Bulletin of the Seismological Society of America, Vol. 99, No. 1, pp. 194–214, February 2009, doi: 10.1785/0120080013

LOTOS Code for Local Earthquake Tomographic Inversion: Benchmarks for Testing Tomographic Algorithms

by Ivan Koulakov

Abstract We present the LOTOS-07 code for performing local earthquake tomographic inversion, which is freely available (see the Data and Resources section for the Web site). The initial data for the code are the arrival times from local seismicity and coordinates of the stations. It does not require any information about the sources. The calculations start from absolute location of sources and estimates of an optimal 1D-velocity model. Then the sources are relocated simultaneously with the 3D-velocity distribution during iterative coupled tomographic inversions. The code allows results to be compared based on node or cell parameterizations. The synthetic dataset used for testing the code is based on source–receiver configurations from a real experiment in Costa Rica. The travel times for this dataset are computed by 3D tracing through a rather complicated synthetic model and are perturbed with realistic noise. We also present a series of synthetic datasets with unknown sources and velocity models (see the Data and Resources section for the Web site) that can be used as blind benchmarks for testing different tomographic algorithms. We encourage other users of tomography algorithms to join the program on creating benchmarks that can be used to check existing codes.

Introduction

Seismic tomography is an effective tool for investigating the deep Earth interior. The results provided by tomographic inversions reveal the mechanisms that control tectonic processes in the Earth. Studies exist at many different scales, from global to local. One of the most complicated tomographic schemes is based on using the arrival times of *P* and *S* seismic waves from local earthquakes with unknown parameters (local earthquake tomography [LET]). In this case, the problem is reduced to a coupled inversion for velocity distributions and source parameters. LET is often performed in high-contrast areas (e.g., subduction zones and volcanic areas) and requires performing iterative nonlinear approaches in which ray paths are traced in updated 3D-velocity models.

During the past few decades, several algorithms for LET schemes have been created by various research groups. One of the most popular LET algorithms is the SIMULPS code and its derivatives, a powerful and simple practical realization tool for iterative tomographic inversion. This code is freely available and is used by many authors routinely. This code was created in the 1980s and 1990s (Thurber, 1983; Eberhart-Phillips, 1986, Thurber, 1993; Thurber *et al.*, 1995) and is now being actively developed and updated by different authors. Dozens of tomographic studies in different regions are based on this algorithm (e.g., Hauksson and Haase, 1997; Eberhart-Phillips and Michael, 1998; Graeber and Asch, 1999; Haslinger *et al.*, 1999; Reyners *et al.*, 1999; Dorbath and Masson, 2000; Hauksson, 2000; Husen *et al.*, 2000; Paul *et al.*, 2001; Eberhart-Phillips and Bannister, 2002; Husen *et al.*, 2002; Chiarabba and Amato, 2003; Husen *et al.*, 2003; Husen and Smith, 2004; Daly *et al.* 2008). Another algorithm for LET was developed by Benz *et al.* (1996) and Hole *et al.* (2000), which has been used in other studies (e.g., Ramachandran *et al.*, 2005; Yang and Shen, 2005). An algorithm created by Roecker and coauthors should be also mentioned here (Roecker *et al.*, 1993; Masturyono *et al.*, 2001; Roecker *et al.*, 2004). Another important contribution to tomographic tools was made by Zhao *et al.* (1992, 1995). Among other codes, which seem to us very promising, we can single out the works by Nakajima *et al.* (2001) and Wagner *et al.* (2005), in which a 3D starting model based on *a priori* information is used. Unfortunately, in this overview, we cannot mention all other researchers who contributed to creating different tomographic algorithms according to the LET scheme. However, we find this variety of codes that allows verification of the results produced by different algorithms very positive.

In this situation, with dozens of tomographic models using different codes appearing every year, the most important and difficult task in tomographic inversion is now not showing the results, but providing convincing arguments that these results are related to structure in the real Earth. The

authors often present the results of synthetic modeling based on datasets they created themselves. Sometimes the descriptions of the tests are very short, and a reader cannot follow the details of the forward and inverse modeling in such inner tests. Therefore, one must trust that the authors performed the tests correctly. However, when the synthetic model is known, there is strong temptation to tune the inversion parameters to achieve the best fit of the results with the model. It is obvious that in a case of real data, this is not a possibility. The most unbiased way to check the algorithm is to perform the inversion using a synthetic dataset created by another person using another algorithm. Ideally, the testing should be performed blindly without any knowledge about the model and source locations. Such a strategy would reveal all the problems of the inversion algorithm and show whether it can be used to obtain reliable results based on real data. In this study, we initiate a program for creating benchmarks for testing LET algorithms. We hope that other researchers will join this program and will create a series of blind tests to be used for testing our algorithms.

Here we present the most recent version of the LOTOS-07 algorithm. In this article we focus on technical details of the new features of the algorithm and present its working ability using a synthetic test. We believe that using synthetic modeling is more appropriate for presenting the algorithm than considering real data results. In this case we forget about the sources and the model and perform the inversion in absolutely the same way as we would for real data. Finally, we can compare our results with the true model and make a conclusion about the working ability of the algorithm, which is not possible for real data.

This article has two main purposes. First, it presents the algorithm LOTOS-07, which can be freely downloaded from our Web site (see the Data and Resources section). This algorithm is simple in practical use and can be easily applied by any person to any datasets according to the LET scheme. Our second purpose is to attract the attention of all tomography users to the problems of synthetic testing and to propose a universal benchmark consisting of a series of blind synthetic tests. We believe that testing existing algorithms with these benchmarks will increase the confidence in tomographic results.

Description of the LOTOS-07 Algorithm

General Remarks

We designed a tomographic algorithm, LOTOS-07 (local tomography software) for simultaneous inversion of *P*- and *S*-velocity structures and source coordinates. The LOTOS-07 algorithm can be directly applied to very different datasets. It has a wide range of possibilities for performing different tests and is easy to operate. Previous versions of the code have already been used to investigate the deep structure beneath the Central Andes (Koulakov *et al.*, 2006), Central Java (Koulakov *et al.*, 2007; Wagner *et al.*, 2007), Costa Rica (Dinc *et al.*, 2008), Toba (I. Koulakov, T. Yudistira, *et al.*, unpublished manuscript, 2008), the Anatolian fault region, Nankai (Japan), Vranchea, and other areas. Since then, the code has been significantly updated. The general structure of the LOTOS-07 code is presented in Figure 1. The most important improvement of the code is the inclusion of a block for 1D-velocity optimization, presented in the Algorithm for 1D-Velocity Optimization section. The main

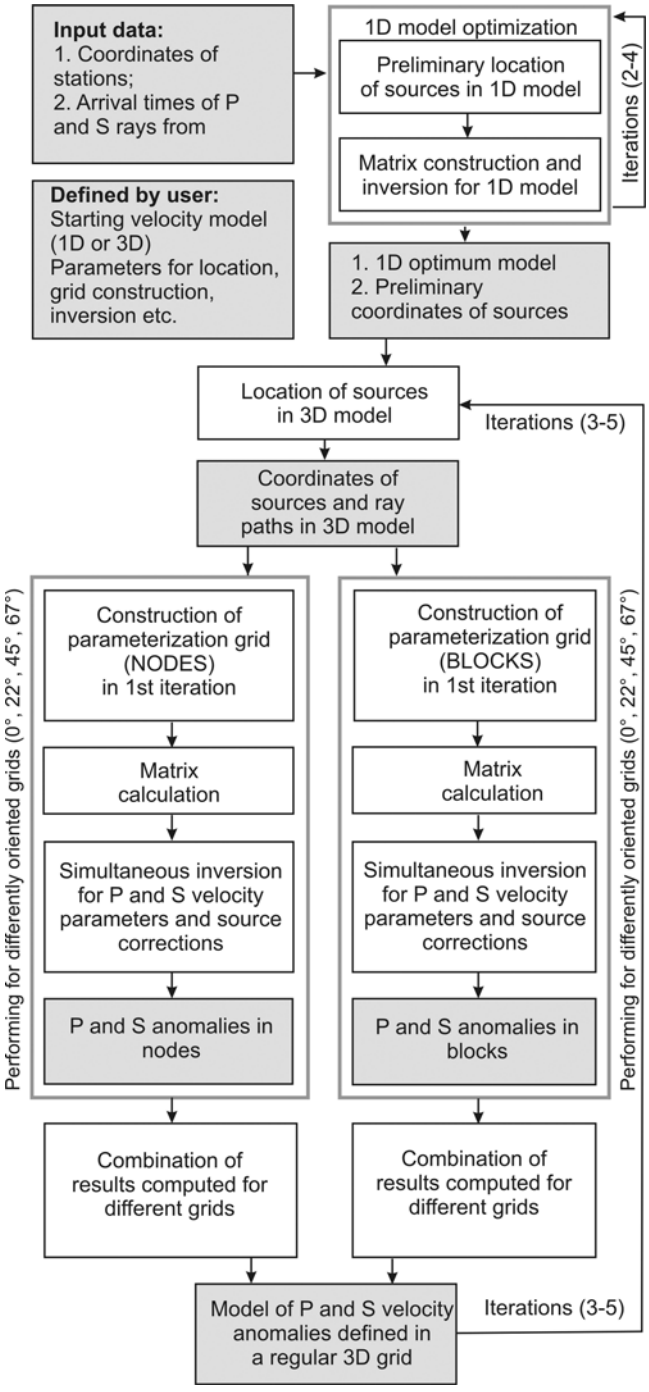

Figure 1. General structure of the LOTOS-07 algorithm. White boxes represent program steps; gray boxes contain the main data used in processing.

steps of the algorithm for 3D-velocity inversion are described in the Bending Algorithm for Ray Tracing in a 3D-Velocity Model section.

In this article, the code is demonstrated using a synthetic dataset. This allows us to quantitatively evaluate the accuracy of source locations and the correctness of the synthetic model reconstruction. The dataset corresponds to a real observation system that was used to compute the 3D-velocity structure of *P* and *S* velocities beneath Costa Rica (Dinc *et al.*, 2008). In total, we used information from more than 1000 events and more than 30,000 *P* and *S* picks. The synthetic model is described in detail in the Synthetic Modeling section. After computing the synthetic travel times, all information about sources and velocity models was forgotten, and we found ourselves in the same situation as for real data. This dataset is one of the benchmark blind tests we created and is freely available (see the Data and Resources section for Web site information). This dataset will be described in detail in the Bench Mark for Testing the Tomographic Algorithms section. The LOTOS-07 code is available online, so anyone can repeat the same calculations and obtain the same images as those presented in this article.

The input dataset for LOTOS-07 includes two files.

1. A file with station geographic coordinates, including longitude, latitude, and elevation of stations.
2. A file with arrival times and initial locations of sources (if available). In the presented dataset, as for real data, coordinates and origin times of sources are not given. Therefore, instead of the source coordinates, an arbitrary point (the center of the study area) is given. The uncertainty of the origin times is modeled by adding a random bias to all travel times for each source.

In addition, a preliminary guess for a starting 1D model and a set of free parameters (for location, parameterization, inversion, visualization, etc.) should be defined by the user.

In the LOTOS-07 code we do not require that sources be located inside the network of stations (having an azimuthal gap $< 180°$), as is the case in many tomographic studies. We suppose that this requirement does not reflect the real importance of an event for tomographic inversion. For example, according to this criterion, a shallow event located close to stations, but outside the network would be rejected, while another event at a 600 km depth with an epicenter projection coinciding with the network area would be used. It is obvious that the contribution of the first event for investigating seismic structure would be much more important. In LOTOS-07 we set the requirements less strictly. For example, an event is rejected if the lateral distance to the nearest station is more than some predefined distance (e.g., 200 km).

Algorithm for 1D-Velocity Optimization

We present an algorithm to evaluate 1D-velocity models that can be used as a starting model for 3D-tomographic inversion. As is stated in our previous studies (e.g., Koulakov *et al.*, 2007), the problem of absolute-velocity determination based on data from natural sources with unknown parameters is very unstable. This instability is mostly due to a trade-off between velocity distribution, origin time, and depth of sources. At the same time, incorrect information about the starting absolute-velocity distribution can cause some bias in the final results. To explore the effects of a starting model on the final results of the tomographic inversion and to reduce these effects, we have developed an algorithm to evaluate 1D-velocity models.

Similar steps have been performed in many other works. One of the most popular algorithms for 1D-velocity evaluation is the VELEST code (Kissling *et al.*, 1994), which has been used in dozens of studies. This code computes a 1D model that is used for a 3D-tomographic inversion. Calculations are performed based on a trial and error method and consist in locating sources in many different 1D models.

Our version of the algorithm is based on the iterative repetition of the following steps:

0. Data selection for an optimization. From the entire data catalog, we select events that should be distributed as uniformly with depth as possible. To do this, we select for each depth interval the events with the maximum number of recorded phases. The total number of events in each depth interval should be less than a predefined value (e.g., four events).
1. Calculation of a travel time table in a current 1D model. In the first iteration, the model is defined manually with the use of possible *a priori* information. The travel times between sources at different depths to the receivers at different epicentral distances are computed in a 1D model using analytical formulas (Nolet, 1981). The algorithm allows the incidence angles of the rays to be defined in order to achieve similar distances between rays at the surface.
2. Source location in the 1D model. The travel times of the rays are computed using tabulated values obtained in step 1. The travel times are then corrected for elevations of stations. The source location is based on calculating a goal function (GF) that reflects the probability of a source location in a current point. The form of the GF is defined in Koulakov and Sobolev (2006). Searching for the GF extreme is performed using a grid search method. We start from a coarse grid and finish our search in a fine grid. This step is performed relatively quickly as it uses the tabulated values of the reference travel times.
3. Calculation of the first derivative matrix along the rays computed in the previous iteration. Each element of the matrix $\mathbf{A}_{ij}$ is equal to the time deviation along the *j*th ray caused by a unit-velocity variation at the *i*th depth level. The depth levels are defined uniformly, and the velocity between the levels is approximated as linear.
4. Matrix inversion is performed simultaneously for the *P* and *S* data using the matrix computed in step 3. In addition to the velocity parameters, the matrix contains the

elements to correct the source parameters (dx, dy, dz, and dt). The data vector contains the residuals computed after the source location (step 2). Regularization is performed by adding a special smoothing block. Each line of this block contains two equal nonzero elements with opposite signs that correspond to neighboring depth levels. The data vector in this block is zero. Increasing the weight of this block smooths the solution. If there is *a priori* information about the existence of interfaces (e.g., Moho), it can be included in the inversion. In this case, the link between the pair of nodes just above and below the interface would be skipped.

Optimum values for free parameters (smoothing coefficients and weights for the source parameters) are evaluated on the basis of synthetic modeling. The inversion of this sparse matrix is performed using the least-squares QR (LSQR) method (Paige and Saunders, 1982; van der Sluis and van der Vorst, 1987). A sum of the obtained velocity variations and the current reference model is used as a reference model for the next iteration, which contains steps 1–4. The total number of iterations is also determined according to the results of synthetic modeling.

Results of the 1D-model optimization are shown in Figure 2. The true-velocity model (dashed curve) is the same in both cases. Here we use two different starting models (black lines) to investigate the stability of the optimization. In model 1 (upper panel), we use a simple starting-velocity model without a low-velocity layer above 7 km depth. For the depth interval of 10–40 km, the fit of the optimized model (gray curve) with the true one seems to be satisfactory. For the shallower and deeper parts, the optimization for P velocity has a rather large misfit. For the second model (lower panel), we included the low-velocity layer in the shallowest part of the starting model, which improved the general fit of the model. At the same time, the inversion produced an artifact at a depth of ~10 km, which causes a low-velocity zone between 10 and 20 km. This could significantly change the ray paths and resulting patterns. However, as will be shown later, we did not see a significant difference between the resulting anomalies at 15 km deep in cases of inversion based on these two 1D models.

These results confirm our statement in our previous works (e.g., Koulakov *et al.*, 2007) that in the case of the LET scheme, reconstructing relative anomalies is much more stable than reconstructing absolute velocities.

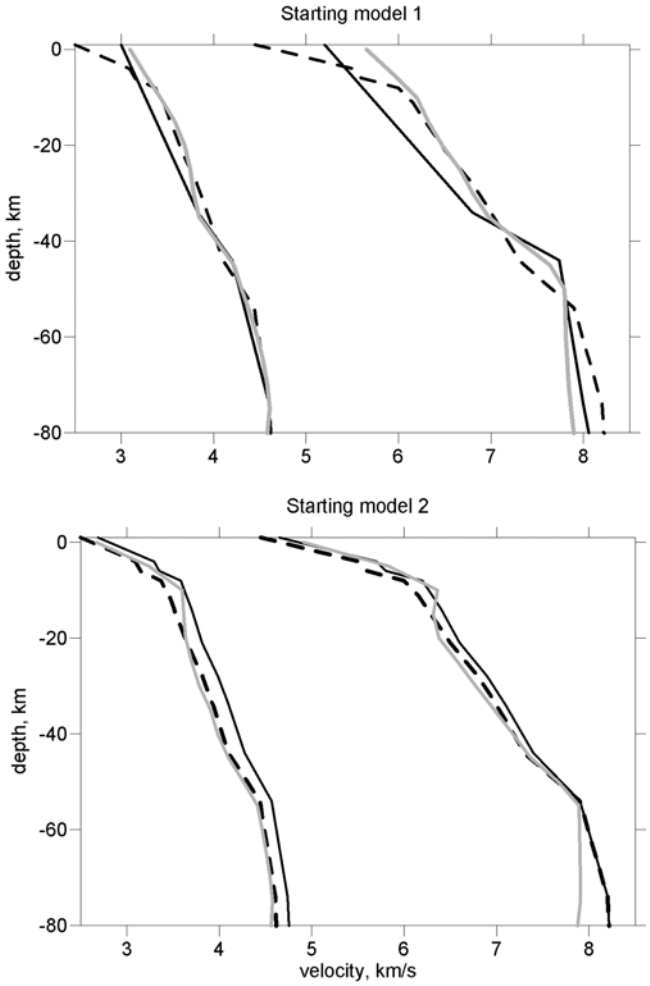

Figure 2. Results of 1D-model optimization. The cases of two different starting models are presented. The dashed curve is the true-velocity distribution in the synthetic model. The black curve represents the starting model. The gray curve is the resulting velocity distribution.

Bending Algorithm for Raytracing in a 3D-Velocity Model

One of the key features of the LOTOS-07 code is a raytracing algorithm based on the Fermat principle of travel-time minimization. A similar approach is used in other algorithms (e.g., Um and Thurber, 1987) and is called bending tracing. We present our own modification of the bending algorithm. An important feature of this algorithm is that it can use any parameterization of the velocity distribution. It is only necessary to define uniquely one positive-velocity value at any point of the study area. It can be done, for example, with nodes, cells, polygons, or analytical laws. The current version of LOTOS-07 includes various options for velocity definition. However, if necessary, any other parameterization can be easily included.

A basic principle of our bending algorithm is shown in Figure 3. In the presented example, we use a model with exaggerated-velocity contrasts. In the vertical direction, the velocity varies from 2.5 to 9 km/sec. The checkerboard anomalies have amplitudes of ±30%. It is obvious that in this model, the ray path has a fairly complicated shape determined by the velocity distribution.

Searching a path with minimum travel time is performed in several steps. The starting ray path is a straight line. In the first step (Fig. 3a), the ends of the rays are fixed (points 1 and 2), and point A in the center of the ray is used for bending. Deformation of the ray path is performed perpendicular to the ray path in two directions: in and across the plane of

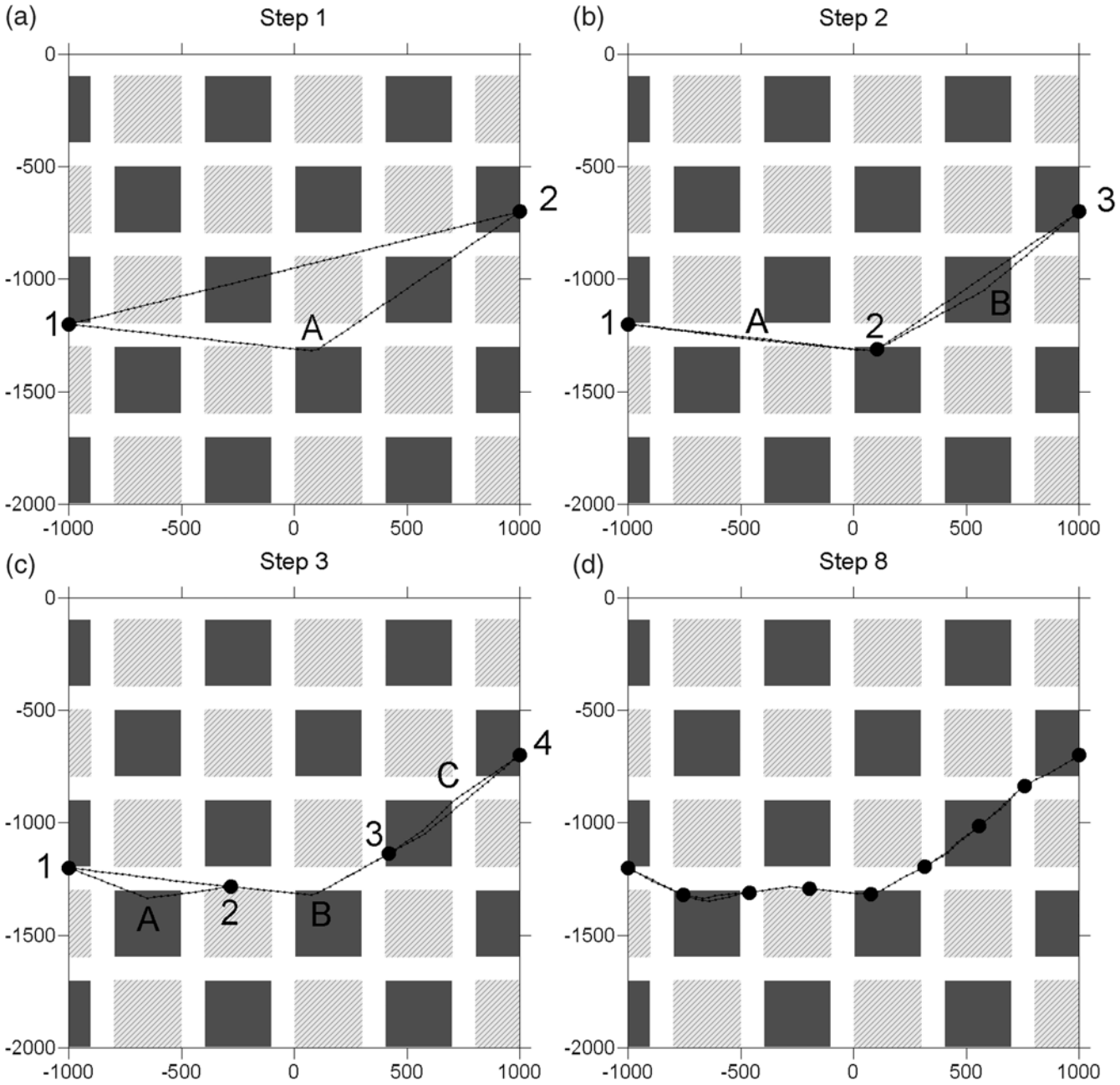

Figure 3. Grounds of the bending algorithm. Ray construction is demonstrated for a model with exaggerated-velocity contrasts. 1D velocity varies from 2500 to 9000 m/sec at 2000 m depth. Hatched light gray patterns represent negative anomalies of −30%; dark gray patterns are positive anomalies of +30%. Details of the bending algorithm are given in the text.

the ray. The values of shift of the new path with respect to the previous path depend linearly on the distance from A to the ends of the segment, as shown in Figure 3. In the second step (Fig. 3b), three points are fixed (points 1, 2, and 3), and deformation of the ray path is performed in two segments (at points A and B). In a third step (Fig. 3c), four points are fixed and three segments are deformed. In the panel labelled Step 8 (Fig. 3d), the results of bending are shown for eight segments. The ray constructed in this way tends to travel through high-velocity anomalies and avoids low-velocity patterns. It should be noted that although a 2D model is shown in Figure 3, the algorithm is designed for the 3D case.

Iterative Tomographic Inversion

The starting 1D-velocity model and initial locations of sources are obtained in the step of 1D-model optimization (see the General Remarks section). The sources are then relocated using a code based on 3D raytracing (bending). As for 1D modification, the location algorithm is based on find-

ing an extreme of a GF. The description of the GF is the same as in the 1D case. However, the grid search method, which is very efficient for 1D models, seems to be too time consuming when 3D raytracing is applied. We therefore use a gradient method (Koulakov *et al.*, 2006) to locate sources in 3D models, which is not as robust as the grid search method but is much faster.

To parameterize the velocity perturbations, two alternative options are used in the LOTOS-07 code. The first method is parameterization with nodes that are installed in the study volume using the algorithm described in Koulakov *et al.* (2006). The nodes are based on vertical lines distributed regularly in map view (e.g., with steps of 5 × 5 km). In each vertical line, the nodes are installed according to the ray distribution. In the absence of rays, no nodes are installed. The spacing between the nodes is chosen to be smaller in areas of higher ray density. However, to avoid excessive concentration of nodes, a minimum spacing is defined (e.g., 5 km). Between the nodes, the velocity distribution is approximated linearly. Examples of node distributions in the depth interval of 10–20 km for the *P* model are shown in the upper panels of Figure 4.

In addition, the LOTOS-07 algorithm allows for the use of another method of parameterization with rectangular cells. The cells are constructed in Cartesian coordinates by subdividing the study volume into parallelepipeds of variable size according to the ray density. The size of the cells is larger if the ray density is lower. Practical realization of grid construction is based on iterative subdivision of parallelepipeds in two equal parts across the x, y, and z directions, one after another. The procedure starts from the entire study volume and ends when the size of the cells becomes smaller than a predefined value. In each iteration, the subdivision of cells into two parts is performed only if the sum of the length of the rays inside a current block is larger than a predefined value. The value of the retrieved-velocity anomaly in each block is presumed to be constant. A similar approach has been used in some regional (e.g., van der Hilst and Engdahl, 1991) and global (e.g., Bijwaard *et al.*, 1998) studies.

In order to reduce the effect of node/cell distributions on the results, we perform the inversion using several grids with different basic orientations (e.g., 0°, 22°, 45°, and 67°). Examples of two different grids for node and cell parameterizations with the basic orientations of 0° and 45° are demonstrated in Figure 4 (left- and right-hand columns, respectively). After computing the results for grids with different orientations, they are stacked into one summary model, reducing any artifacts related to grid orientation.

It is important to note that the total number of nodes/cells can be larger than the ray number. This does not cause any obstacles for performing the inversion, because in our case, the unknown parameters associated with the parameterization nodes/cells are not independent but are linked through a smoothing block that will be described later for the step of inversion. If the parameterization spacing is significantly smaller than the sizes of the expected anomalies, the results of the inversion are almost independent of the distribution of nodes/cells. In this sense, our parameterization can be considered quasi continuous. The construction of the parameterization grids is performed only in the first iteration. In the next iterations, the algorithm uses the same node/cell configurations.

The first derivative matrix is calculated using the ray paths computed after the source locations in the 3D model. Each element of the matrix, $\mathbf{A}_{ij} = \partial t_i / \partial v_j$, is equal to the time deviation along the ith ray due to a unit-velocity perturbation in the jth node/block.

Inversion of the entire sparse $\mathbf{A}$ matrix is performed using an iterative LSQR algorithm (Paige and Saunders, 1982; van der Sluis and van der Vorst, 1987). In addition to *P*- and *S*-velocity parameters, the matrix contains the elements responsible for the source (dx, dy, dz, and dt), and station corrections. The amplitude and smoothness of the solution is controlled by two additional blocks. The first block is a diagonal matrix with only one element in each line and zero in the data vector. Increasing the weight of this block reduces the amplitude of the derived *P*- or *S*-velocity anomalies. The second block controls the smoothing of the solution. Each line of this block contains two equal nonzero elements of opposite sign that correspond to all combinations of neighboring node/cells in the parameterization grid. The data vector in this block is also zero. Increasing the weight of this block reduces the difference between solutions in neighboring nodes, resulting in smoothing of the computed-velocity fields.

The steps of grid construction, matrix calculation, and inversion are performed for several grids with different basic orientations. The resulting velocity anomalies derived for all grids are combined and computed in a regular grid. This model is added to the absolute-velocity distributions used in a previous iteration. New iterations repeat the steps of source location, matrix calculation, and inversion.

One of the most delicate problems in tomographic inversion is correctly defining the free parameters for inversion (smoothing and amplitude coefficients, weights for source and station corrections, number of iterations, etc.). In many tomographic studies (e.g., Eberhart-Phillips, 1986; Graeber and Asch, 1999; Paul *et al.*, 2001; Husen *et al.*, 2003; Wagner *et al.*, 2005; Yang and Shen, 2005), in order to estimate the optimal regularization level, the authors investigate the relationships between the amplitudes of the solution, root mean square (rms) of residuals, and regularization coefficients (so-called trade-off curves [TOCs]). Visually analyzing the TOCs, they declare that the best solution corresponds to the corner point of the L-shaped curve. However, this approach seems not always adequate. First, in many cases, the visual analysis of TOCs does not uniquely reveal the point of maximal curvature. Second, in most of studies the damping parameter is evaluated based on a TOC computed in the first iteration. However, it is clear that the amplitudes of the solutions in second, fifth, and tenth iterations would be different. The TOCs computed in the first iteration cannot pro-

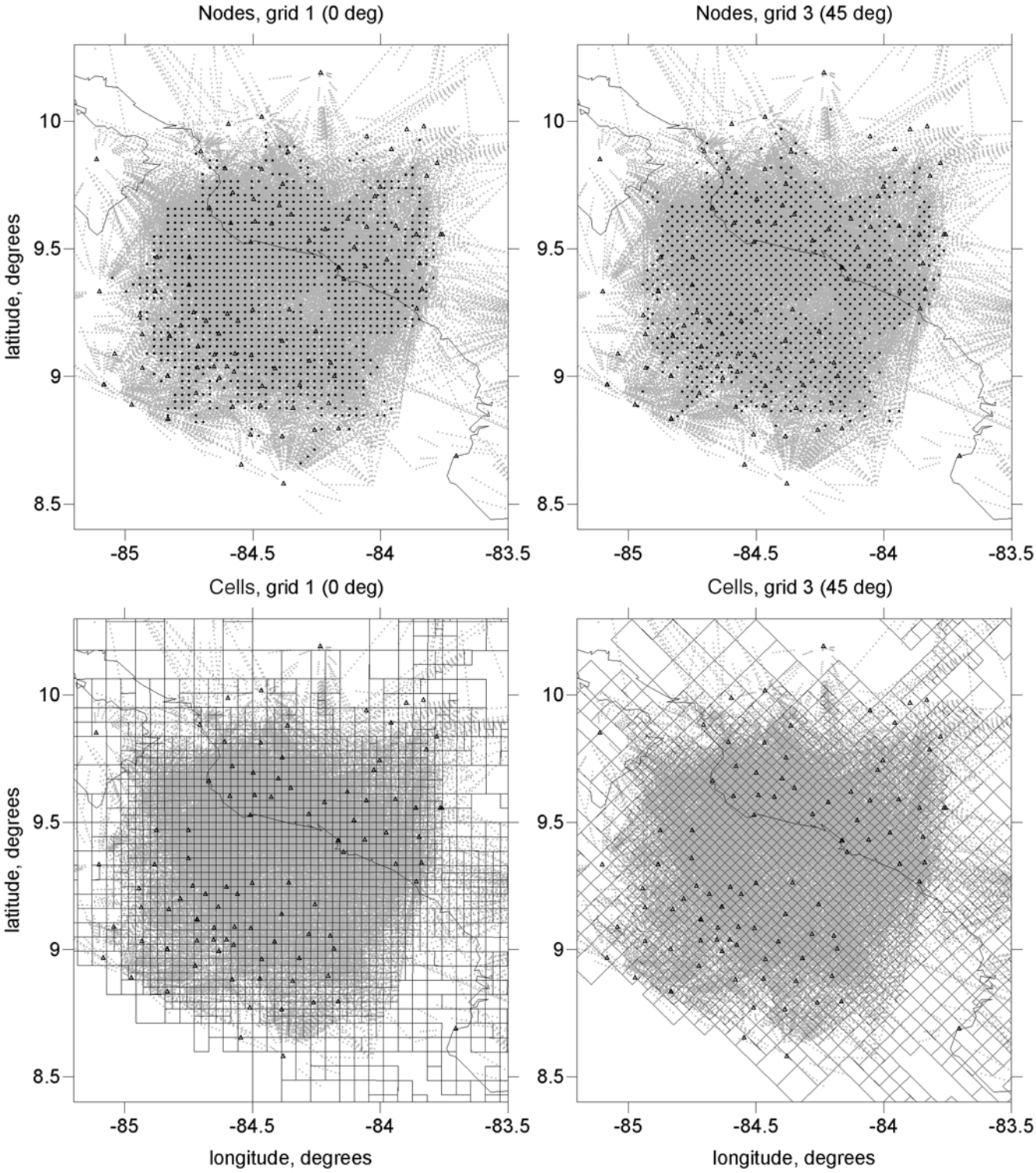

Figure 4. Two types of parameterization provided by LOTOS-07 code (upper row, nodes; lower row, cells). For both cases, two orientations of grids are shown (0° and 45° in the left-hand and right-hand columns, respectively). Gray points show the paths of *P* rays in the depth interval 10–20 km. The grids are used to compute the *P* model and are presented for a depth of 20 km. The triangles represent seismic stations.

vide the answer as to which of them is closer to the reality. Computing the TOCs for the final iteration is time consuming and seems, to us, not efficient. These and other arguments against using TOCs in iterative nonlinear tomographic inversion are discussed in more detail in Appendix B.

In our opinion, the most effective and unbiased way to evaluate optimal values of free inversion parameters is by performing synthetic modeling that reproduces the real situation. This also allows qualitative estimates of amplitudes of seismic anomalies in the real Earth. This approach is investigated in detail in Koulakov *et al.* (2007) and I. Koulakov, T. Yudistira, *et al.* (unpublished manuscript, 2008).

Additional Options

The LOTOS-07 algorithm contains a much larger spectrum of possibilities than are presented in the online version. For various reasons, several options are not yet included in the Web site. The following additional possibilities have been developed in the framework of the LOTOS-07 code.

Inversion for V_P and V_P/V_S Ratio. The inversion for the V_P and V_P/V_S ratio can be performed in parallel with V_P and V_S inversion. This algorithm is described in Koulakov *et al.* (2007).

Attenuation Tomography. The distribution of attenuation in the study area is computed based on values of t^*. Calculations are performed within one iteration based on the velocity model, locations of sources and the ray paths obtained after iterative inversion for V_P, V_S, and source parameters. An older version of the algorithm was used to invert t^* data in Central Chile (Koulakov *et al.*, 2006). The most recent version of the code was applied to process data in the area of the Marmara Sea (Bindi *et al.*, unpublished manuscript, 2008).

Anisotropic Tomography. Based on the general architecture of the LOTOS-07 code, we have developed an algorithm for performing the anisotropic tomographic inversion. The inversion provides 3D anisotropic *P*- and isotropic *S*-velocity distributions based on *P* and *S* travel times from local seismicity. For the *P* anisotropic model, we determine four parameters in each parameterization block. These represent an orthorhombic anisotropy with one predefined direction oriented vertically. Three of the parameters describe slowness variations along three horizontal orientations with azimuths of 0°, 60°, and 120°, and one is a perturbation along the vertical axis. This algorithm has been implemented for investigating the crustal and uppermost mantle structure beneath Central Java (I. Koulakov, A. Jakovlev, *et al.*, unpublished manuscript, 2008).

Planning of Network Deployments. We have developed a tool that can be used to plan optimum deployment of seismic networks. The algorithm produces a realistic distribution of synthetic events, either on the basis of the worldwide catalog or manually. Then, several installations of seismic stations are proposed, and picks are generated based on 3D tracing in a synthetic model. Comparing the results, one can select the network that provides the best resolution for the target objects. This tool has already been used to plan projects in Chile, Japan, Sumatra, and other areas.

Synthetic Modeling

In this article, we illustrate the working ability of the LOTOS-07 code with synthetic examples that allow quantitative estimates of the method limitations. The synthetic models in our works are presented as a superposition of a 1D-velocity model and 3D-velocity anomalies. One of the configurations of the velocity anomalies, here used as the main model for testing the code, is presented in Figure 5. This model is represented by a combination of vertical prisms of complex shapes that compose the portrait of Simon Bolivar. In map view, this model does not change with depth. Amplitudes of anomalies in black and white areas are +7% and −7%, respectively.

The configuration of source–receiver pairs in this test corresponds to a real passive seismic experiment performed in Costa Rica (Dinc *et al.*, 2008). The travel times between the sources and receivers are computed by tracing through the synthetic model using the 3D-bending algorithm described in the Bending Algorithm for Raytracing in a 3D-Velocity Model section. The travel times can be perturbed with noise. Here we consider datasets both with and without noise to explore the stability of the algorithm toward noise.

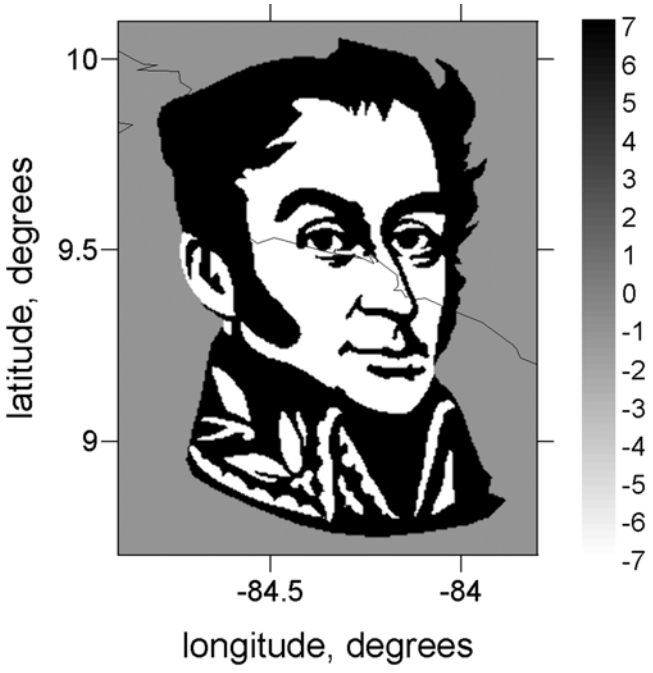

Figure 5. Initial model used to create the dataset. Black and white patterns that compose the portrait of Simon Bolivar are represented by unlimited vertical prisms with velocity perturbations of ±7% with respect to the reference model (dashed curve in Fig. 2).

The LOTOS-07 algorithm has several options for adding the noise. In the considered dataset, noise is produced according to the remnant residuals after real data inversion for the same source–receiver pairs. We suppose that the remnant residuals after performing full tomographic inversion of real data are partly related to picking errors and partly related to the limited resolution capacity of the algorithm. In this test, to model the picking error in the initial data, the noise is defined as 80% of values of real remnant residuals. We believe that this noise definition is better than the random noise generators used in most tomographic studies because it takes into account not only the observation system configuration but also the different receiving conditions in different stations. In the considered dataset, the rms of noise for *P* and *S* data were 0.113 and 0.132 sec, respectively.

After computing the travel times, we forget everything about the velocity model and sources and perform full reconstruction using only two files with the coordinates of stations and arrival times from unknown sources. This causes the synthetic inversion to adequately reflect real data processing. In the synthetic dataset, the initial locations of sources are presumed to be unknown and are fixed in the central point of the study area (longitude −84.5°, latitude 9.5°, and depth of 0). The processing starts from the preliminary source location with simultaneous optimization of the 1D-velocity model. Results of optimization for the 1D model are presented in Figure 2 and described in the Algorithm for 1D-Velocity Optimization section. The distribution of events after preliminary location using tabulated travel times is shown in Figure 6 in the map and cross section corresponding to iteration 0. In this step, the average error of locations with respect to the true sources is 4.81 km. Results for the first iteration correspond to the locations of sources in the optimized 1D model, the same as in the case for iteration 0, but using the bending tracing algorithm. In this case, the average error of the source location is 4.68 km. The results for iterations 3 and 5 correspond to source locations in the 3D models obtained after two and four iterations of tomographic inversion. For these cases, the errors of the source locations decreased to 3.57 and 3.37 km, respectively.

Figure 7 shows the results of reconstruction of the Simon Bolivar model after performing iterations 1, 3, and

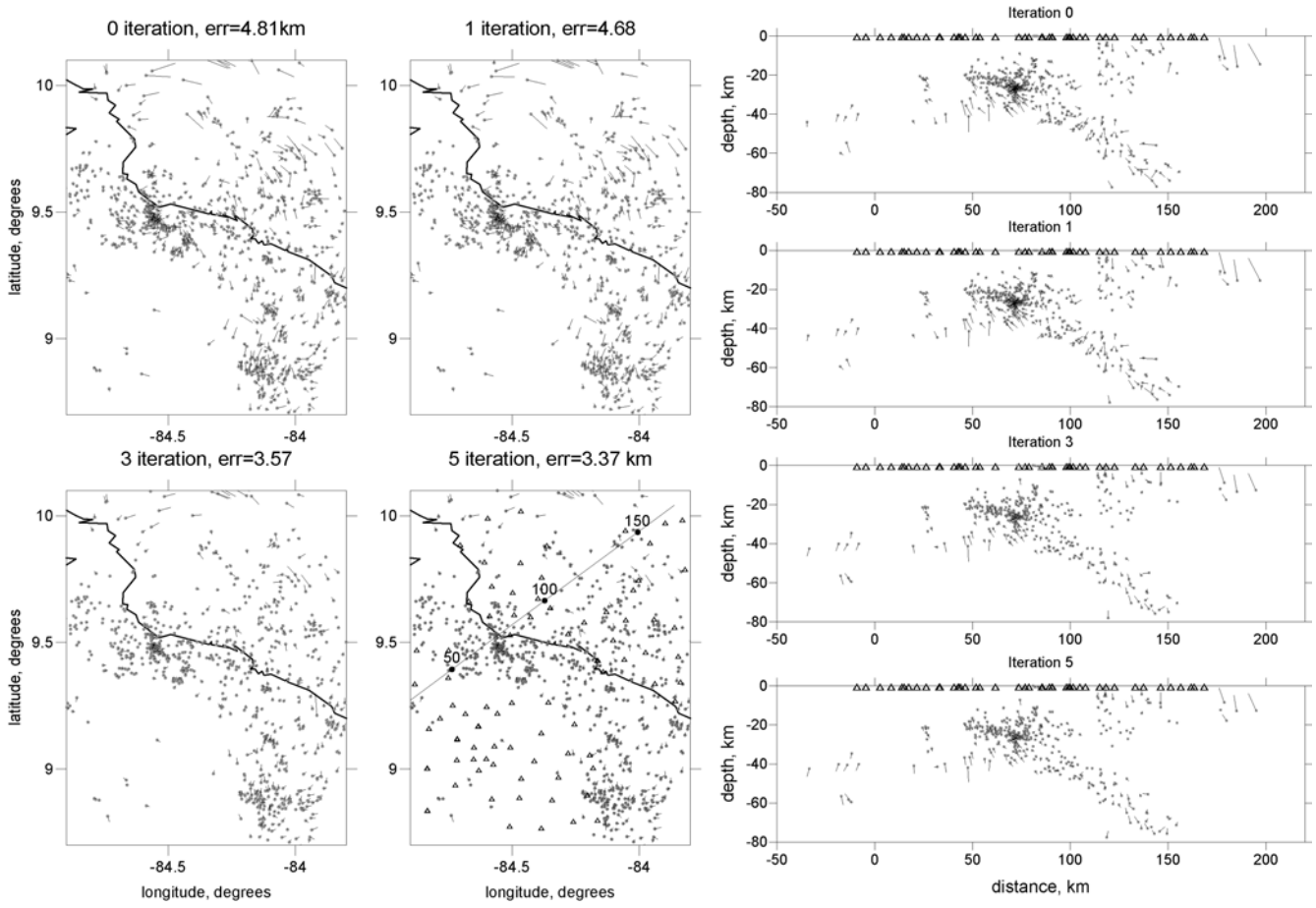

Figure 6. Events after preliminary location (iteration 0) and one, three, and five iterations presented in map view (left-hand panels) and in a cross section (right-hand panels). Gray dots are the resulting locations; bars are the mean errors with respect to true locations. Triangles show the stations in map view and the stations projected to the profile. The location of the profile is indicated in the map corresponding to iteration 5.

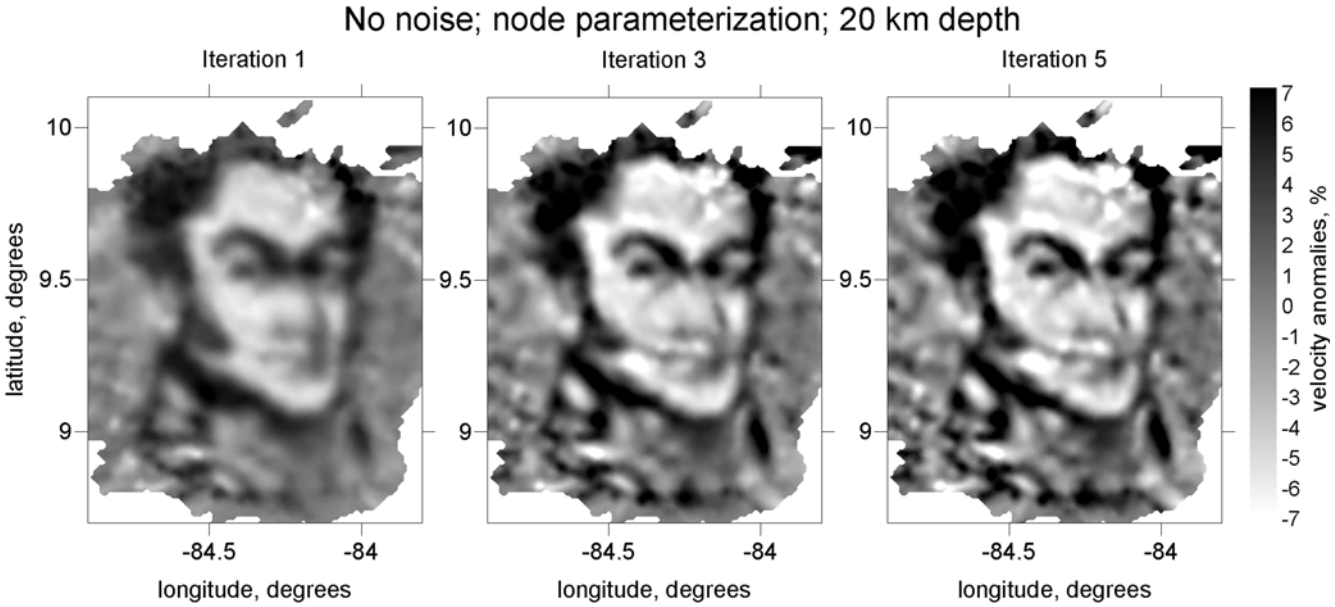

Figure 7. Reconstruction of the Bolivar synthetic model using node parameterization. The results after one, three, and five iterations are presented. *P*-velocity anomalies with respect to the optimized 1D-velocity model (gray curve in the upper panel of Fig. 2) are presented for a depth of 20 km.

5 using the node parameterization. Increasing the iteration number causes the amplitude to increase and instabilities to appear. The inversion allows reconstruction of even very fine features, such as lines of the nose and mouth with thicknesses less than 10 km.

For inverting the noisy dataset (Fig. 8), the resolution appears to be lower. When the matrix was inverted, we had to increase the smoothing coefficient from 0.3 to 0.5 in order to reduce the effect of noise. As a result, the reconstruction of the model in Figure 8 is not as clear as in Figure 7. However, features greater than 15 km size are retrieved stably. The values of residual rms before and after inversion for both clean and noisy data are presented in Table 1, in rows 1 and 2. In general, the noisy data represent similar values of variance reduction as are observed after real data processing.

As was already mentioned, the LOTOS-07 algorithm allows inversions to be performed using two different parameterization methods: with nodes or with cells. Results of inversions of clean and noisy datasets with the use of cell parameterization after iterations 1, 3, and 5 are presented in Figures 9 and 10. It should be noted that the sizes of the cells

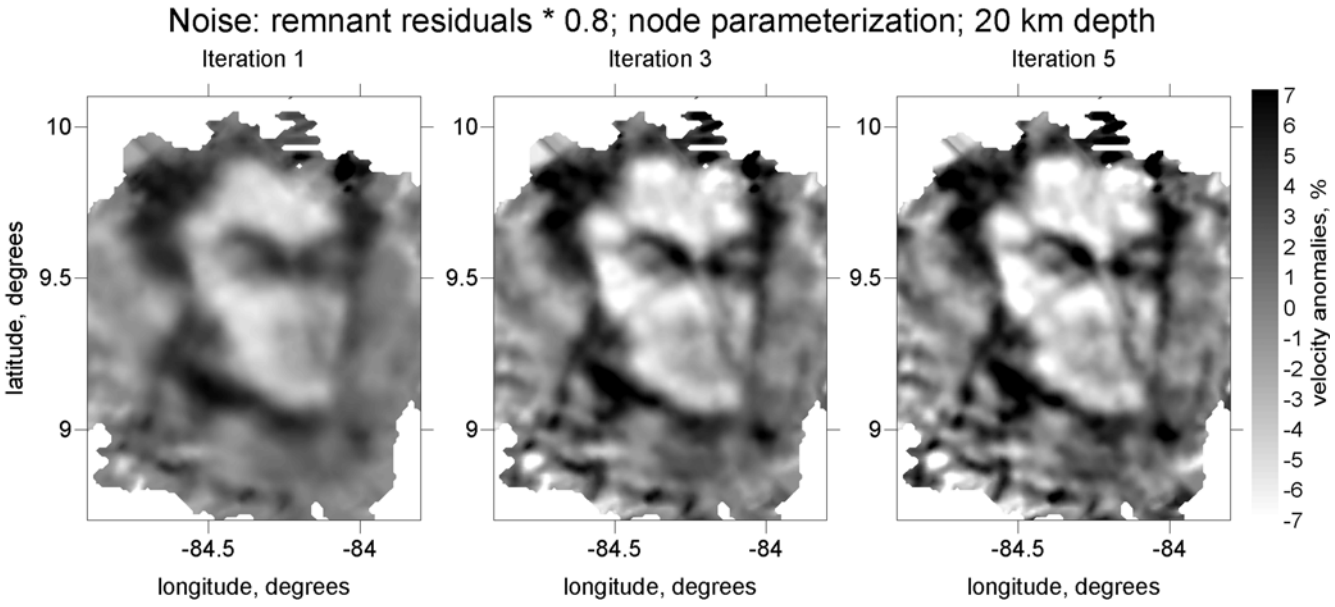

Figure 8. The same as in Figure 7 but with the noisy dataset.

Table 1

Values of rms of Residuals Before and After the Inversion for Two Starting Reference Models and Using the Node and Cell Parameterizations

Bolivar Synthetic Model	rms (sec) *P* Residuals, 1 Iteration	ms (sec) *S* Residuals, 1 Iteration	rms (sec) *P* Residuals, 5 Iterations	ms (sec) *S* Residuals, 5 iterations
Reference model 1, nodes, no noise	0.139	0.197	0.058	0.076
Reference model 1, nodes, real noise	0.181	0.248	0.121	0.144
Reference model 2, nodes, no noise	0.150	0.212	0.078	0.109
Reference model 1, cells, no noise	0.139	0.197	0.065	0.088
Reference model 1, cells, real noise	0.181	0.248	0.119	0.144
Reference model 2, cells, no noise	0.150	0.212	0.098	0.140

were similar to the node spacing in the examples discussed previously (Figs. 7 and 8). Cell parameterization generally provides a lower resolution than node parameterization. Table 1 shows that the data fit for cell parameterization is slightly worse than when using nodes. Cell parameterization provides less stable solutions toward noise than nodes. Figure 10 shows that the solution after iteration 5 is strongly perturbed by artifacts. Based on the comparison of these two parameterization methods, we conclude that node parameterization appears to be preferable for tomographic inversion with regard to stability and resolution.

As was mentioned in the Description of the LOTOS-07 Algorithm section, the final solution is obtained by stacking several models computed in grids with different basic orientations. In Figure 11 we present examples of solutions obtained for grids with 0° and 45° orientations for both node and cell parameterizations. Here, we present the results of inversion of the clean dataset after five iterations. For both parameterization methods, in the central parts of the study area, the solutions in the different grids are identical. Some minor differences are observed in the margins, but they disappear after stacking of the results.

We performed several runs of the programs using different starting 1D models. In Figure 12 we present the results based on the 1D model 2 (Fig. 2). Although this model seems to be closer to the true 1D distribution, the data fit for the final result appears worse than for model 1. This paradox is related to the fact that for model 2, the 1D-distribution optimization produces a low-velocity layer that significantly changes the ray paths with respect to the true model. At the same time, the results of inversion obtained for starting model 2 (Fig. 12) seem to be similar to the results based on model 1, presented in Figure 7. This supports the statement mentioned in our previous studies that for tomographic inversion with unknown sources, the reconstruction of relative-velocity anomalies is much more stable than retrieving absolute velocities.

In the previous examples, we mostly explored the horizontal resolution. The vertical resolution in the LET schemes is usually worse. Figure 13 shows the results of reconstruction of two checkerboard models defined in a vertical section. The sizes of the anomalies in Figure 13a,c and 13b,d were 30×30 and 20×20 km, respectively. Across the profile, the anomalies remain unchanged in the band of

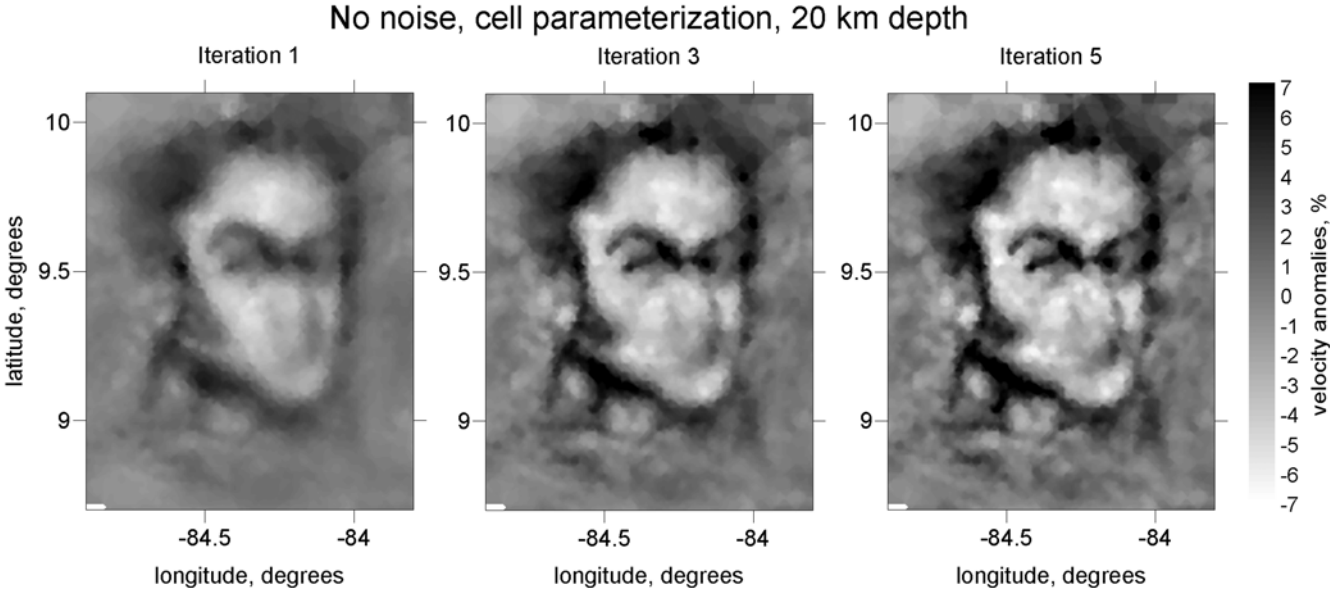

Figure 9. Reconstruction of the Bolivar synthetic model using cell parameterization. Results after one, three, and five iterations are presented. *P*-velocity anomalies with respect to the optimized 1D-velocity model (gray curve in upper panel of Fig. 2) are presented for a depth of 20 km.

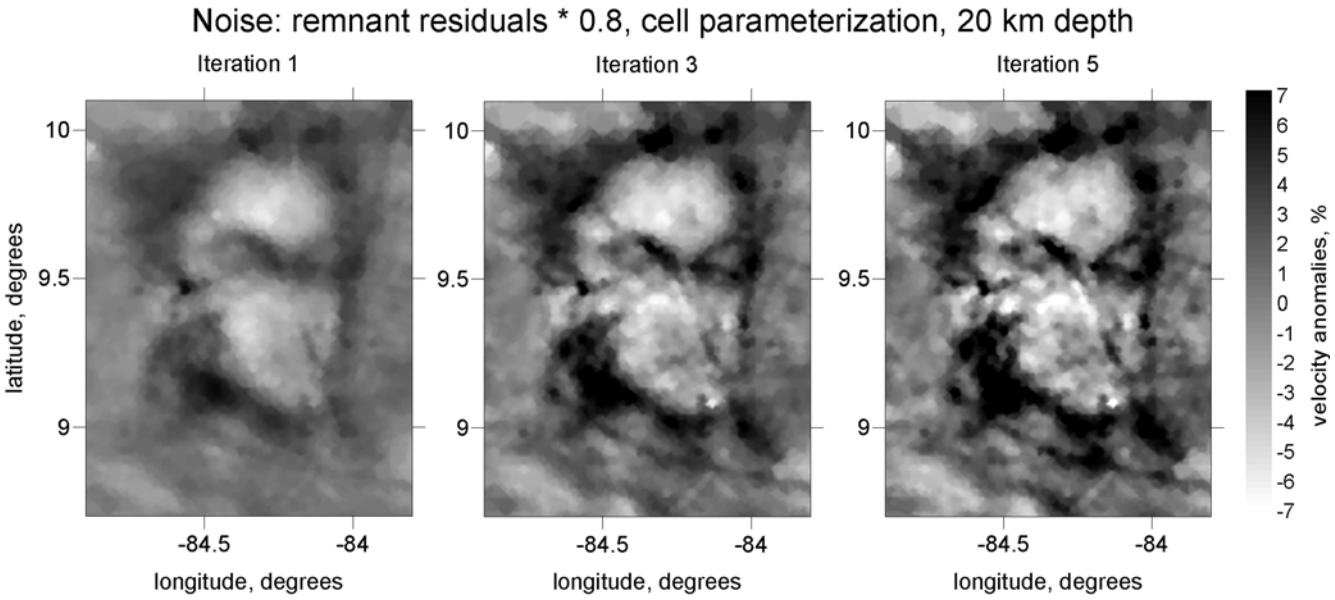

Figure 10. The same as in Figure 9 but with the noisy dataset.

±90 km. The values of the anomalies in both cases are ±7%. These synthetic tests were performed for the same source–receiver configuration and in exactly the same way as in the case of previously described Bolivar tests. The noise for these synthetic datasets was produced by a generator of random numbers with the same histogram shape as real residuals. The rms of *P* and *S* errors were 0.15 and 0.20 sec, respectively. In the presented model, we use node parameterization. The reconstruction results show that the algorithm can correctly retrieve the interface of sign change at depths of 25 and 15 km for VER_BRD_1 and VER_BRD_2, respectively. However, the general configuration of the retrieved anomalies is less clear than for horizontal resolution testing.

Benchmark for Testing the Tomographic Algorithms

The main purpose of any tomographic study is not to show maps and sections with images of seismic structures but to convincingly argue that these images represent real features in the Earth. As a rule, tomographic studies are supplemented with various tests, and the synthetic reconstructions usually take a special part. Dataset creation and reconstructions of synthetic models are usually performed by the same authors, and tests are often described in passing, without details. This leaves the reader with no way to determine whether the test was performed correctly or not.

The basic assumption of synthetic modeling should be maximum relevancy to real data. In general, we start processing real data with only two data files: station coordinates and arrival times of seismic rays. Neither sources nor velocity model are presumed to be known. Synthetic modeling should use the same data types and follow the same procedure as in real data processing. Analyzing hundreds of different tomography studies (it makes no sense to refer to all of them here), we concluded that in some cases the synthetic modeling does not adequately reflect the processing of real data. We single out the following major problems:

1. In some algorithms (see, e.g., Chiarabba and Amato, 2003; Husen and Smith, 2004; Dias *et al.*, 2007) the synthetic model is defined in the same cells that are used for parameterization (e.g., the checkerboard is defined in the alternating parameterization cells/nodes). In this case the model is predefined in a most convenient way for the inversion, and the reconstruction results look much nicer than the real resolution capacity of the algorithm. For example, if a checkerboard with 20 km size patterns is resolved with 20 km size cells, it is claimed that 20 km is a minimal size that can be resolved with this inversion. However, it is easy to predict, that larger sizes of patterns, if they are not divisible by 20 km (e.g., 23 km), would hardly be resolved with the same grid of 20 km spacing. Similarly, the reconstruction would probably fail if the parameterization grid is shifted to a half step with respect to the synthetic patterns. This problem is discussed in more details in Appendix A.
2. In many studies, real data inversion starts with optimization of the 1D model, but for some reason, in synthetic modeling, this step is forgotten. In most cases, the same reference model that is used to calculate the synthetic data is implemented for the synthetic inversion. This is obviously inadequate for the real situation when the reference 1D model is unknown. Correct synthetic reconstruction should start with a different 1D model other than the true one.
3. In many cases, after computing the synthetic times, the authors do not forget about the locations and origin times

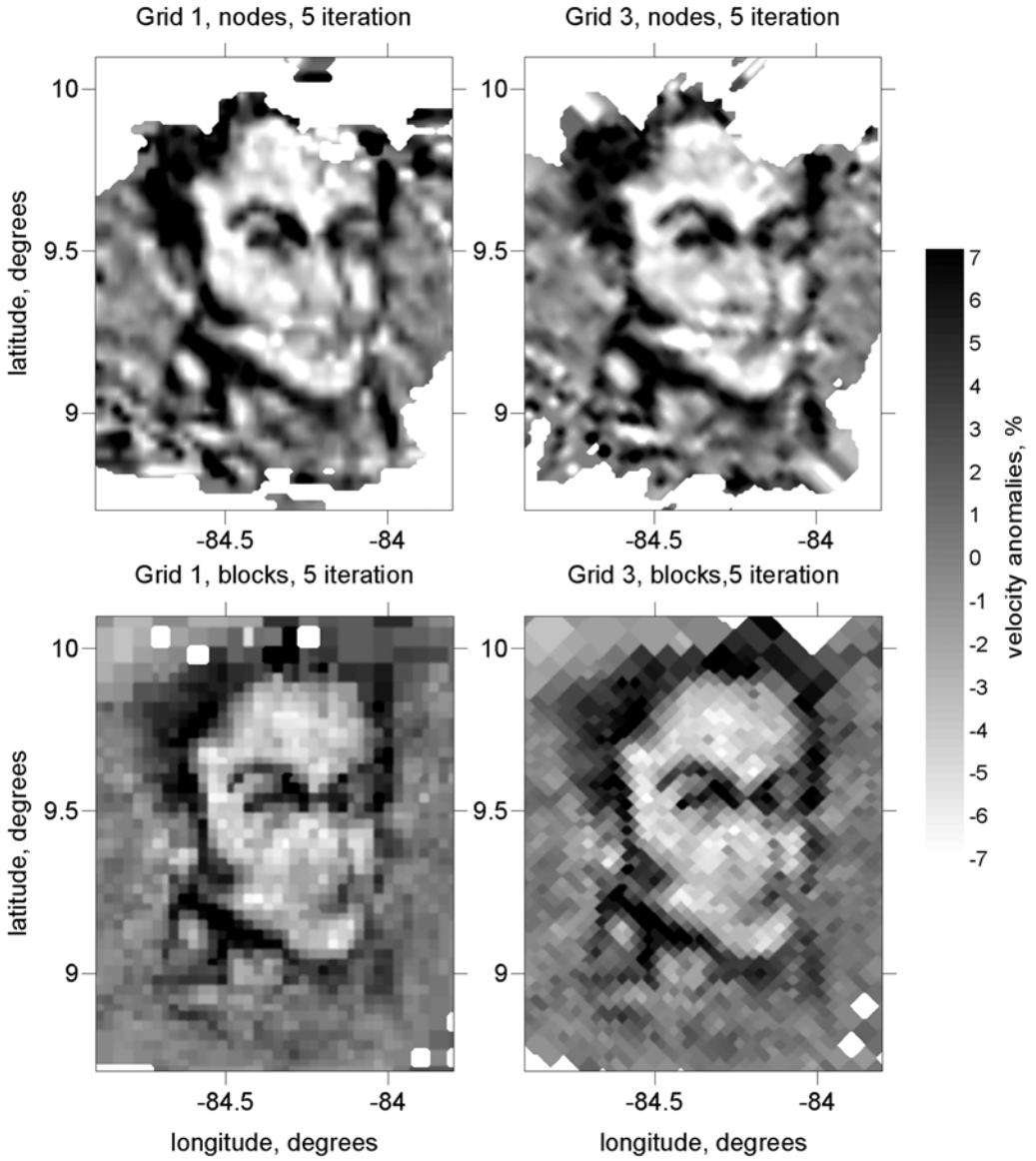

Figure 11. Reconstruction results for two differently oriented parameterization grids (0° and 45°) for node (upper) and cell (lower) types of parameterization. *P*-velocity anomalies after 5 iterations are presented for a depth of 20 km.

of the sources. In these studies, the synthetic reconstruction starts from the same source parameters as in forward modeling. Even if the procedure presumes coupled inversion, the source would hardly shift from its true location. In other words, if we start from true source locations, it means that half of the job for the synthetic modeling is done, and the problem becomes much easier than the real case. Therefore, it is more suitable to start the synthetic modeling without any information about sources, for example, placing them in an arbitrary point of the study area or using approximate locations in a 1D model.

In addition, two other requirements should be taken into account:

1. The data should be perturbed with noise with a realistic statistical distribution and amplitude. The level of noise must be defined according to the values of remnant residuals after the final iteration of the real data tomographic inversion. After performing the real and synthetic inversions, the level of remnant residuals should be similar.
2. Inversion parameters (smoothing coefficients, regularization level, iteration numbers, parameterization, etc.) should be identical in the real and synthetic cases.

A possibility for external control would be very advantageous. If the synthetic dataset and inversion are produced and performed by a single author, evaluation of the reconstruction quality is based on trust in the author that all the steps were accomplished correctly. In some cases, there is a strong temptation to make adjustments in order to make the results more attractive. It is obvious that checking the working ability of codes would be more unbiased and con-

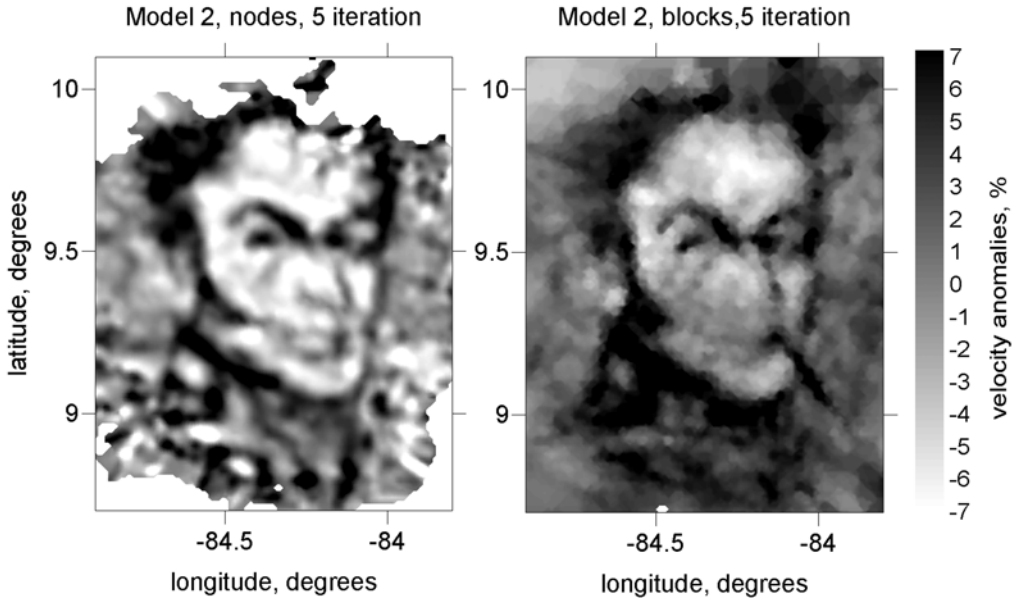

Figure 12. Reconstruction results with the node and cell parameterizations using another starting reference model (model 2, lower panel in Fig. 2). *P*-velocity anomalies after 5 iterations are presented for a depth of 20 km.

vincing if it were based on external datasets. In other words, there should be a series of standard tests, benchmarks, created by different authors, which can be used to test and compare tomographic codes.

In this study, we present our first contribution to this benchmark and encourage all authors and users of tomographic algorithms to take part in its updating. The rules for creating data should correspond to the requirements listed previously. First, these datasets should include only the station coordinates and arrival times. Any information about source parameters and velocity model should not be presented. Ideally, reconstruction of the benchmark should be performed blindly. The author of the dataset provides the information about the model and sources only after receiving the reconstruction results. In this case, comparison of the results and the model would show the efficiency of the algorithm for performing the real data inversion.

At the same time, we understand that some of the existing tomographic codes use preliminary catalog locations as input data. They do not perform absolute location of sources and optimization of the 1D-velocity model. In order to attract users of such codes to the benchmark testing, we have created several datasets with information about the sources, which were preliminarily located in the 1D-optimal model. For these datasets we also provide the 1D models that are slightly different from the true ones.

In our internet site, see the Data and Resources section, we provide several datasets that can be used by anyone to test tomographic codes according to the LET scheme. The datasets were produced by 3D raytracing, as described in the Bending Algorithm for Raytracing in a 3D-Velocity Model section, through different synthetic models. We present four datasets for different regions, data amounts, and model complexities: (1) Costa Rica (data with and without noise), (2) Turkey, (3) Toba (N. Sumatra), and (4) Chile (offshore area at 21° latitude).

Datasets 1, 2, and 3 correspond to real configurations of observation systems, while dataset 4 is a synthetic dataset that was used to plan the deployment of a seismic network. The dataset for each area contains two files: (1) a file with station geographic coordinates and (2) a file with initial locations of sources and travel times. In some of the presented datasets, coordinates and origin times of sources are not given. Instead of source coordinates, an arbitrary point (the center of the study area) is fixed. The uncertainty of the origin times is modeled by adding a random bias to all travel times from each source.

The Costa Rica datasets (with and without noise) are used as examples to illustrate the working ability of the LOTOS-07 code in this article. The initial models for these datasets, source distributions, and results are shown in Figures 5–12. Anyone can easily repeat the calculations for these datasets using the executable version of the LOTOS-07 code, which can be found at the Web site listed in the Data and Resources section. Following rather simple instructions helps the user execute the entire inversion procedure based on the same two files of initial information. After performing the inversion, the same images for the Simon Bolivar test as presented in this article will be obtained. Total execution time for processing this dataset (more than 1000 events and more than 30,000 picks) using a regular computer will take 3–5 hr. The same version of the LOTOS-07 code can be used to invert the other three datasets.

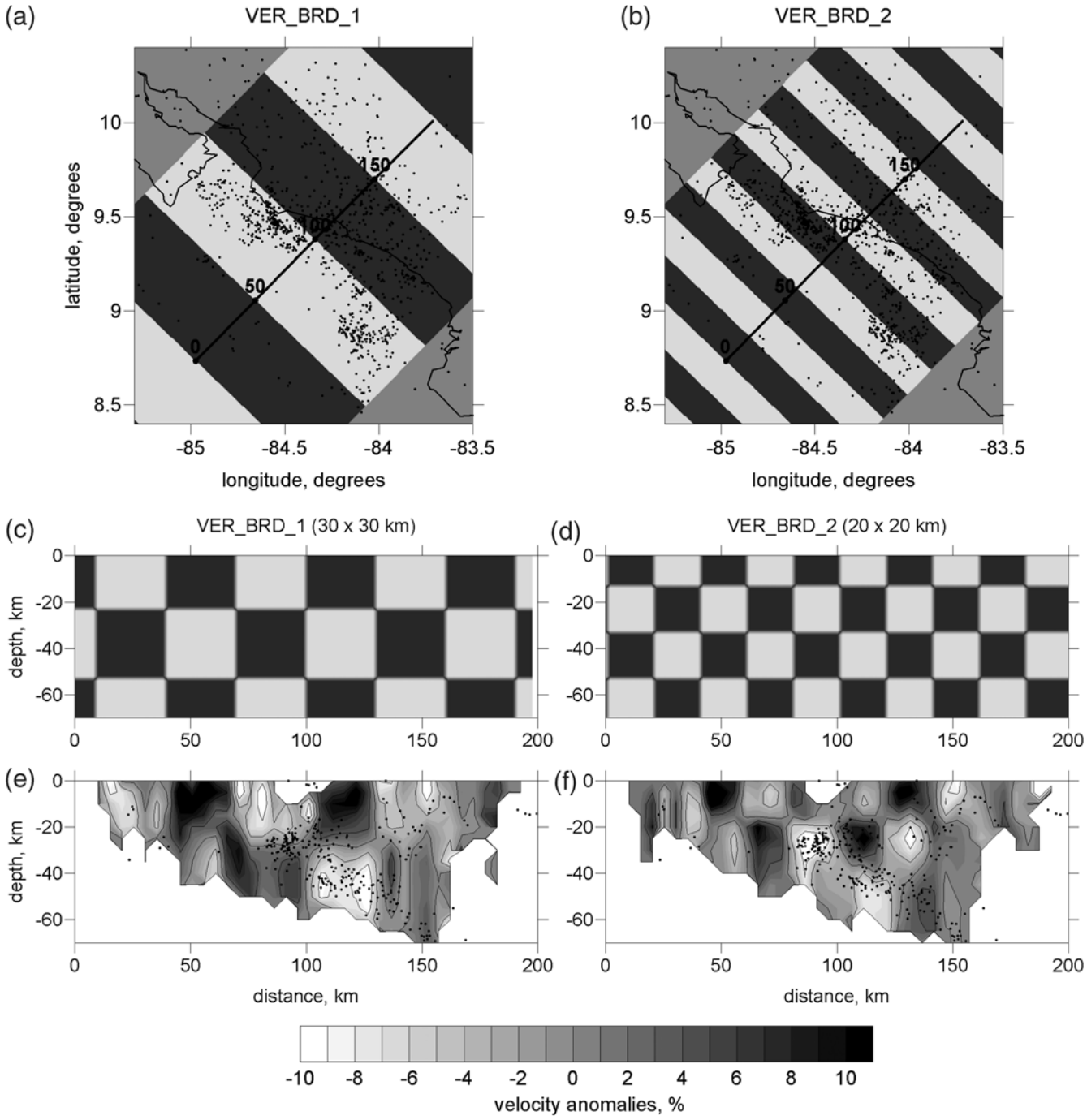

Figure 13. Testing the vertical resolution using the checkerboard model defined in a vertical section. The location of the profile used to define and visualize the model is shown in the upper maps. Two models with block sizes of 30×30 and 20×20 are presented. Synthetic models are shown in map view at 10 km depth (upper row) and in vertical section (middle row). Reconstruction results for these two cases in the same profile are shown in the lower row. Projections of the events located less than 40 km from the profile are shown with black dots. Contour lines indicate the levels of $\pm 4\%$ and $\pm 8\%$.

Conclusions

In this article, we present the most recent version of the LOTOS-07 algorithm, which can be freely downloaded from our Web site. Using the example datasets provided at the same Web site, one can easily produce one's own datasets according to the proposed format and correctly define the initial parameters. The calculation can be executed for any real and/or synthetic dataset. We will continue to work on developing the code and making it more user-friendly. We encourage all interested colleagues to regularly follow the changes on our Web site.

The main features of the LOTOS-07 algorithms with respect to other LET codes are the following:

1. The LOTOS-07 algorithm uses the minimal input information (arrival times and coordinates of stations), which is available just after the phase picking. It performs full processing starting with 1D model determination and initial source location and ending with iterative coupled inversion for 3D velocity and source location. In the case of using other codes, these steps are usually performed with different programs.
2. LOTOS-07 contains powerful tools for easily defining various synthetic models. The synthetic reconstructions use the same input files (arrival times and station coordinates) and the same program steps as in the case of the real data processing. This represents a more realistic situation than in the cases of using the data with known 1D model and sources, as done in many tomographic studies.
3. Parameterization with dense rotated grids makes the velocity model grid independent. In this sense, the parameterization used in LOTOS-07 can be considered as quasi continuous. The resolution is controlled by damping in the inversion step, and not by the grid spacing, as done in some other tomographic algorithms. Furthermore, LOTOS-07 allows one to choose between two parameterization methods, with grids or cells.
4. The LOTOS-07 code uses the original version of bending raytracing. An important feature of this algorithm is that it can use any description of the velocity distribution. It can be done with nodes, cells, polygons, or analytical laws. This code is faster than most of the existing analogs.
5. The LOTOS-07 code has a rather simple file structure and is easy for practical use.

We are ready to collaborate with all authors and users of different tomographic codes. In particular, for any configuration of the observation systems, we can produce synthetic models of any complexity and compute blind datasets using our algorithms. We can do this work for any colleagues who are interested in external testing of their algorithms.

We encourage everyone to contribute in producing their own synthetic datasets according to the rules listed previously. These datasets will be placed on our Web site and will be available for any interested party as a part of the benchmark.

We expect that such collaboration of different specialists in seismic tomography would be fruitful for improving the robustness of algorithms and results.

Data and Resources

In the article we present synthetic data based on real configuration of sources and receivers corresponding to a real experiment in Costa Rica (Dinc *et al.*, 2008).

The LOTOS-07 code, which is presented in this article, is freely available online at www.ivan-art.com/science/LOTOS_07 (last accessed November 2008). This Internet site provides detailed description of the code, manuals, listing of all FORTRAN codes, examples of real and synthetic datasets, and other necessary information. It can be used both by beginners (running executable files and changing the files and parameters) and advanced users, who can modify the source codes. Different options can be offered to any interested person through personal communication with the author of this article.

The synthetic datasets for the benchmark presented in this article are available at www.ivan-art.com/science/benchmark (last accessed November 2008). We invite all colleagues to make their contribution in creating open testing datasets. They will be placed on the same Internet site.

Acknowledgments

This study is supported by the Russian Foundation for Basic Research (RFBR) Grant Number 08-05-00276-a. We are very grateful to the Associate Editor Jeanne Hardebeck and two anonymous reviewers for friendly and constructive suggestions that have significantly improved the manuscript.

References

Benz, H. M., B. A. Chouet, P. B. Dawson, J. C. Lahr, R. A. Page, and J. A. Hole (1996). Three-dimensional *P* and *S* wave velocity structure of Redoubt Volcano, Alaska, *J. Geophys. Res.* **101,** 8111–8128.

Bijwaard, H., W. Spakman, and E. R. Engdahl (1998). Closing the gap between regional and global travel time tomography, *J. Geophys. Res.* **103,** 30,055–30,078.

Chiarabba, C., and A. Amato (2003). V_P and V_P/V_S images in the M_w 6.0 Colfiorito fault region (central Italy): a contribution to the understanding of seismotectonic and seismogenic processes, *J. Geophys. Res.* **108,** no. B5, 2248, doi 10.1029/2001JB001665.

Daly, E., D. Keir, C. J. Ebinger, G. W. Stuart, I. D. Bastow, and A. Ayele (2008). Crustal tomographic imaging of a transitional continental rift: the Ethiopian rift, *Geophys. J. Int.* **172,** 1033–1048.

Dias, N. A., L. Matias, N. Lourenço, J. Madeira, F. Carrilho, and J. L. Gaspar (2007). Crustal seismic velocity structure near Faial and Pico Islands (AZORES), from local earthquake tomography, *Tectonophysics* **445,** 301–317.

Dinc, A. N., I. Koulakov, M. Thorwart, W. Rabbel, E. Flueh, and I. Arroyo (2008). A combined tomographic inversion of two independent amphibious networks in Central Costa Rica, *Geophys. J. Int.* (in press).

Dorbath, C., and F. Masson (2000). Composition of the crust and uppermantle in the Central Andes (19.5° S) inferred from *P* wave velocity and Poisson's ratio, *Tectonophysics* **327,** 213–223.

Eberhart-Phillips, D. (1986). Three-dimensional velocity structure in northern California Coast Ranges from inversion of local earthquake arrival times, *Bull. Seismol. Soc. Am.* **76,** 1025–1052.

Eberhart-Phillips, D., and S. Bannister (2002). Three-dimensional crustal structure in the Southern Alps region of New Zealand from inversion of local earthquake and active source data, *J. Geophys. Res.* **107,** no. B10, 2262, doi 10.1029/2001JB000567.

Eberhart-Phillips, D., and A. J. Michael (1998). Seismotectonics of the Loma Prieta, California, region determined from three-dimensional V_P, V_P/V_S, and seismicity, *J. Geophys. Res.* **103,** 21,099–21,120, doi: 10.129/98JB01984.

Graeber, F. M., and G. Asch (1999). Three-dimensional models of *P* wave velocity and *P*-to-*S* velocity ratio in the southern central Andes by simultaneous inversion of local earthquake data, *J. Geophys. Res.* **104,** 20,237–20,256.

Haslinger, F., E. Kissling, J. Ansorge, D. Hatzfeld, E. Papadimitriou, V. Karakoastas, K. Makropoulos, H. G. Kahle, and Y. Peter (1999). 3D crustal structure from local earthquake tomography around the Gulf of Arta (Ionian region), NW Greece, *Tectonophysics* **304,** 201–218.

Hauksson, E. (2000). Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California, *J. Geophys. Res.* **105,** 13,875–13,903.

Hauksson, E., and J. S. Haase (1997). Three-dimensional V_P and V_P/V_S velocity models of the Los Angeles basin and central Transverse Ranges, California, *J. Geophys. Res.* **102,** no. B3, 5423–5454.

Hole, J. A., T. M. Brocher, S. L. Klemperer, T. Parsons, H. M. Benz, and K. P. Furlong (2000). Three-dimensional seismic velocity structure of the San Francisco Bay area, *J. Geophys. Res.* **105,** 13,859.

Husen, S., and R. B. Smith (2004). Probabilistic earthquake relocation in three-dimensional velocity models for the Yellowstone National Park region, Wyoming, *Bull. Seismol. Soc. Am.* **94,** 880–896.

Husen, S., E. Kissling, and E. R. Flueh (2000). Local earthquake tomography of shallow subduction in north Chile: a combined onshore and offshore study, *J. Geophys. Res.* **105,** 28,183–28,198.

Husen, S., R. Quintero, and E. Kissling (2002). Tomographic evidence for a subducted seamount beneath the Gulf of Nicoya: the cause of the 1990 $M_w = 7.0$ Gulf of Nicoya earthquake, *Geophys. Res. Lett.* **29,** no. 8, 1238, doi 10.1029/2001GL014045.

Husen, S., R. Quintero, E. Kissling, and B. Hacker (2003). Subduction-zone structure and magmatic processes beneath Costa Rica constrained by local earthquake tomography and petrological modeling, *Geophys. J. Int.* **155,** 11–32.

Kissling, E., W. Ellsworth, D. Eberhart-Phillips, and U. Kradolfer (1994). Initial reference models in local earthquake tomography, *J. Geophys. Res.* **99,** 19,635–19,646.

Koulakov, I., and S. Sobolev (2006). Moho depth and three-dimensional *P* and *S* structure of the crust and uppermost mantle in the Eastern Mediterranean and Middle East derived from tomographic inversion of local ISC data, *Geophys. J. Int.* **164,** no. 1 218–235.

Koulakov, I., M. Bohm, G. Asch, B.-G. Lühr, A. Manzanares, K. S. Brotopuspito, Pak Fauzi, M. A. Purbawinata, N. T. Puspito, A. Ratdomopurbo, H. Kopp, W. Rabbel, and E. Shevkunova (2007). *P* and *S* velocity structure of the crust and the upper mantle beneath central Java from local tomography inversion, *J. Geophys. Res.* **112,** B08310, doi 10.1029/2006JB004712.

Koulakov, I., S. V. Sobolev, and G. Asch (2006). *P*- and *S*-velocity images of the lithosphere-asthenosphere system in the Central Andes from local-source tomographic inversion, *Geophys. J. Int.* **167,** 106–126.

Masturyono, R. McCaffrey, D. A. Wark, S. W. Roecker, Fauzi, G. Ibrahim, and Sukhyar (2001). Distribution of magma beneath Toba Caldera, North Sumatra, Indonesia, Constrained by 3-dimensional *P*-wave velocities, seismicity, and gravity data, *Geochem. Geophys. Geosys.* **2,** no. 4, 1014, doi 10.1029/2000GC000096.

Nakajima, J., T. Matsuzawa, A. Hasegawa, and D. Zhao (2001). Three-dimensional structure of V_P, V_S, and V_P/V_S beneath northeastern Japan: implication for arc magmatism and fluids, *J. Geophys. Res.* **106,** 21,843–21,857.

Nolet, G. (1981). Linearized inversion of (teleseismic) data, in *The Solution of the Inverse Problem in Geophysical Interpretation*, R. Cassinis (Editor), Plenum Press, New York, 9–37.

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

Paul, A., M. Cattaneo, F. Thouvenot, D. Spallarossa, N. Bethoux, and J. Frechet (2001). A three-dimensional crustal velocity model of the south-western Alps from local earthquake tomography, *J. Geophys. Res.* **106,** no. B9, 19,367–19,389.

Ramachandran, K., S. E. Dosso, G. D. Spence, R. D. Hyndman, and T. M. Brocher (2005). Forearc structure beneath southwestern British Columbia: a three-dimensional tomographic velocity model, *J. Geophys. Res.* **110,** B02303, doi 10.1029/2004JB003258.

Reyners, M., D. Eberhart-Phillips, and G. Stuart (1999). A three dimensional image of the shallow subduction: crustal structure of the Raukumara Peninsula, New Zealand, *Geophys. J. Int.* **137,** 873–890.

Roecker, S. W., T. M. Sabitova, L. P. Vinnik, Y. A. Burmakov, M. I. Golvanov, R. Mamatkanova, and L. Munirova (1993). Three-dimensional elastic wave velocity structure of the western and central Tien Shan, *J. Geophys. Res.* **98,** 15,779–15,795.

Roecker, S., C. Thurber, and D. McPhee (2004). Joint inversion of gravity and arrival time data from Parkfield: new constraints on structure and hypocenter locations near the SAFOD drill site, *Geophys. Res. Lett.* **31,** L12S04, doi 10.1029/2003GL019396.

Schurr, B. (2001). Seismic structure of the Central Andean subduction zone from local earthquake data, *PhD Thesis*, GeoForschungsZentrum Potsdam.

Thurber, C. (1983). Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California, *J. Geophys. Res.* **88,** 8226–8236.

Thurber, C. (1993). Local earthquake tomography: velocities and V_P/V_S—theory, in *Seismic Tomography, Theory and Practice*, H. Iyer and K. Hirahara (Editors), CRC Press, Boca Raton, Florida, 563–583.

Thurber, C. H., S. R. Atre, and D. Eberhart-Phillips (1995). Three-dimensional V_P and V_P/V_S structure at Loma Prieta, California, *Geophys. Res. Lett.* **22,** 3079–3082.

Um, J., and C. H. Thurber (1987). A fast algorithm for two-point seismic ray tracing, *Bull. Seismol. Soc. Am.* **77,** 972–986.

van der Hilst, R. D., and E. R. Engdahl (1991). On ISC PP and pP data and their use in delay-time tomography of the Caribbean region, *Geophys. J. Int.* **106,** 169–188.

van der Sluis, A., and H. A. van der Vorst (1987). Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in *Seismic Tomography*, G. Nolet (Editor), Reidel, Dordrecht, 49–83.

Wagner, L. S., S. Beck, and G. Zandt (2005). Upper mantle structure in the south central Chilean subduction zone (30° to 36°S), *J. Geophys. Res.* **110,** B01308, doi 10.1029/2004JB003238.

Wagner, D., I. Koulakov, W. Rabbel, B.-G. Luehr, A. Wittwer, H. Kopp, M. Bohm, and G. Asch, and MERAMEX Scientists (2007). Joint inversion of active and passive seismic data in Central Java, *Geophys. J. Int.* **170,** no 2, 923–932, doi 10.1111/j.1365-246X.2007.03435.x.

Wang, C.-Y., W. W. Chan, and W. D. Mooney (2003). Three-dimensional velocity structure of crust and upper mantle in southwestern China and its tectonic implications, *J. Geophys. Res.* **108,** no. B9, 2442, doi 10.1029/2002JB001973.

Yang, T., and Y. Shen (2005). *P*-wave velocity structure of the crust and uppermost mantle beneath Iceland from local earthquake tomography, *Earth Planet. Sci. Lett.* **235,** 597–609.

Zhao, D., D. Christensen, and H. Pulpan (1995). Tomographic imaging of the Alaska subduction zone, *J. Geophys. Res.* **100,** 6487–6504.

Zhao, D., A. Hasegawa, and S. Horiuchi (1992). Tomographic Imaging of *P* and *S* wave velocity structure beneath northeastern Japan, *J. Geophys. Res.* **97,** 19,909–19,928.

Appendix A

About Defining the Synthetic Model

Synthetic tests must be presented in each tomographic study in order to show the realistic resolution capacity provided by the ray coverage and the algorithm. Definition of the model and the modeling setup should be as adequate for the real case as possible. In this sense, performing the synthetic modeling in some tomographic studies seems incorrect to me. One of the key points of synthetic modeling is defining the true model. In many studies, the synthetic model is defined in the same nodes/cells as used for inversion (e.g.,

Chiarabba and Amato, 2003; Wang, *et al.*, 2003; Husen and Smith, 2004; Dias *et al.*, 2007; and many others). We claim that such a predefinition of the model is the wrong way. Although in most of cases good reconstruction quality is demonstrated, these tests do not represent the realistic resolution.

To ground this statement, we present a series of tests in Figure A1 created using a simplified 2D tomographic code. In this code the sources are fixed, the rays are the straight lines, and the parameterization blocks are the rectangular cells with constant velocity. Starting velocity is constant. Despite these simplifications, this code is very useful to show

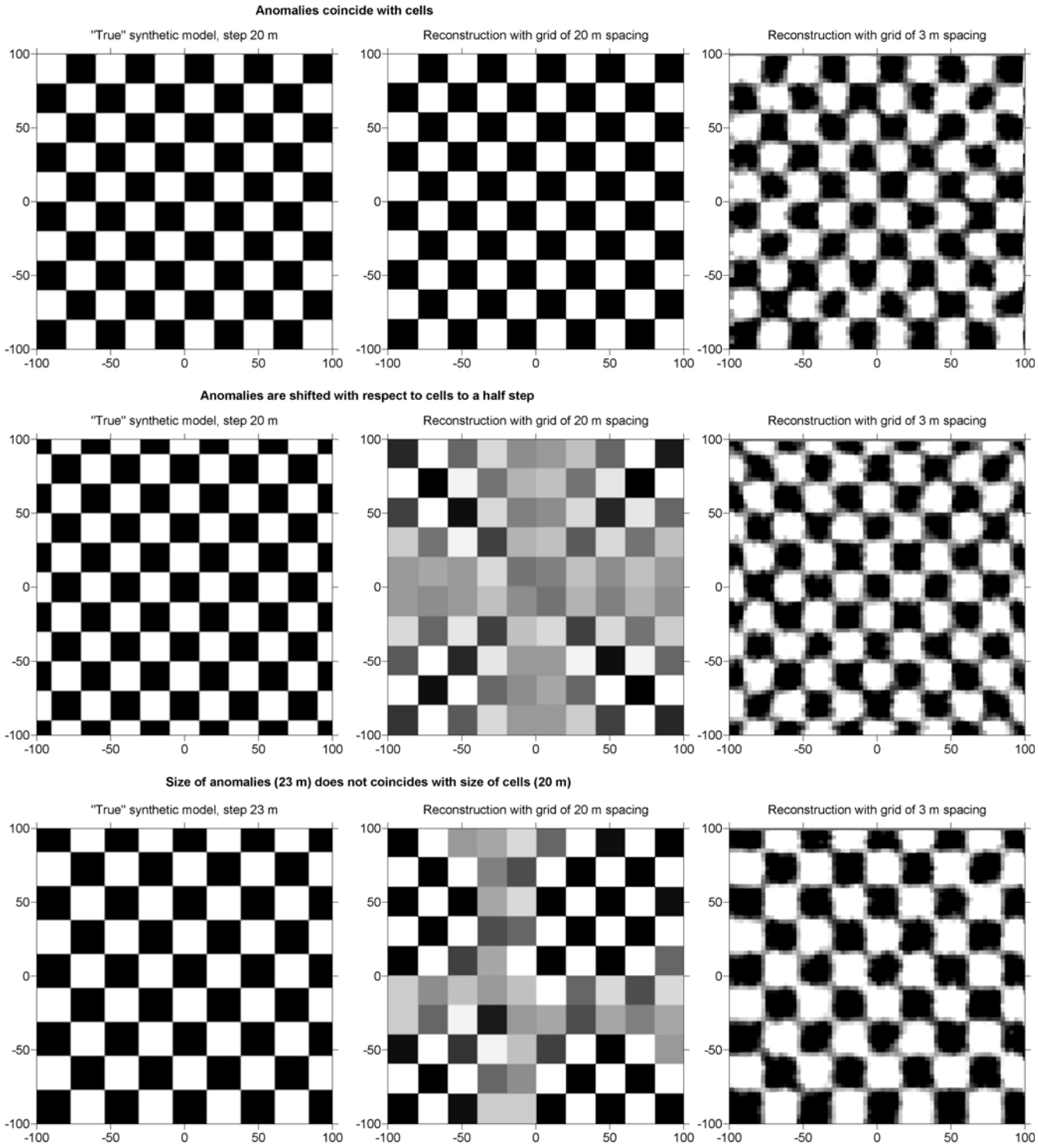

Figure A1. A series of tests performed using a simplified tomographic code. The left-hand column represent the true synthetic models. The middle and right-hand columns are reconstruction results using the cell spacing of 20 and 3 m, respectively.

some fundamental effects of tomographic inversion. The test in Figure A1 demonstrates the effect of the synthetic model predefinition according to the parameterization. In the upper row, the left-hand plot represents a true synthetic model with a checkerboard of 20 m spacing. The central plot is the result of reconstruction with the parameterization cells coinciding with synthetic patterns. This situation is a simplified representation of the modeling concept used in many of the previously mentioned tomographic studies, when the synthetic model is defined in the same nodes/cells as used for inversion. The reconstruction in this case is perfect and one could get the impression that patterns of 20 m size and larger are reliably resolved in this model. However, this is not true. The presented situation is not an adequate way for testing, and it does not represent an experiment for the real Earth where such predefinition is impossible. It is easy to check that by using a grid shifted with respect to the synthetic anomalies by a half step (middle row); the inversion fails and we do not obtain any reasonable result. The same is observed when the size of the anomalies is not divisible with the grid spacing (lower row): reconstructing the anomalies that are 23 m in size using a grid of 20 m, the spacing fails as well.

In LOTOS-07 we propose using a grid with spacing that is much smaller than the size of the expected anomalies. The right-hand column of Figure A1 presents the inversion results for the same synthetic models using the grid spacing of 3 m. Although in all these cases the synthetic model and parameterization grids are independent, the inversion provides reasonable results for all models.

It seems to be a matter of principle that definition of the synthetic model must be free of any link with the parameterization grid used for inversion. In this sense, the ideal would be a blind test with an unknown model, which is proposed in this article.

Appendix B

About Using the Trade-Off Curve for Defining Damping Parameters

A key problem of any tomographic study is ambiguity of determination of the damping parameters used for inversion. To formalize optimal damping searching, Eberhart-Phillips (1986) has proposed using TOCs, which presents the data misfit versus amplitude of the resolved model computed for different damping parameters. It was stated that damping parameters evaluated using TOCs provide realistic values of velocity anomalies in the recovered model. Presently TOC is used as a standard tool in most of tomographic studies. However, after inspecting dozens of articles and several Ph.D. theses, we are not convinced that TOC is really useful for tomographic inversion, at least as it is presented in most of articles.

Here is a typical description of using TOCs from Wagner *et al.* (2005): "The other method proposed by Eberhart-Phillips (1986) involves running the tomography for one iteration (1_IK) over a range of damping parameters and plotting data variance versus model variance. The ideal damping value is at the corner of the resulting L curve (2_IK), beyond which increases in model variance do not produce significant decreases in data variance, and vice versa. In our case, this method produces an ideal damping value of 20 for both the P and S and the P and V_P/V_S inversions. Changing this damping value ± 10 does little to change the basic patterns in the results, but it does affect the amplitude of the deviations. These amplitudes are also affected by the number of iterations used. In order to find the optimal number of iterations, we use our previously determined damping values over 10 iterations, and again plot data variance versus model variance, this time for each iteration. In our case, 2 iterations (3_IK) are optimal for both the P and S and P and V_P/V_S inversions." If this description is correct, there are, at least, three reasons why we do not believe in TOCs.

Reason 1

It is indicated that TOC is constructed based on the first iteration of inversion (1_IK). We have found only two studies where TOC presents the results after final iteration: a Ph.D. thesis by Schurr (2001) and our own work (Koulakov *et al.*, 2007). In other studies TOC is computed after first iteration, similarly as in the presented description. We claim that TOC computed in this way is probably not adequate for evaluation of the optimal damping parameters.

The amplitudes of velocity anomalies in the recovered model are controlled by, at least, two parameters: damping and number of iterations. It means that the similar amplitude of the anomaly can be achieved using weak damping in one iteration and strong damping in several iterations. In our strong opinion, using strong damping and many iterations is more preferable because moving with small steps minimizes the nonlinear effect in each iteration. TOC provides the estimate for optimal damping in the first iteration. If we use two, five, or ten iterations, we would obtain different amplitudes of the solution. Which of them is closer to the reality? TOCs computed in the first iteration cannot provide the answer.

As for exploring the curve data variance versus iterations (3_IK) it does not seem to be an adequate argument. In our experience, the iterative inversion always provides the largest changes of the model and data fit after the first iteration (e.g., fig. 5a in Koulakov *et al.*, 2007 and Fig. B1c of this study). Each subsequent iteration provides smaller changes than a previous one, regardless the value of damping. Thus, the corner of the L-shaped curve that corresponds to the best number of iterations will be always in steps 1 or 2.

Reason 2

We do not like the visual analysis of L-shaped curvature of TOCs (2_IK). For example, looking at the curves presented

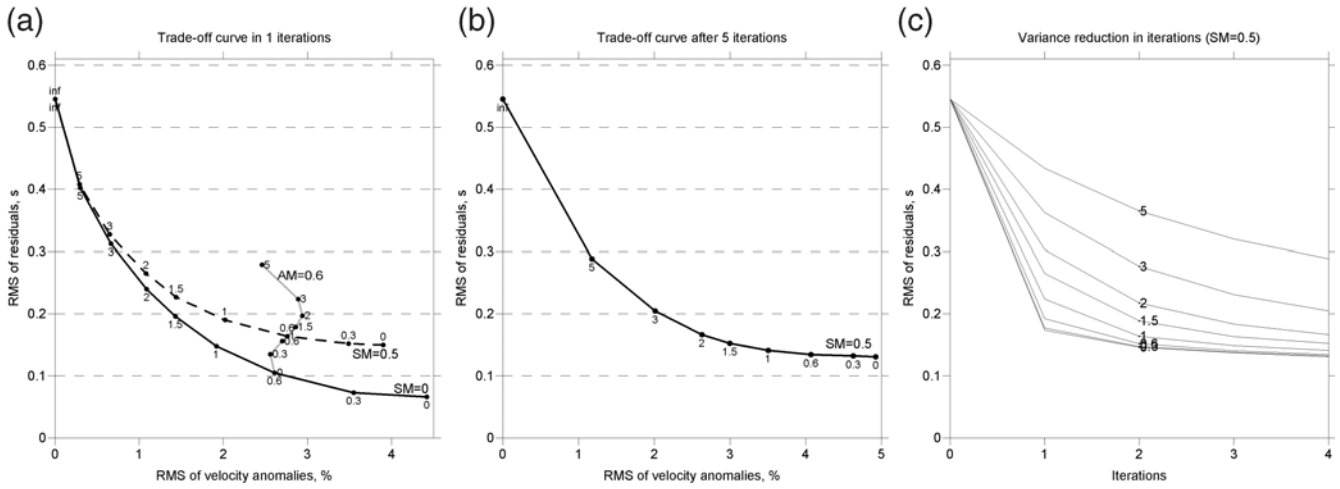

Figure B1. Different formal parameters corresponding to the synthetic model presented in Figure B2. (a) The TOC just after inversion in first iteration (without raytracing). Solid and dashed curves represent the TOC corresponding to fixed values of smoothing: SM = 0 and SM = 0.5, respectively. Numbers indicate values of AM. The gray curve shows the TOC for variable smoothing (SM, indicated with numbers) and fixed AM = 0.6. (b) The TOC computed for the fifth iteration using a fixed value of smoothing, SM = 0.5. (c) Values of variance reduction after raytracing in four iterations for different values of AM (indicated with numbers) and fixed smoothing, SM = 0.5.

in figure 5c in Koulakov *et al.* (2007), it is hard to define which of the values are optimal. For the *P* model (blue line) the curve is almost linear, and 1.5, 1, and 0.75 seem to be equally good. For the *S* model (red line) the situation is similar. It is hard to select the best value between 4, 3, 2, and 1.5.

Reason 3

Even if we found a correct damping, it is not always helpful in finding the true amplitudes of velocity anomalies in the entire area. The problem is that the solution amplitude depends very much on the data coverage. It is especially dramatic in regional studies where the ray density varies significantly in different parts of the study area. In this case, one damping parameter can provide over and underdamped solutions in parts with lower and higher ray coverage, respectively, within one model. As a result, the amplitude of the velocity anomalies might be lower than the real value in one part and higher in another part.

Furthermore, in the case of significant noise in the data, the optimal damping is controlled by the noise level. In this case, the solution is stable only if damping is sufficiently strong, and the resulting velocity anomalies can be weaker in amplitude that the real ones.

To illustrate some of the previous statements about TOCs, we have performed a synthetic test that is presented in Figure B2. Some formal parameters for this test are presented in Figure B1. The dataset for this test was generated artificially. We consider a seismic network consisting of 50 stations (plot A1) placed in a roughly isometric area of 200 km. 148 events were distributed randomly in the depth interval of 0–100 km (plot A2). For each event the number of recorded *P* picks varied randomly from 25 to 50. The *S* picks were generated for 40%–80% of the existing *P* picks. In total, 8597 pairs of source–receivers have been produced (5525 *P* and 3244 *S* rays). The synthetic model contains two positive (+8%) and two negative (−8%) anomalies defined inside ameba shaped polygons (plot A1), which remain unchanged in all depths. The synthetic travel times were computed using 3D raytracing and were perturbed by random noise of 0.1 and 0.15 sec rms for *P* and *S* data, respectively. The reconstruction has been performed using the LOTOS-07 code in the way described in the Synthetic Modeling section. Here we present the results for the *P* model only; for the *S* model the results are similar.

It should be noted that besides the number of iterations and amplitude damping (AM) in LOTOS_07, we have another parameter, smoothing coefficient (SM), which controls the properties of the solution. In Figure B1 we present different TOCs with fixed AM or SM corresponding to the first iteration just after inversion (without performing raytracing). This is a way for presenting TOCs, which is used in most of tomographic studies. It can be seen, that if we fix SM and consider different values of AM, the TOCs for SM = 0 (solid curve) andSM = 0.5 (dashed curve) have similar hyperbola-like shapes. If we use a criterion of finding a corner of an L-shaped curve, the best damping value would correspond to AM = 1.5 for both values of smoothing. In Figure B2 in plots C1–C3 and D1–D3, we present the inversion results corresponding to this value and SM = 0.5 and SM = 0, respectively. In both cases the amplitude of the solution is much lower that the true values, that means that AM = 1.5 provides overdamped solutions.

In Figure B1b we present the TOCs corresponding to fifth iteration and fixed smoothing, SM = 0.5. It can be seen that in this curve the corner point corresponds to the value of AM = 3. It is clear that if the value of AM = 1.5 provides the overdamped solution, the value of AM = 3 would result in a worse model with lower amplitudes of anomalies.

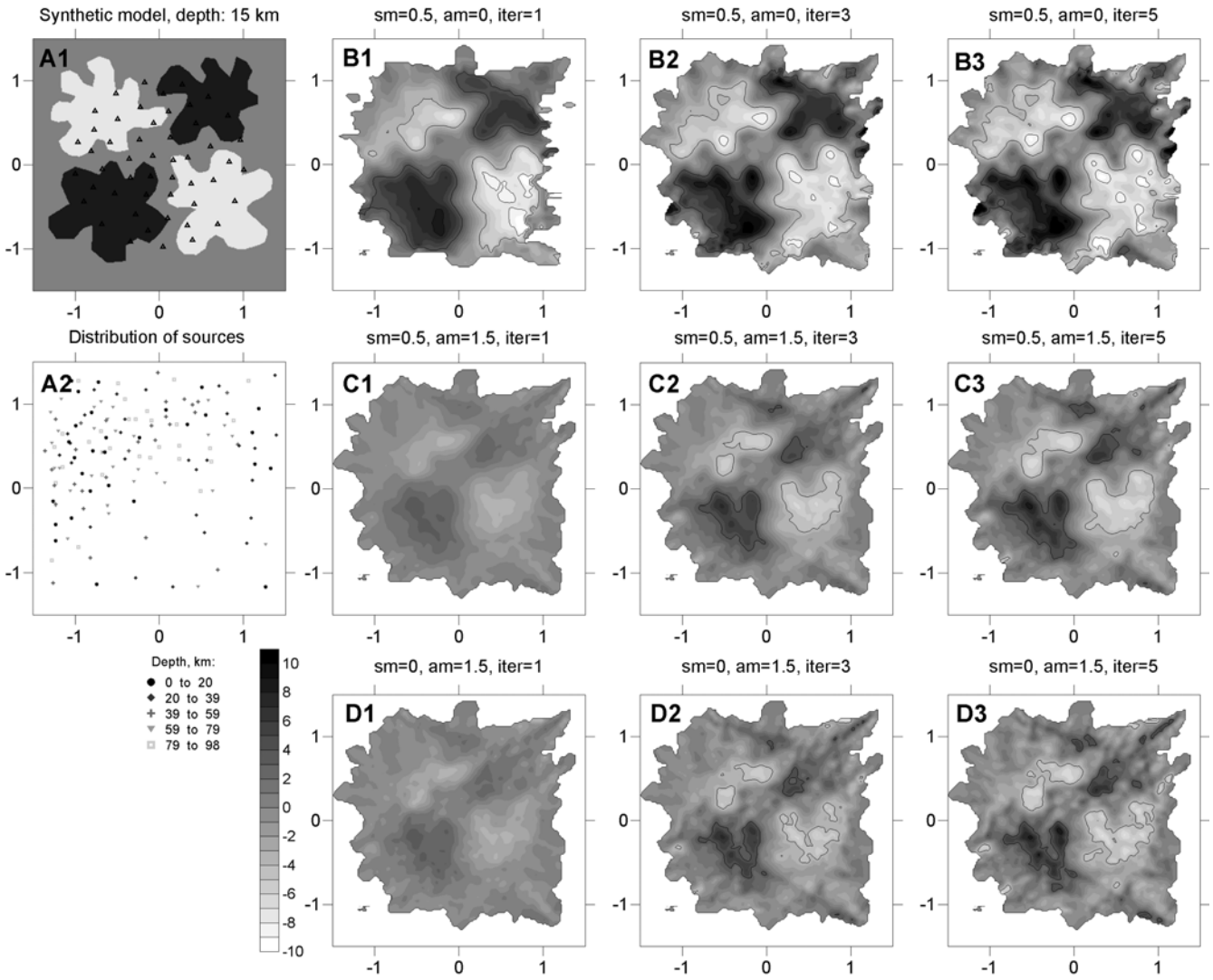

Figure B2. Synthetic model for testing the efficiency of using the TOCs. A1 is a distribution of true synthetic anomalies, which remain unchanged in all depths. The triangles depict the stations. A2 is a distribution of events. Different symbols indicate the depths of th eevents. B1–B3 are the reconstruction results in one, three, and five iterations, respectively, using smoothing SM = 0.5 and AM = 0. C1–C3 represent the results with AM = 1.5 and SM = 0.5. D1–D3 represent the results with SM = 0 and AM = 1.5. Contour lines indicate the levels of $\pm 4\%$ and $\pm 8\%$.

According to our estimations, the best solution that shows the most realistic amplitudes of anomalies (plots B1–B3 in Fig. B2) is achieved with AM = 0 and SM = 0.5. It means that for this case, the most realistic solution is obtained without AM, only using smoothing damping.

This test demonstrates that the method of finding the optimal damping using TOCs proposed by Eberhart-Phillips (1986) fails, at least for this test. First of all, the damping values corresponding to corner points of TOCs in the first and final iterations are not the same. If we use five iterations, it is more logical to use the value corresponding to the final iteration. Thus, the suggestion of Eberhart-Phillips (1986) of using damping found in the first iteration for the entire iterative procedure seems to be incorrect. Second, both corner points for first and fifth iterations do not provide the optimal damping values.

In LOTOS-07 we search for optimal values for smoothing and amplitude damping based on synthetic modeling. To do it correctly, we perform the synthetic modeling in the way that is described in the Synthetic Modeling section and shown in Figure B2. It should be noted, that we prefer playing more with SM than with AM. However, this is not a strict requirement and is just a matter of taste and case.

Institute of Petroleum Geology and Geophysics
SB RAS, Prospekt Akademika Koptuga, 3
Novosibirsk, 630090, Russia
KoulakovIY@ipgg.nsc.ru

Manuscript received 2 April 2008