Petascale Computing for Future Breakthroughs in Global Seismology

L. Boschi [a] J.-P. Ampuero [a] D. Peter [a] P. M. Mai [a] G. Soldati [b]
D. Giardini [a]

[a] *Institute of Geophysics, E.T.H. Hönggerberg-HPP, 8093 Zürich, Switzerland*

[b] *Istituto Nazionale di Geofisica e Vulcanologia, via di Vigna Murata 605, 00143 Roma, Italy*

Abstract

Will the advent of "petascale" computers be relevant to research in global seismic tomography? We illustrate here in detail two possible consequences of the expected leap in computing capability. First, being able to identify larger sets of differently regularized/parameterized solutions in shorter times will allow to evaluate their relative quality by more accurate statistical criteria than in the past. Second, it will become possible to compile large databases of sensitivity kernels, and update them efficiently in a non-linear inversion while iterating towards an optimal solution. We quantify the expected computational cost of the above endeavors, as a function of model resolution, and of the highest considered seismic-wave frequency.

Key words: tomography, finite-frequency, resolution, computational seismology, inverse theory

1 Introduction

In the past decade there have been a number of claims by tomographers that lead to a re-consideration of certain aspects of the theory of the Earth. Those claims have been justified by improvements in tomographic resolution. van der Hilst *et al.* (1997), for example, inverted a very large database on an unprecedentedly dense voxel grid, making use of an inversion algorithm that exploited the inherent sparsity of the linear inverse problem. They found very sharp images of fast, deep heterogeneities, that, because of their geographic distribution, were explained in terms of subducted material, sinking into the lower mantle. This finding, while subject of debate (is "resolution" really as high as claimed?), has been a strong argument in favour of whole-mantle vs.

Preprint submitted to Elsevier *11 January 2007*

layered convection. More recent examples are the work of Ishii and Dziewonski (2003), who mapped an "innermost inner core" of only 300 km in radius, suggesting that it be "the oldest fossil left from the formation of Earth", and the controversial article of Montelli *et al.* (2004), who improved global resolution by means of a more accurate approach to the calculation of sensitivity functions, and found "clear evidence that a limited number of hotspots are fed by plumes originating in the lower mantle"; this claim is clearly relevant to the current debate on the nature of mantle plumes (e.g., Sleep, 2006), involving all disciplines in the Earth sciences, and stirred a very animated debate.

The controversy originated by these publications, and, in general, the lack of correlation at short spatial wavelengths between tomographic images derived in different approaches (e.g., Becker and Boschi, 2002) indicate that the next important challenge in global seismic tomography is that of finding effective ways to improve the images' resolution. We describe in the following the role that high-performance computing might play, in reference to developments in tomographic imaging and the subsequent interpretation of mapped Earth structure.

2 Limiting factors of tomographic resolution

Tomographic resolution, or the smallest lateral extent of a velocity anomaly that can be correctly mapped by an inversion algorithm, is limited by (i) the geographic coverage of inverted seismic observations, (ii) the resolving power of the selected parameterization, (iii) the accuracy of the theoretical formulation, or the equation relating seismic data to the velocity field. The latter problem has been explored, for example, in the recent works of Montelli *et al.* (2004, 2006), Boschi *et al.* (2006), and Boschi (2006), and while differences between ray-theory and finite-frequency models exist, they do not seem to be as important as those caused, at this stage, by (i) or (ii). (i) has been a major limiting factor in the past: the density of the parameterization is proportional to the number of basis functions (i.e. number of model coefficients, or "free parameters") used to describe the tomographic image, which in turn defines the size of the inverse problem to be solved. Pioneers of global seismic tomography like Dziewonski (1984) or Woodhouse and Dziewonski (1984), even using what at the time were regarded as very powerful computers, could only afford a model parameterization in terms of $\sim 10^2$ free parameters. The mid-90s breakthroughs of Grand (1994) and van der Hilst *et al.* (1997) consisted in employing a *voxel* parameterization (as opposed to the harmonic one of Dziewonski (1984) and Woodhouse and Dziewonski (1984)), resulting in a sparse inverse problem, solvable by iterative algorithms–hence, lower RAM and computation time requirements (e.g. Trefethen and Bau, 1997), allowing to invert for $\sim 10^5$ model coefficients. A decade later, this issue is not

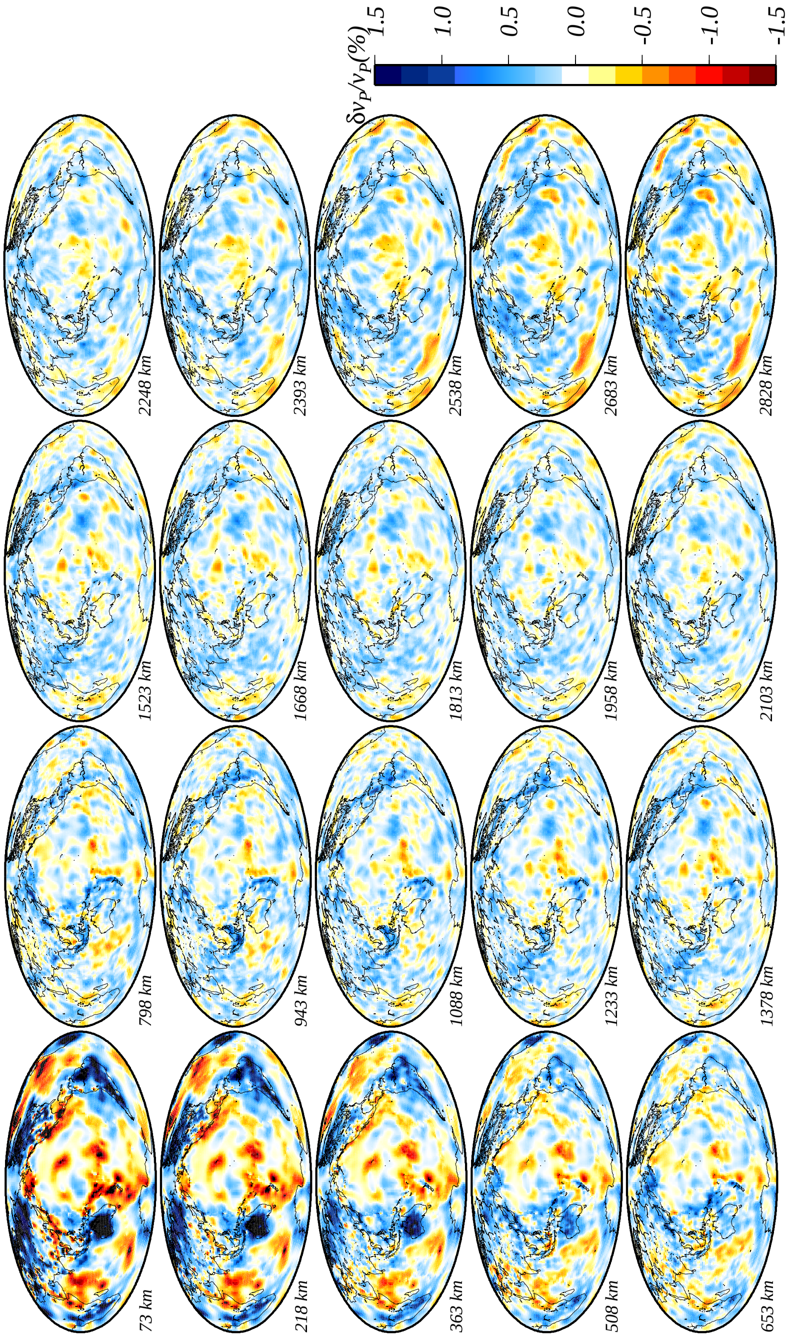


Fig. 1. P-velocity model *vox1.5p*, shown in its entirety. The mean depth of each layer is given below and to the left of the corresponding panel. *vox1.5p* was derived from Antolik *et al.*'s (2001)database of P-wave travel times.

as relevant anymore. Owing to an adaptive-grid approach, Bijwaard *et al.* (1998) have been able to make use of a parameterization locally as fine as 0.6°, while keeping the total number of free parameters relatively low. More recently, model $vox1.5p$ of Boschi et al. (2007) (figure 1) based on an approximately equal-volume grid, achieves instead a constant nominal resolution of 1.5°. It consists of 366,640 free parameters, and yet one inversion requires only minutes on a 1-CPU desktop computer.

We infer that at the current stage of global seismic tomography, the main factor limiting resolution is data coverage, which, without a large network of ocean-bottom receivers, will remain poor in regions underlying oceans. In the absence of uniform station coverage, the main challenge for seismic tomographers is to establish appropriate parameterization/regularization criteria, to damp instabilities caused by lack of data, without obscuring valuable information.

3 Statistically sound, linearized ray-theory tomography

Establishing a criterion to identify the highest-likelihood model in a family of solutions that would intuitively be considered "acceptable" has been a major problem–and limiting factor for resolution–in global seismic tomography, with the choice of a "best" model left to the author's subjective consideration.

Since the seminal work of Akaike (1974), rigorous "information criteria" have been derived (e.g., Burnham and Anderson, 2002; Hurvich and Tsai, 1989; Leonard and Tsu, 1999) to determine the actual number of free parameters needed to explain a given seismic database; they have not been applied often to global seismic tomography, probably because of their high computational cost. They require that many inversions be performed on grids of various density (nominal resolution); the "number of degrees of freedom" associated with each inversion must also be found, evaluating the model resolution matrix $\mathbf{R}$ and its trace (Boschi *et al.*, 2006). This is the most time-consuming step, but can be perfectly parallelized as explained e.g. by Soldati *et al.* (2007).

We have experimented with Antolik *et al.*'s (2001) database of P-wave travel-time observations, inverting them for isotropic, 3-D structure in mantle P-velocity. The CPU-time needed to conduct a family of such inversions, spanning a broad range of solution-model complexity values, is shown in figure 2 as a function of parameterization density. The CPU-time for one inversion at the highest resolution considered here ($\sim 10^5$ voxels of 1.5° horizontal extent) is $\sim 10^2$s in the "acceptable"-solution region, and to find $\mathbf{R}$ we must complete $\sim 10^5$ such inversions, resulting in a total single-CPU time of $\sim 10^7$s, or, from our benchmark of the CPU on which the exercise was conducted (speed

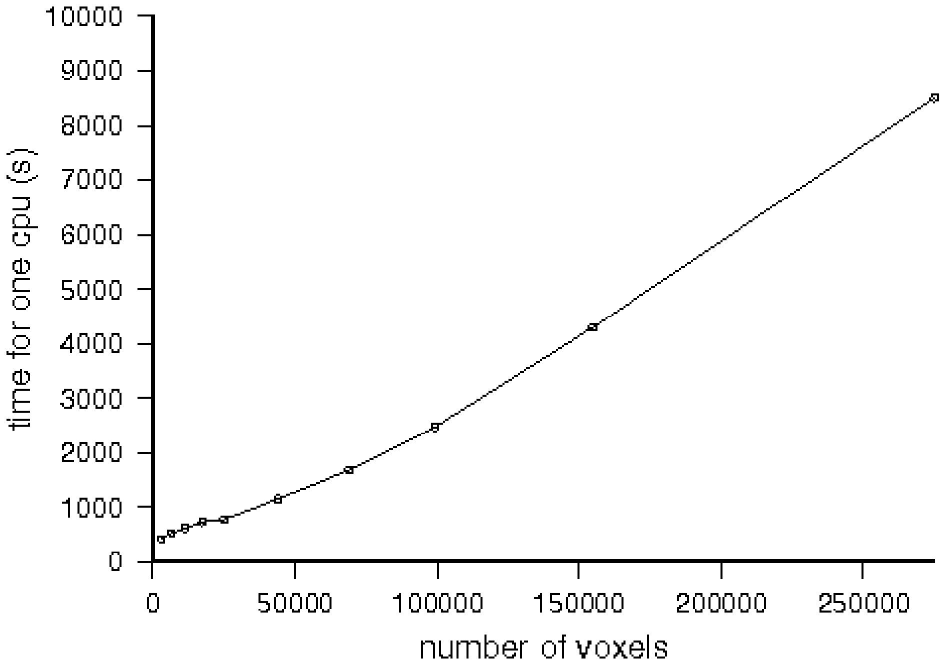


Fig. 2. At each parameterization level (horizontal axis, from 15° to 1.5° nominal resolution; 15 layers) we conduct 27 LSQR inversions, each with a different regularization parameter. The time needed to complete this exercise is plotted on the vertical axis. We find 27 solutions of variable roughness, ranging between the strongly underdamped and strongly overdamped regions.

$\sim 1 \times 10^9$ Flop per second), $\sim$ 10Petaflop. Two such computations will need to be performed.

We applied AICC, or Akaike corrected information criterion (Hurvich and Tsai, 1989; Dal Forno *et al.*, 2005) to the mentioned, global mantle P-velocity inverse problem. The densest grid we employed has 3.75° horizontal spacing, while the vertical parameterization remains constant (15, $\sim$ 200km-thick layers). Calculations of $\mathbf{R}$ were conducted on a 20-CPU Linux cluster. Results shown in figure 3 indicate that the information content of both weakly and strongly regularized solutions continues to grow with growing number of degrees of freedom. To find the curve's maximum, the exercise needs to be iterated on even denser grids, requiring in practice (as one could estimate from figure 2) petascale capacities.

AICC is a subjective choice, and we plan to explore other information criteria, the most popular alternative to AICC being perhaps the Bayesian information criterion, employed for example by Oda and Shibuya (1996), or Sambridge (2006).

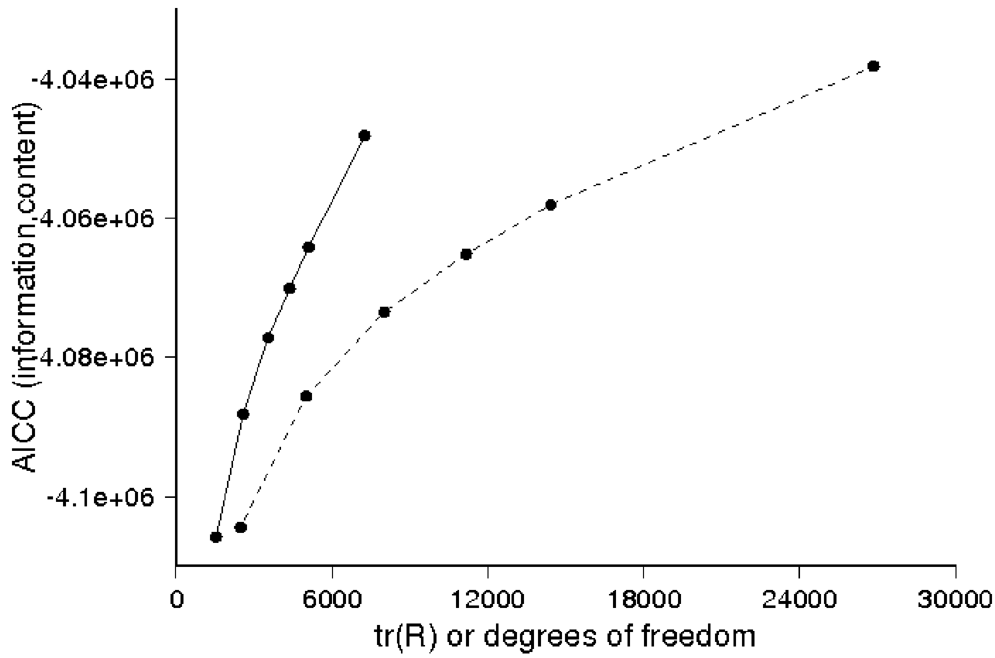


Fig. 3. Corrected Akaike-criterion likelihood as a function of the trace of the resolution matrix. The latter is a measure of the number of degrees of freedom of the solution. We change it by leaving the regularization constraints fixed, but varying the parameterization density (15 vertical layers with decreasing 15°, 10°, 7.5°, 6°, 5° and 3.75° horizontal gridsize). The solid line corresponds to strongly damped, but acceptable solutions; the dashed to weakly damped but acceptable.

4 Numerical finite-frequency tomography

An increasing number of authors in global seismology are beginning to use finite-frequency sensitivity kernels rather than simple ray theory to develop higher-resolution tomographic images of the Earth's mantle, inverting seismic observations made at relatively long periods, where finite-frequency effects might be more relevant and affect tomographic resolution strongly (Boschi (2006) for a list of more or less recent works in global finite-frequency tomography). High-performance computers allow to compute sensitivity kernels numerically, by means of the adjoint method (Tromp *et al.*, 2005; Peter *et al.*, 2007) and/or the scattering integral method (Chen *et al.*, 2006). As opposed to the analytical approach (e.g. Dahlen *et al.*, 2000), numerical methods are more flexible with respect to changes in the reference model, whose lateral heterogeneities will be properly accounted for.

An example of the effects of lateral heterogeneities on sensitivity kernels is shown in figure 4, where the finite-element "membrane wave" approach (Tanimoto, 1990; Tape, 2003; Peter *et al.*, 2007; Tape *et al.*, 2007) is used to compute the sensitivity of Love-wave phase anomalies to phase velocity at a period of 150s. Differences between spherical- and aspherical-Earth kernels are small, but comparable to the kernels themselves; while not affecting the long-wavelength character of our global tomographic images, they become increasingly relevant as features of shorter wavelength are to be resolved (Peter *et al.*, 2007).

4.1 Implementation and computational cost

The computation of sensitivity kernels is by far the most expensive step of any finite-frequency tomography algorithm. There exists one kernel per source-receiver couple, i.e. one kernel per observation, and in principle the adjoint method requires that two simulations be conducted to compute each kernel. However, the total number of simulations to compute all kernels associated with a given database can be reduced in various ways (e.g., Capdeville *et al.*, 2005; Tromp *et al.*, 2006). Most recently, Chen *et al.* (2006) show that this number can be reduced to $3n_R + n_S$, where n_R denotes the number of (3-component) receivers, and n_S the number of sources.

Today, the most widely used, and possibly most efficient algorithm for numerical simulations of global seismic wave propagation is the spectral-element software package *Specfem* (e.g., Komatitsch *et al.*, 2002). Ampuero and Nissen-Meyer (2007) show that the cost of one run of*Specfem* is related to the shortest (most expensive) period to be accurately modeled, T_{min}, by

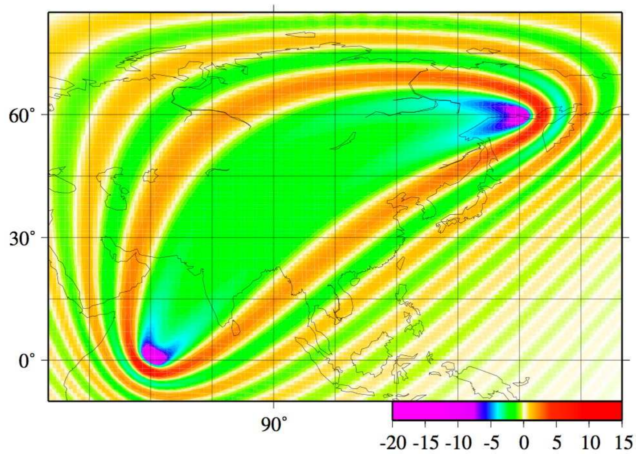
$$\text{cost in Flop} = \left(\frac{2\pi\Delta}{c_0 T_{\text{min}}}\right)^4 \times \Gamma, \tag{1}$$

where Δ denotes epicentral distance and c_0 reference (mean) phase velocity, and the parameter Γ depends on the largest tolerated error, which we define as arrival time error normalized by total travel time. Choosing the latter to be $\sim 10^{-4}$, the curve in figure 5 is found.

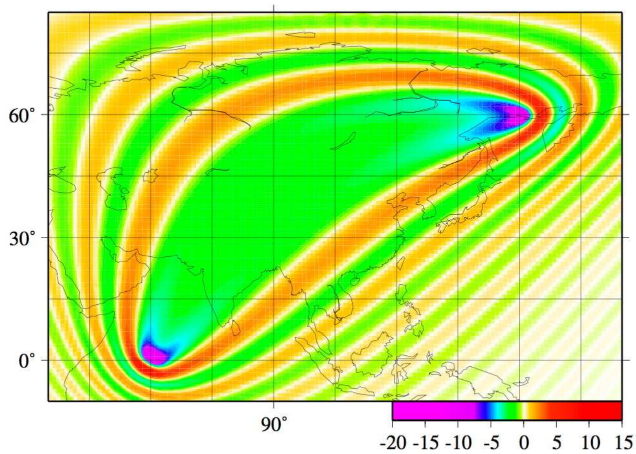
If only minor-arc phase-anomaly observations are considered, then $\Delta \leq 180°$, and with $T_{\text{min}} = 20$s (Qin *et al.*, 2006) (hence the cost of one simulation $\sim 10^7$Gigaflop from figure 5) and $n_R \sim n_S \sim 10^2$, we can expect the cost of computing all necessary kernels to be $\sim 10^3$ Petaflop. This figure does not include the cost of input/output operations (which might become necessary as the shortest modeled period is diminished, parameterization refined, and RAM subsequently becomes insufficient), or reconstruction of the *forward* wavefield by solving the wave equation backwards in time (Tromp *et al.*, 2005; Chen *et al.*, 2006). In our preliminary runs of *Specfem* on a 20-CPU cluster, we have found the backpropagation of the adjoint wavefield to take roughly as long as three normal forward propagations with the same source-receiver geometry.

Fig. 4. Example of numerical kernels (dimensionless) derived with the adjoint method for 150 s Love waves in (a) homogeneous and (b) heterogeneous starting phase-velocity models. (c) Difference between (a) and (b). (From Peter *et al.*, 2007.)

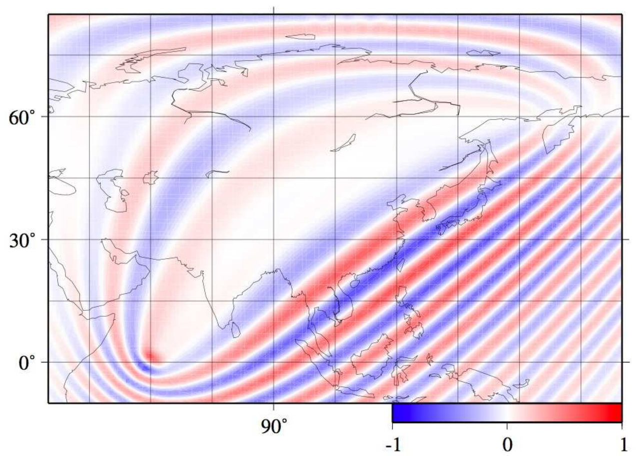
(a)



(b)



(c)



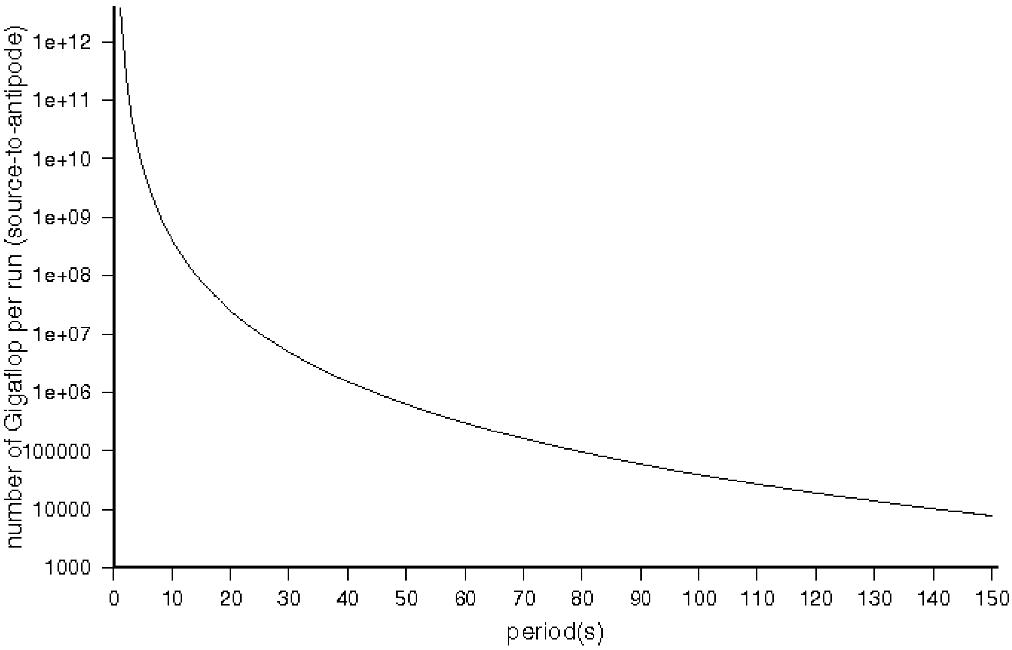


Fig. 5. Expected cost of one global spectral-element simulation of waves propagating from a source to its antipode, as a function of the shortest (and most expensive) accurately modeled period. We chose accuracy, defined as arrival time error normalized by total travel time, to be $\sim 10^{-4}$. (Based on Ampuero and Nissen-Meyer, 2007.)

Sensitivity kernels also need to be updated a few times, repeating each time the same number of *Specfem* runs, and taking the result of each inversion as the starting point for the next, until convergence is reached.

Specfem has been shown to perform and scale extremely well (Komatitsch *et al.*, 2003). Additionally, once an optimal number of processors per simulation has been found, the computation can be further parallelized by performing a number of simulations at the same time, each on a different chunk of the cluster (recall that $3n_R + n_S$ simulations have to be performed at each iteration).

5 Summary

Improving the resolution of tomographic maps is crucial to answer important questions on the nature of the Earth's mantle–the best current example being perhaps the debate on the origin of hotspots and on the very existence of mantle plumes (e.g., Sleep, 2006), presumed narrow features that need high-resolution tomography to be properly mapped.

The RAM and speed of computers available to the scientific community are now sufficient to solve very large inverse problems in a short time, making it easy to derive very finely parameterized seismic images of the Earth. Nevertheless, as tomographers strive to enhance resolution, questions that still need to be addressed are (i) how to identify appropriate parameterization and/or regularization schemes, and (ii) how to surpass the resolution limit implicitly posed by the ray-theory approximation, still adopted by many researchers today.

We propose here to tackle (i) by means of computationally expensive statistical approaches like the Akaike criterion (Akaike, 1974; Hurvich and Tsai, 1989), now made feasible by the advent of petascale computing. We indicate the numerical approach to finite-frequency (Born-approximation) tomography as the best currently available answer to (ii), and analyze its cost as a function of increasing modeled/inverted seismic-wave frequency. The availability of petascale hardware will be integral to the implementation of numerical finite-frequency tomography at increasingly high resolution.

References

[1] Akaike, H., 1974. A new look at the statistical model identification, *IEEE Trans. Autom. Control*, **19**, 716–723.

[2] Ampuero J.-P., and T. Nissen-Meyer, 2007. High order conservative time schemes in spectral element methods for seismic wave propagation, in preparation.

[3] Antolik, M., G. Ekström, and A. M. Dziewonski, 2001. Global event location with full and sparse data sets using three-dimensional models of mantle P-wave velocity, *Pure Appl. Geophys.* **158**, 291–317.

[4] Becker, T. W., and L. Boschi, 2002. A comparison of tomographic and geodynamic mantle models *Geochem. Geophys. Geosyst.* **3**, 1003.

[5] Bijwaard, H., W. Spakman, and E. R. Engdahl, 1998. Closing the gap between regional and global travel time tomography *J. geophys. Res.,* **103**, 30055–30078.

[6] Boschi, L., 2006. Global multiresolution models of surface wave propagation: comparing equivalently regularized Born and ray theoretical solutions, *Geophys. J. Int.,* **167**, 238–252.

[7] Boschi, L., T. W. Becker, G. Soldati, and A. M. Dziewonski, 2006. On the relevance of Born theory in global seismic tomography, *Geophys. Res. Lett.,* **33**, L06302.

[8] Boschi, L., T. W. Becker, and B. Steinberger, 2007. Mantle plumes: seismic images and dynamic models, *Earth Planet. Sci. Lett.*, submitted.

[9] Burnham, K. P., and D. R. Anderson, 2002. *Model selection and multimodel inference. A practical information-theoretic approach*, Springer, New York, second edition.

[10] Capdeville, Y., Y. Gung, and B. Romanowicz, 2005. Towards global Earth tomography based on the spectral element method: a technique based on source stacking, *Geophys. J. Int.,* **162**, 541–554.

[11] Chen, P., T. H. Jordan, and L. Zhao, 2007. Full 3D waveform tomography: A comparison between the scatteringintegral and adjoint-wavefield methods, *Geophys. J. Int.*, in press.

[12] Dal Forno, G., P. Gasperini, and E. Boschi, 2005. Linear or nonlinear rheology in the mantle: A 3D finite-element approach to postglacial rebound modeling, *J. Geodyn.*, **39**, 183–195.

[13] Dziewonski, A. M., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.,* **89**, 5929–5952.

[14] Grand, S. P., 1994. Mantle shear structure beneath the Americas and the surrounding oceans, *J. geophys. Res.,* **99**, 11591–11621.

[15] Hurvich, C. M., and C.-L. Tsai, 1989. Regression and time series model selection in small samples, *Biometrika*, **76**, 297–307.

[16] Ishii, M., and A. M. Dziewonski, 2003. Distinct seismic anisotropy at the centre of the Earth. *Phys. Earth Planet. Inter.*, **140**, 203–217.

[17] Komatitsch, D., J. Ritsema, and J. Tromp, 2002. The spectral-element method, Beowulf computing and global seismology, *Science*, **298**, 1737–1742.

[18] Komatitsch, D., S. Tsuboi, Ji. Chen, and J. Tromp, 2003. A 14.6 billion degrees of freedom, 5 teraflops, 2.5 terabyte earthquake simulation on the Earth Simulator, Proceedings of the Supercomputing 2003 Conference.

[19] Leonard, T., and S. J. Hsu, 1999. *Bayesian methods. An analysis for statisticians and interdisciplinary researchers*, Cambridge University Press, Cambridge, U.K.

[20] Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung, 2004. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science* **303**, 338–343.

[21] Montelli, R., G. Nolet G, F. A. Dahlen, and G. Masters, 2006. A catalogue of deep mantle plumes: New results from finite-frequency tomography, *Geochem. Geophys. Geosyst.*, **7**, Q11007.

[22] Peter, D., C. Tape , L. Boschi, and J. H. Woodhouse, 2007. Surface wave tomography: global membrane waves and adjoint methods, *Geophys. J. Int.*, submitted.

[23] Qin, Y., Y. Capdeville, V. Maupin, and J.-P. Montagner, 2006. A SPICE blind test to benchmark global tomographic methods, abstract, AGU Fall meeting.

[24] Sambridge, M., K. Gallagher, A. Jackson, and P. Rickwood, 2006. Trans-dimensional inverse problems, model comparison and the evidence. *Geophys. J. Int.,* **167**, 528–542.

[25] Sleep, N. H., Mantle plumes from top to bottom, 2006. *Earth-Science Reviews*, **77**, 231–271.

[26] Soldati, G., L. Boschi, and A. Piersanti, 2007. Global seismic tomography and modern parallel computers, *Annals of Geophysics*, in press.

[27] Tanimoto, T., 1990. Modelling curved surface wave paths: membrane surface wave synthetics, *Geophys. J. Int.,* **102**, 89–100.

[28] Tape, C., 2003. *Waves on a spherical membrane*, M.Sc. Thesis, University of Oxford, U. K.

[29] Tape, C., Q. Liu, and J. Tromp, 2007. Finite-frequency tomography using adjoint methods–Methodology and examples using membrane surface waves, *Geophys. J. Int.,* submitted.

[30] Trefethen, L. N., and D. Bau III, 1997. *Numerical Linear Algebra*, Soc. for Ind. and Appl. Math., Philadelphia, Penn.

[31] Tromp, J., C. Tape, and Q. Liu, 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**, 195–216.

[32] Tromp, J., Q. Liu, C. Tape, and A. Maggi, 2006. Towards Seismic Tomography Based Upon Adjoint Methods, *Eos Trans. AGU*, **87**(52), Fall Meet. Suppl., Abstract S53C-05.

[33] van der Hilst, R. D., S. Widiyantoro, and E. R. Engdahl, 1997. *Nature*, **386**, 578–584.

[34] Woodhouse J. H., and A. M. Dziewonski, 1984. Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.,* **89**, 5953–5986.