrix in both model and data spaces, using exactly the same procedure practiced in traditional traveltime tomography. Since we do have comparisons between simulated and observed seismograms, the data space resolution matrix is of less interest. But model space resolution matrix tells us the trade-offs between model parameters in the inversion, which reveals uncertainty of the inversion, if any.*

**Reply**: Thanks for the comments. In the revised manuscript, we have addressed the point that the computation of individual kernels enables us to examine resolution in both data and model spaces. Please see the last paragraph of Section 7.

***Comment 6****: Page 2525 Line 14: 'To take into account ..........'*
*It seems to me this long long sentence is grammatically incomplete.*

**Reply**: Thanks for pointing out the mistake. We have rewritten this sentence.

***Comment 7****: Page 2525 Line 28: 'full numerical method'*
*Maybe 'fully numerical method'?*

**Reply**: Thank you. It is revised to 'fully numerical method'.

***Comment 8****: Page 2531 Equation 4, 5, 9, 31, one line above Equation 32*
*I don't like the partial derivative being written as* $\partial u(0,x)/\partial t$. *Prefer* $\frac{\partial u(0,x)}{\partial t}$. *Just my personal taste.*

**Reply**: Thanks for this suggestion. We use $\frac{\partial u(0,x)}{\partial t}$ instead in the revision.

***Comment 9****: Page 2533 Equation 12, 13, 24, 26*
*Most often,* $S$ *is used for the integral, but these are the cases where* $\Omega$ *is used.*

**Reply**: Thanks for this comment. Our forward modelling is conducted in a 2-D vertical plane, and therefore we preferred to use $S$ to emphasize that the integral region is a 'surface' instead of a 'volume' $\Omega$.

***Comment 10***: *Page 2537 Equation 17*
*A windowing function is missing in the second integral in the denominator.*

**Reply**: Thanks very much. We have added the window time function in Eq. (17).

***Comment 11***: *Page 2539 Equation 19*
*Maybe allowing different smooth lengths along $x$ and $z$ directions?*

**Reply**: Thanks for this great suggestion. We now redefine the 2-D Gaussian smoothing function as $G(x,z) = 4e^{-4x^2/\sigma_x^2 - 4z^2/\sigma_z^2}/(\pi\sigma_x\sigma_z)$, where $\sigma_x$ and $\sigma_z$ are the averaging scale lengths along $x$ and $z$ directions, respectively. Different smooth lengths are thus allowed. Please see the first paragraph of Section 4.1.

***Comment 12***: *Page 2542 Line 24*
*Entrywise product and summation afterwards?*

**Reply**: Thanks for this question. The operator $\circ$ denotes a two-step operation which first gets the entrywise product of two matrices and then sums up all the entries of the produced matrix. we have revised the description of this operator to make it more clear.

***Comment 13***: *Page 2545 Line 14*
*'Singular' value decomposition.*

**Reply**: Thanks. Revised.