



Parallel, 'large', dense matrix problems: Application to 3D sequential integrated inversion of seismological and gravity data

R. Tondi^{a,*}, C. Cavazzoni^b, P. Danecek^c, A. Morelli^a

^a Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Bologna, Via Donato Creti 12, 40128 Bologna, Italy

^b CINECA, Interuniversity Computing Centre, Via Magnanelli 6/3, 40033 Casalecchio di Reno (BO), Italy

^c Univ Granada, Inst Andaluz Geofis, E-18071 Granada, Spain

ARTICLE INFO

Article history:

Received 20 December 2011

Received in revised form

21 May 2012

Accepted 22 May 2012

Available online 31 May 2012

Keywords:

Parallel

Dense matrix

Block-cyclic distribution

Inverse problem

Probability density function

ScaLAPACK

Gravity field

Shear-wave velocity structure

Density structure

ABSTRACT

To obtain accurate and reliable estimations of the major lithological properties of the rock within a studied volume, geophysics uses the joint information provided by different geophysical datasets (e.g. gravimetric, magnetic, seismic). Representation of the different types of information entering the problem using probability density functions can provide the mathematical framework to formulate their combination. The maximum likelihood estimator of the resulting joint posterior probability density functions leads to the solution of the problem. However, one key problem appears to limit the use of this solver to an extensive range of real applications: information coming from potential fields that implies the presence of dense matrices in the resolving estimator. It is well known that dense matrix systems rapidly challenge both the algorithms and the computing platforms, and are not suited to high-resolution 3D geophysical analysis. In this study, we propose a procedure that allows us to obtain fast and reliable solutions of the joint posterior probability density functions in the presence of large gravity datasets and using sophisticated model parametrization. As it is particularly CPU-consuming, this 3D problem makes use of parallel computing to improve the performance and the accuracy of the simulations. Analysis of the correctness of the results, and the performance on different parallel environments, shows the portability and the efficiency of the code. This code is applied to a real experiment, where we succeed in recovering a 3D shear-wave velocity and density distribution within the upper mantle of the European continent, satisfying both the seismological and gravity data. On a multiprocessor machine, we have been able to handle forward and inverse calculations with a dense matrix of 215.66 Gb in 18 min, 20 s and 20 min, 54 s, respectively.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Geophysical exploration has developed a series of powerful tools for tomography of the Earth, which commonly consists of the inversion of a single geophysical dataset to estimate a geophysical-related property: e.g. compressional and transverse wave velocities from arrival times (Aki and Richards, 1980), conductivity from electromagnetic data (Newman and Alumbaugh, 1997), mass density from gravity data (Camacho et al., 1997), and magnetic susceptibility from magnetic data (Li and Oldenburgh, 1996). However, conventional geophysical tomographic methods face the problem of irregular data coverage over the surface of the studied volume, which can produce irregular image resolution. This problem is difficult to address for each isolated geophysical technique, and it demands an effort

for the integration of different geophysical methods into a single inversion scheme. Classically, integration can be carried out through the construction of independent models and the retention of the mutually compatible ones. This method is not always successful. Moreover, the search for a combined optimization requires that many models are constructed—a task that is costly and time-consuming. Alternatively, the inversion for a particular dataset can provide the initial model estimate for the inversion of a second dataset, or as a further possibility, we can perform a simultaneous least-squares inversion of two weighted datasets by perturbation of the related model parameters. For these approaches, (a) the structural or geological boundaries must be common to both inversions (Gallardo and Maju, 2004), or (b) a possible exact physical relationship is required to relate the different geophysical properties (Bosch, 1999): this can be quite challenging. Recently, due to the growth of computing power, methods of pure joint inversion of different geophysical datasets (Tiberi et al., 2003), as well as stochastic approaches that explore the whole model space (Moorkamp et al., 2010), have become

* Corresponding author. Tel.: +39 0514151439.

E-mail address: tondi@bo.ingv.it (R. Tondi).

more used. However, the often subjective process of finding an optimal regularization value of the different geophysical responses (Lines et al., 1988) in the first case, and the huge computational resources needed in the second case, can limit the effectiveness of these methods.

In this study, we present a fast and effective method to integrate the information given by seismological and gravity datasets. In particular, as we deal with a potential field, we face the problem of the presence of dense matrices in the resolving estimator. Dense matrices represent systems that are firmly coupled and that generally require a prohibitive amount of computer memory. A typical approach to solving dense systems is to use the LU factorization (Press et al., 1986) (each entry of the matrix is computed as an interaction between two boundary elements; often, many integrals must be computed) or Cholesky factorization (a decomposition of a symmetric, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose). However, these strategies are not well suited to large systems, as in many instances, the time required to compute the matrix is considerably larger than the time needed for the solution. As explained by Edelman (1993), the approaches towards large problem solving can be of three kinds: (a) develop numerical techniques to ensure that the best possible accuracy is obtained by users who want to trust the software; (b) get the highest possible performance out of some machine; and (c) exploit the structure inside the large dense matrices to develop new strategies of computation and storage of these objects. ScaLAPACK (Choi et al., 1996), the software library that we use to adapt our codes to a parallel machine environment, focuses on the third approach. In the paragraphs that follow, we describe how we have successfully managed to perform the computation required for our optimization problem over a parallel environment. Our serial code has already been tested and validated for several geophysical applications (Tondi and de Franco, 2006; Tondi et al., 2009). Therefore, in this report, we want to (a) guarantee the reliability and effectiveness of our parallel code through two synthetic experiments and (b) show that the handling of large dense systems is attainable with a field experiment. As the field experiment, the 3D modelling of densities and seismic velocities beneath the European continent are presented. The ability of the method to provide an improved geophysical model is evaluated on the basis of the data residuals.

2. The inversion algorithm

To establish the basis for the discussion of the parallel implementation of our serial code, in this paragraph we report the key elements of the inversion procedure. Readers are invited to refer to the previous papers for further details (Tondi et al., 2000; Tondi and de Franco, 2006). The approach does not perform a joint inversion; instead it pursues a multi-step strategy. The two datasets are inverted independently, provided they share the same structural boundaries and exist in a physical relationship between model densities and velocities. In the first step, only the seismic data – which can resolve detailed 3D structures – are inverted for a seismic wave-speed model. In general, any seismic inversion strategy based on various phase types can be performed in this step. In the second step, together with a space-dependent velocity–density relationship, the velocity model serves as a priori information for the integration of the gravity information. The joint posterior probability density function (JPP) of the problem is defined and solved for the maximum likelihood (Tarantola, 2005). These steps are iteratively repeated until the algorithm converges towards the final model. The flow chart in Fig. 1 schematically

summarises the processing steps and the data-flow of the procedure.

2.1. Initial settings

The 3D distribution of the seismic velocity structure and density represent the space-dependent functions $V(\mathbf{r})$ and $\rho(\mathbf{r})$, respectively, where \mathbf{r} is the position vector. For the scope of the inversion, these functions need to be approximated by a finite set of values: the parameters of the model. We choose to describe the model by representative values V_m and ρ_m at M nodes of a Cartesian grid, where M is the total number of model parameters and the index variable $m \in [1, M]$. For the description of the inversion algorithm, these model parameters are conveniently rewritten as vectors $\mathbf{v}^{(i)} = \{V_1, V_2, \dots, V_M\}$ and $\boldsymbol{\rho}^{(i)} = \{\rho_1, \rho_2, \dots, \rho_M\}$, where i is the iteration index of the inversion procedure, and where we define the model adjustment vectors $\Delta\boldsymbol{\rho}^{(i)} = \boldsymbol{\rho}^{(i+1)} - \boldsymbol{\rho}^{(i)}$ and $\Delta\mathbf{v}^{(i)} = \mathbf{v}^{(i+1)} - \mathbf{v}^{(i)}$ as well.

The input data for the integrated inversion comprise two distinct datasets: the seismological dataset (further detailed in the next paragraph of this section), and the gravity dataset of dimension N , where $\mathbf{g} = \{g_1, g_2, \dots, g_N\}$, $n \in [1, N]$, \mathbf{g}^{obs} denotes the input data vector, $\mathbf{g}^{(i)}$ the i -th's model predicted gravity response, and $\Delta\mathbf{g}^{(i)} = \mathbf{g}^{obs} - \mathbf{g}^{(i)}$ the corresponding gravity residual vector.

The procedure is initialized with an appropriate seismic velocity starting model (which will be referred to as the *a priori model* hereinafter). This *a priori* model is transformed into a mass density model using a linear space-dependent relationship:

$$\rho(\mathbf{r}) = \alpha(\mathbf{r})V(\mathbf{r}) + \beta(\mathbf{r}) \quad (1)$$

To account for this in the inversion procedure, we define the vector $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_M\}$ for the velocity–density gradient in the same parametrization as the model.

2.2. Surface-wave analysis

The seismological dataset consists of group velocity measurements $\{U_l^{obs}(T, w)\}$ of fundamental mode Love and Rayleigh waves at different periods T for a large set of source–receiver paths, where we use the index variable $l \in [1, L(T, w)]$ for the l -th path, and $L(T, w)$ for the total number of paths for the given period T and wave type w . The surface-wave inversion for the S-wave velocity structure follows the scheme by Schivardi and Morelli (2009, 2011). It is performed in two steps:

1. A linear inversion of group velocity measurement for Rayleigh and Love waves at different periods T yields a set of regional group velocity maps; i.e. a laterally varying function of geographic coordinates. For each wave type and geographic location, this set of maps defines the Rayleigh and Love wave group velocity dispersion curves; i.e. the local functions of period T .
2. A non-linear inversion takes these dispersion curves as input and yields the best-fitting depth profiles of the SV and SH velocities (V_{SV} and V_{SH} , respectively), and hence the isotropic $V_S = (\frac{2}{3}V_{SV} + \frac{1}{3}V_{SH})$ at each location.

This procedure first requires the calculation of the Rayleigh and Love wave group velocities $U_l^{(i)}(T, w)$ – for the set of periods T corresponding to the measurement dataset – as predicted by the current velocity model $\mathbf{v}^{(i)}$; i.e. the creation of the *synthetic* phase velocity maps. These maps are then used together with the group velocity data as the reference for the linear tomographic inversion. The updated group velocity maps serve as a priori

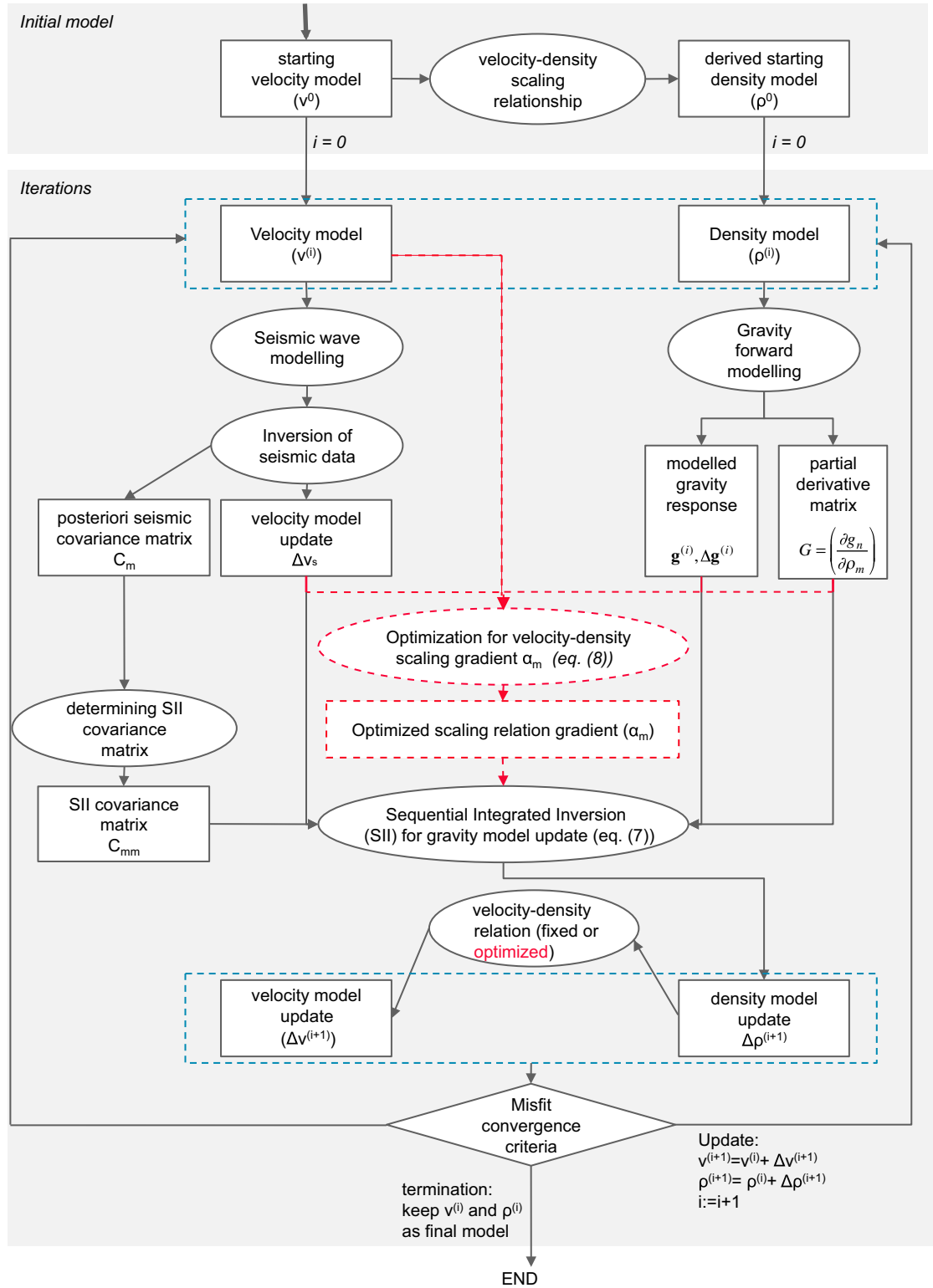


Fig. 1. Flow chart of the inversion procedure, including optimization of density and velocity model parameters through the inversion of the seismic wave field and the vertical component of the gravity field. Red dashed lines indicate the steps related to the optional optimization of the α parameter. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

information for the second inversion step, in which the combined group velocity maps define the local dispersion curve to be fitted by the shear-wave velocity profile of the new model at the given position. The resulting isotropic 3D shear-wave velocity model $\tilde{\mathbf{v}}^{(i)}$ is compared to the reference model $\mathbf{v}^{(i)}$, to obtain the intermediate

model update $\Delta \mathbf{v}_s^{(i)} = \tilde{\mathbf{v}}^{(i)} - \mathbf{v}^{(i)}$, which subsequently serves in the integrated inversion.

The inversion procedure also yields the a posteriori seismic covariance matrix \mathbf{C}_m and a misfit criterium for the seismic data $\Delta \mathbf{U}$. The a posteriori seismic covariance matrix is calculated from

the expression:

$$\mathbf{C}_m = \mathbf{C}_{ms} - \mathbf{C}_{ms} \mathbf{A}^T [\mathbf{A} \mathbf{C}_{ms}^{-1} \mathbf{A}^T + \mathbf{C}_d^{-1}]^{-1} \mathbf{A} \mathbf{C}_{ms} \quad (2)$$

where \mathbf{C}_{ms} is the a priori seismic covariance matrix, \mathbf{C}_d is a diagonal matrix of a posteriori errors of slowness (calculated) for each period, and matrix \mathbf{A} contains the partial derivatives of the group wave velocity with respect to the S-wave velocities (see Schiavardi and Morelli, 2009 for details). The a posteriori seismic covariance matrix \mathbf{C}_m is used as a priori information in \mathbf{C}_{mm} of the integrated inversion. The seismic data misfit is expressed as the root-mean-square of the group velocity residuals:

$$\overline{\Delta U}(T, w) = \sqrt{\frac{1}{L(T, w)} \sum_{l=1}^{L(T, w)} [U_l^{obs}(T, w) - U_l^i(T, w)]^2} \quad (3)$$

2.3. Gravity modelling

To proceed with the integrated inversion, we use the current mass density model $\rho^{(i)}$ (Fig. 2) to calculate the predicted gravity response $\mathbf{g}^{(i)}$. The input data \mathbf{g}^{obs} provide the vertical component of the gravity acceleration measurements, and thus the analysis considers only this component, and we use a scalar symbol g in the notation. Moreover, the partial derivatives of each gravity measurement g_n with respect to each density model parameter ρ_m are required for the procedure of integrated inversion and define the N by M partial derivative matrix

$$\mathbf{G} = \left[\left(\frac{\partial g_n}{\partial \rho_m} \right)_{n,m} \right]^{N \times M}$$

The algorithm by Pohánka (1998), for a polyhedral body where the density is linearly dependent on some coordinate, is applied to perform the forward gravity modelling and to estimate the partial derivatives in the current gravity model. After calculating

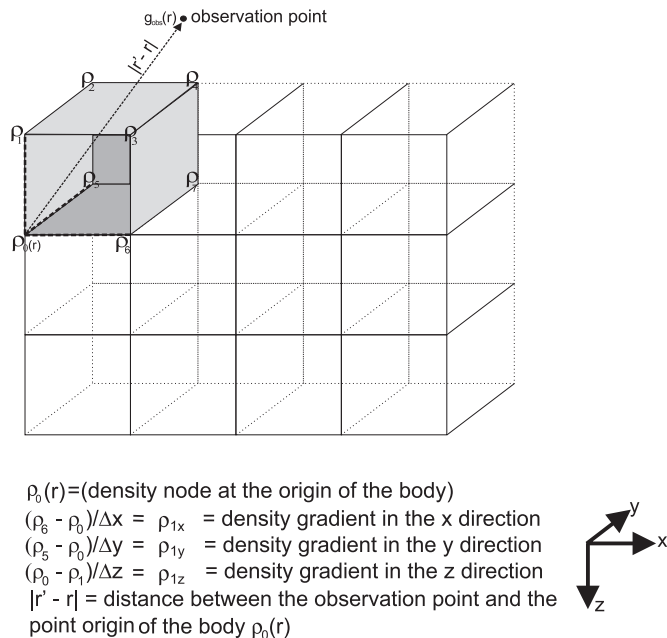


Fig. 2. Parametrization of the density model. For each density block, we assign a value for the density at the origin of coordinates $\rho_0(r)$ and the values of the density gradients in the x, y, and z directions. Density can be expressed as $\rho(\mathbf{r}) = \rho_0 + \rho_1 \cdot \mathbf{r}$, where ρ_1 is the arbitrarily oriented density gradient (Pohánka, 1998, Eq. (3)). As input data, \mathbf{g}^{obs} provides only the vertical component of the gravity acceleration vector; the density optimization is related to $\rho(z) = \rho_0 + \rho_1 \cdot z$.

the gravity data residual $\Delta \mathbf{g}^{(i)}$, the gravity data misfit is expressed as the root-mean-square over the residual vector:

$$\overline{\Delta g} = \sqrt{\frac{1}{N} \sum_{n=1}^N [g_n^{obs} - g_n^i]^2} \quad (4)$$

2.4. Sequential integrated inversion of seismological and gravity data

The JPP, which expresses the likelihood of the possible solutions of our problem (Tarantola, 2005), reads as

$$L(\mathbf{d}\mathbf{g}, \mathbf{d}\rho, \mathbf{d}\mathbf{v}) = \exp\left\{-\frac{1}{2}[\mathbf{d}\mathbf{g} - \mathbf{G} \mathbf{d}\rho]^T \mathbf{C}_{gg}^{-1} [\mathbf{d}\mathbf{g} - \mathbf{G} \mathbf{d}\rho]\right\} \times \exp\left\{-\frac{1}{2}[\mathbf{d}\rho - \alpha \mathbf{d}\mathbf{v}_s]^T \mathbf{C}_{mm}^{-1} [\mathbf{d}\rho - \alpha \mathbf{d}\mathbf{v}_s]\right\} \quad (5)$$

This is the product of two contributions: the first term provides the probability associated with the gravity model space, while the second term uses the node-dependent linear relationship between the wave velocity and density to integrate the best-fit seismic velocity model, which is obtained in the analysis above and used here as a priori information. The a priori covariance matrices \mathbf{C}_{gg} and \mathbf{C}_{mm} are used to account for possible errors associated with the gravity data and the seismic velocity model. The \mathbf{C}_{mm} matrix takes into account the error propagation from the velocity to the density model:

$$\sigma(\rho_m) = \sqrt{\mathbf{C}_{mm}} = \mathbf{v}_m \sigma(\alpha_m) + \alpha_m \sigma(\mathbf{v}_m) + \sigma(\beta_m) \quad (6)$$

where the symbol $\sigma(\cdot)$ describes the uncertainty associated with the given quantity, $\sigma(\mathbf{v}_m)$ is taken from the square roots of the diagonal elements of the a posteriori seismic covariance operator (Eq. (2)), and $\sigma(\alpha_m)$ and $\sigma(\beta_m)$ are the uncertainties in the coefficients of the scaling $\rho_m - \mathbf{v}_m$ linear relationship. The values on the main diagonal of \mathbf{C}_{mm} (under the assumption that there is no correlation between the errors in each node) determine the trade-off between the data variance and the solution variance. Very small values of \mathbf{C}_{mm} correspond to high values of \mathbf{C}_{mm}^{-1} , hence representing over-regularized solutions where the density model keeps close to the starting model (the seismological model).

The maximum likelihood point of the JPP (Eq. (5)) with respect to the density parameter $\max_{(\mathbf{d}\rho)} L(\mathbf{d}\mathbf{g}, \mathbf{d}\rho, \mathbf{d}\mathbf{v})$ leads to the solution of the problem:

$$\Delta \rho^{(i)} = (\mathbf{G}^T \mathbf{C}_{gg}^{-1} \mathbf{G} + \mathbf{C}_{mm}^{-1})^{-1} (\mathbf{G}^T \mathbf{C}_{gg}^{-1} \Delta \mathbf{g} + \alpha \mathbf{C}_{mm}^{-1} \Delta \mathbf{v}_s) \quad (7)$$

Additionally, as shown in the flow chart (Fig. 1), JPP (Eq. (5)) can also be maximized with respect to the α parameter:

$$\alpha_m^{(i)} = (\mathbf{G}^T \mathbf{C}_{gg}^{-1} \mathbf{G} (\mathbf{v}_0 + \Delta \mathbf{v}_s))^{-1} (\mathbf{G}^T \mathbf{C}_{gg}^{-1} \mathbf{g}) \quad (8)$$

In this way, the starting $\rho_m - \mathbf{v}_m$ relationship is modified according to the datasets. The update to the density model is transformed into the update to the velocity model through the assumed (or optimized) linear relationships, and the two models are checked for gravity data misfit (Eq. (4)) and seismic data misfit (Eq. (3)). The procedure is repeated, following the flow chart in Fig. 1, until a satisfactory fit to the observed datasets (data misfit) is achieved (usually not more than 2–3 iterations).

3. Parallel solution strategy

As anticipated in the Introduction, to match seismological to gravity analysis, our algorithm of inversion requires the 3D velocity and density models to share the same parametrization. However, while the matrix Green's function used for the seismological analysis is a sparse matrix (a matrix populated primarily

with zeros), which is by nature easily compressed and can benefit from specialized algorithms for the storage and calculation, the \mathbf{G} matrix is dense.

The serial algorithm for the computation of the gravity field \mathbf{g} (Fig. 6) first cycles over the N measurement points, and for each point \mathbf{g}_n ($n=1, \dots, N$) it sums the mass contribution given by each block of the model; contextually, the elements $\mathbf{G}_{n,m}$ of the \mathbf{G} matrix are calculated inside the loop of each block. As the dimensions of the problem increase, these two inner loops can become very time consuming (e.g. for a \mathbf{G} sized 7455×7500 , the code takes 1 h, 7 min to execute on a Supermicro server with a 2.2 GHz Dual Core AMD Opteron Processor) and render large solutions impracticable.

However, we can exploit these two loops and the ScaLAPACK strategy of the block-cyclic decomposition scheme to divide the computational load among a 2D grid of processors. ScaLAPACK is a collection of routines written in Fortran77 for dense real or complex linear system solving, and least squares and eigenproblems. It extends the underlying concept of LAPACK (Linear

Algebra PACKage) (Anderson et al., 1992) to distributed memory concurrent supercomputers, through the use of block-partitioned algorithms, to minimize data movement between different levels in hierarchical memory. Block-partitioned algorithms are ground on the block-cyclic distribution of a matrix, and hence the key argument to start a computational process with ScaLAPACK is to distribute the vectors and matrices into a 2D grid of processors (Fig. 3).

In this way, we can attain two goals at the same time: we divide the execution time by a factor approximately equal to the total number of used cores, and we divide the matrix in a number of submatrices to be used directly by the parallel version of the optimization code built with the use of ScaLAPACK routines. The pseudo-code shown in Fig. 4 explains in detail each step of the implemented parallel code.

The code is implemented in Fortran90 and the MPI library, together with the BLACS communication environment, allow the distribution of the computational load among the decided grid of processors, and the production and writing of

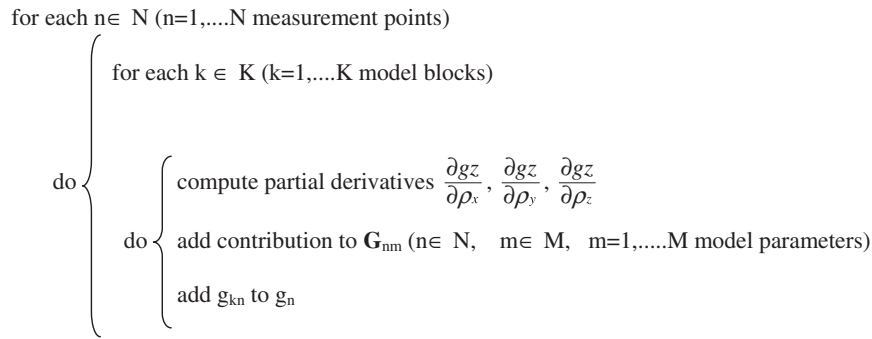


Fig. 3. Pseudo-code algorithm for serial computation of the gravity information.

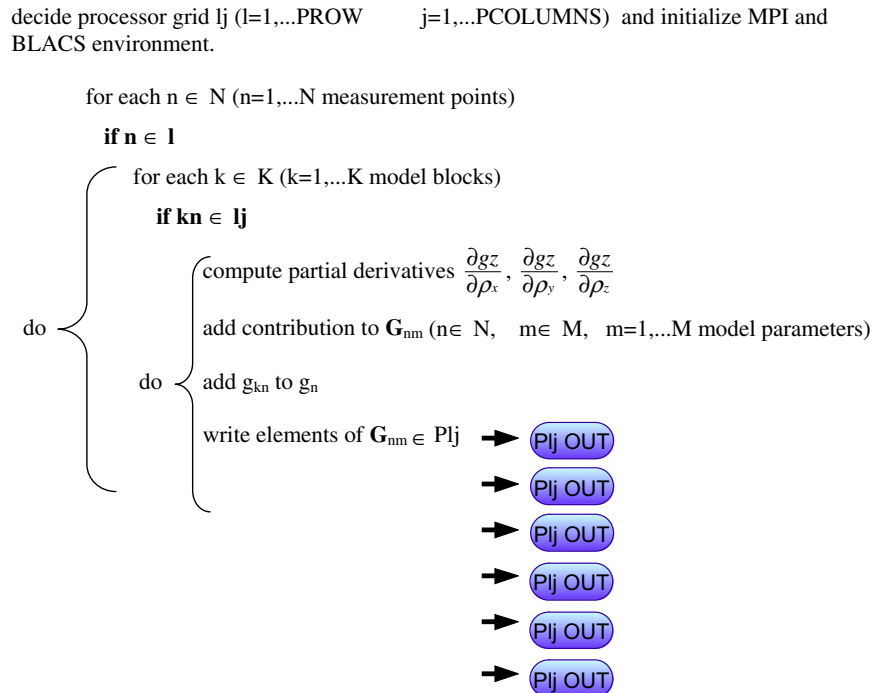


Fig. 4. Pseudo-code algorithm for parallel computation of the gravity information. Pl_j ($l=1, P_{\text{ROW}} \times j=1, P_{\text{COLUMN}}$) is the size of the processor grid.

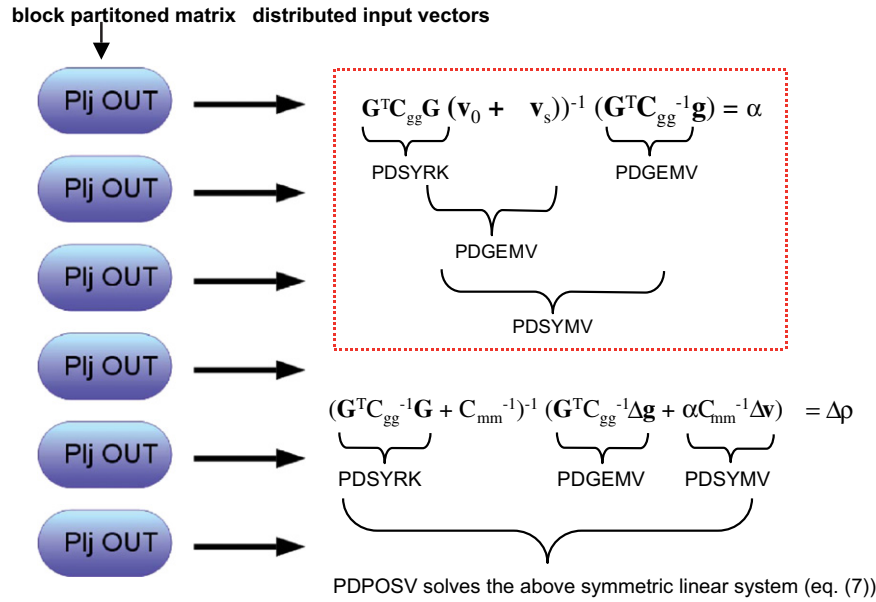


Fig. 5. Scheme showing the I/O operations of the ScaLAPACK PDPOSV routine used to implement the parallel version of the optimization algorithm. PDSYRK computes the Parallel Rank-K update of a Real Symmetric Matrix in Double precision. PDGEMV computes the Parallel Matrix–Vector product for either a real General matrix or its transpose in Double precision. PDSYMV computes the Parallel Matrix–Vector product for a real Symmetric matrix in Double precision (Choi et al., 1994). The red dashed rectangle indicates the optional computations for the optimization of the α parameter (Eq. (8)). The matrices and vectors are explained in detail in the text. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

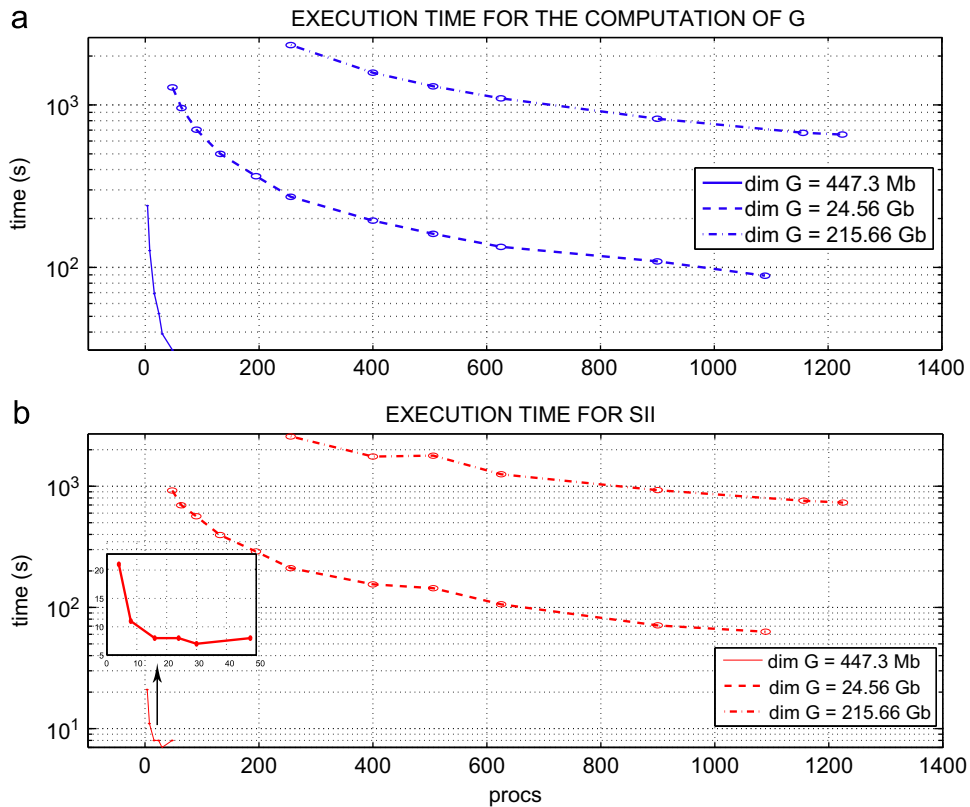


Fig. 6. (a) Execution times related to the production of the dense matrix G , and for (b) to the optimization process (SII). The results are on a log scale and are expressed as a function of the number of processors used for the computation, as related to three matrices of different order (sizes indicated in the legend). For the experiment with the small matrix (continuous line), the serial code takes 1 h, 7 min to produce the matrix, and 42 s to solve the inverse problem with the use of LAPACK library (Anderson et al., 1992).

the desired number of submatrices that contain the mass contribution of the specified block to the specific gravity measurement point.

Once the Plj ($l=1, P_{ROW} \times j=1, P_{COLUMN}$) files with the assigned portions of the block-partitioned matrix have been produced and the gravity residual vector Δg has been calculated, it is

straightforward to feed this information into the parallel version of the optimization algorithm (Eq. (7)). As illustrated in Fig. 5, each submatrix is read by the sequential integrated inversion (SII) algorithm, which uses the Parallel Basic Linear Algebra Subprogram (PBLAS) routines (Choi et al., 1994) for the distributed vector–vector (PDSYMV), Matrix–vector (PDGEMV) and Matrix–matrix (PDSYRK) operations, and the ScaLAPACK routine PDPOSV to compute the solution to the system of linear equations. All of the computation is performed in double precision, and when possible, we exploit the symmetry of the matrices.

With this strategy, problems of storage and of I/O operations and computations are easily approached, even with massive datasets.

4. Results analysis

The parallel code has been evaluated through two experiments. In the first experiment, we evaluated the *speed-up factor*, a measure of the relative execution time between an n -processor

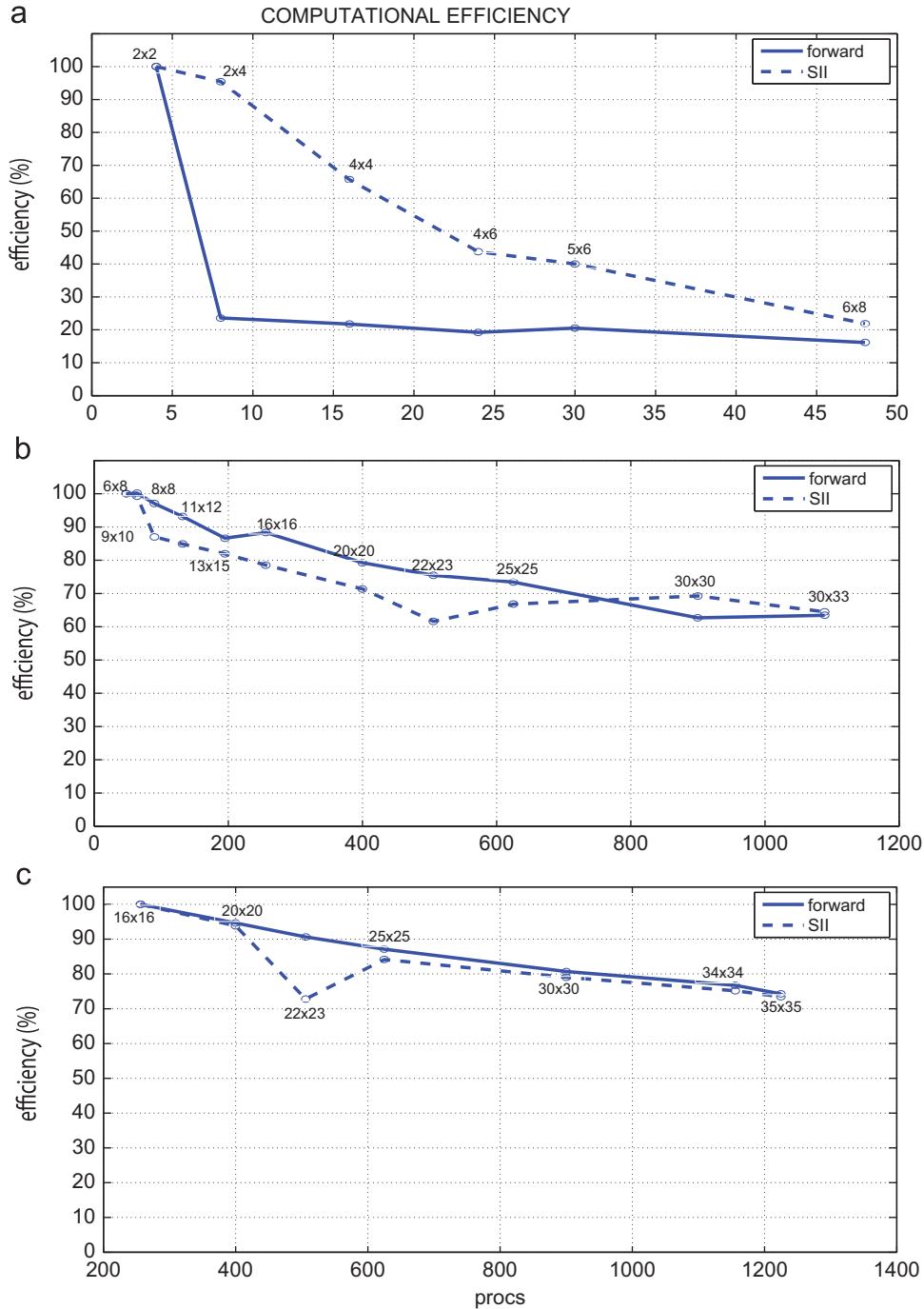


Fig. 7. Percentage computational efficiency (total execution time compared to time needed with the minimum size of process grid) related to the production of the dense matrix \mathbf{G} (continuous line) and to the optimization process (SII) (dashed line) for (a) 7455×7500 (447.3 Mb), (b) $57\,350 \times 57\,481$ (24.56 Gb), and (c) $170\,255 \times 170\,016$ (215.66 Gb) matrices. The results are expressed as a function of the number of processors used for the computation. The processor grid ($P_{\text{ROWS}} \times P_{\text{COLUMNS}}$) is also indicated beside each curve.

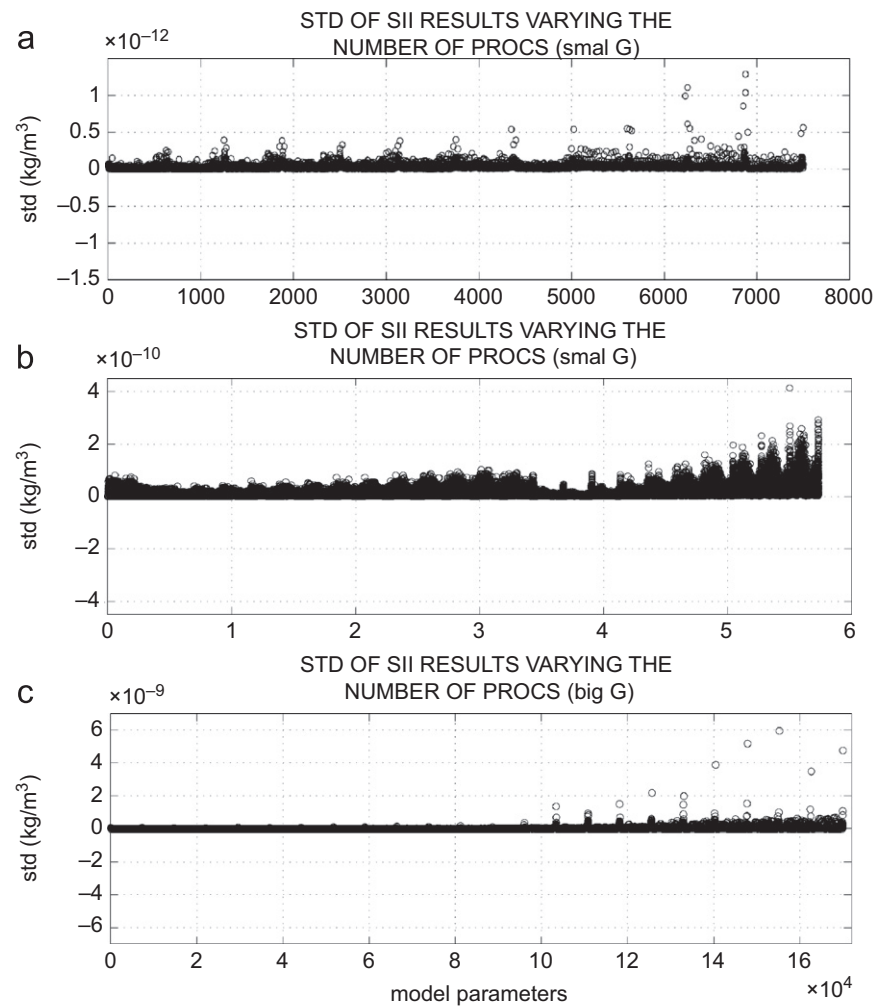


Fig. 8. Evaluation of the standard deviation of each parameter, calculated with the SII solver and varying the number of processors. The number of executions is the same as that used to test the performance: (a) six runs for the small matrix (447.3 Mb), (b) eleven runs for the medium matrix (24.56 Gb), and (c) seven runs for the big matrix (215.66 Gb).

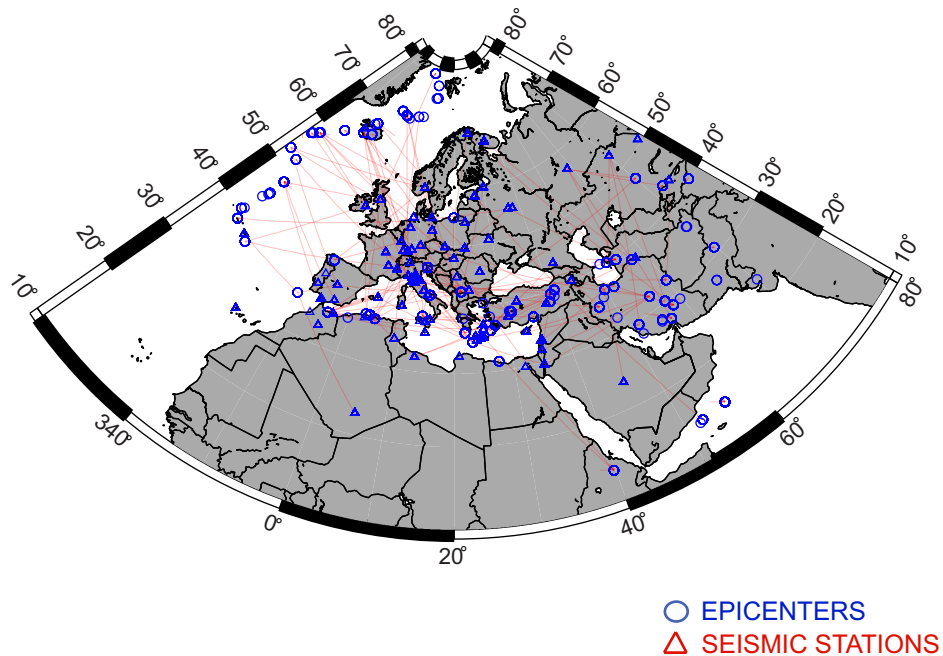


Fig. 9. Rayleigh ray path distribution ($T=60$ s) related to the seismic dataset used for this study, with superimposed earthquakes (blue circles) and seismic stations (red triangles). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

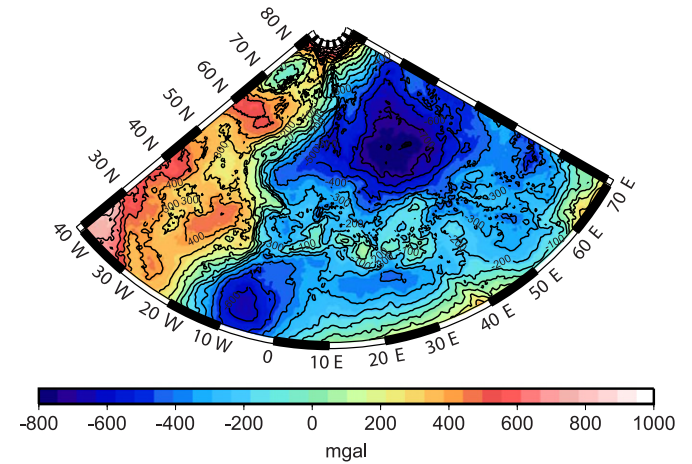
system and a single processor system, varying the size of the input matrices. The second experiment is related to the correctness of the results. The results are discussed as a function of the size of the processor grid, and were obtained over the parallel environment IBM Power6-575 installed at the Consorzio Interuniversitario Calcolo Elettronico (CINECA). The cluster is composed of 5376 4.7 GHz IBM Power6 4 Gb RAM processors on 168 nodes, connected by Infiniband $\times 4$ DDR. AIX 6 is the operating system.

4.1. Performance

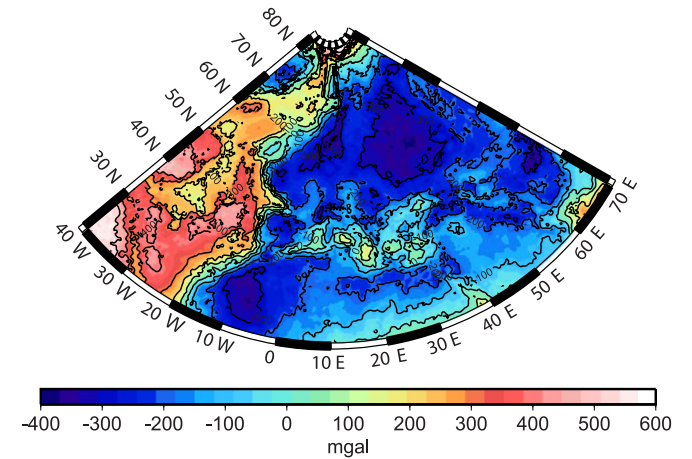
Three \mathbf{G} matrices of different sizes were used as the test cases: (a) 7455×7500 (memory to be allocated 447.3 Mb), (b) 57350×57481 (memory to be allocated 24.56 Gb), and (c) 170255×170016 (memory to be allocated 215.66 Gb). For each of these matrices, the executions were performed with the number of processors varying from a minimum to a maximum, appropriately chosen with respect to the size of the matrix [4; 48], [48; 1089], [256; 1225]. The performances were evaluated on the basis of the execution times needed to solve the problem. Due to memory issues, it was possible to evaluate the efficiency of the algorithms with respect to the serial solvers only in the case of \mathbf{G} sized 7455×7500 . Fig. 6a compares the relative execution times for the computation of the three \mathbf{G} of different sizes (forward problem). The lower boundary was set considering the size of the matrix with respect to the RAM of the processors; the upper boundary was fixed for when the relative execution time started to increase. Fig. 6b shows the execution time for the optimization problem. As we can observe, the trend of the time curve shows the same behaviour, for both the forward and the optimization problems. The proposed parallel solutions show good scalability up to a corner value; beyond this, the execution time reaches an asymptote and it is not convenient to increase the number of processors. If we compare the efficiency (calculated with respect to the minimum size of the process grid) that was achieved by the three implementations (Fig. 7), it is possible to note an increment in the performance with the growth of the

computational grid, with the efficiency passing from an average of 20% in the small-sized problem to an average of 75% in the large-sized problem. It is interesting that there is an anomalous loss of

a EUROPE Bouguer residual field for the starting model (Birch)



b EUROPE Bouguer residual field for the starting model (Karato)



c EUROPE Bouguer residual field for the starting model (Karato & Karki)

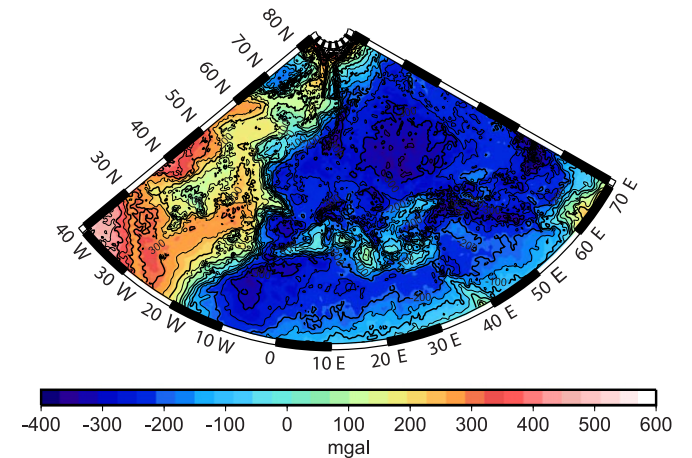


Fig. 11. Gravity data misfit ($g_{\text{OBS}} - g_{\text{CALC}}$) related to the starting density model calculated with the parallel procedure visualized in Fig. 4. The 170 255 measure points are regularly distributed on the study area with different starting density-velocity scaling relationships: (a) the empirical Birch's (1964) law (mean RMS=484.3982 mgal), which was chosen for our final model; (b) the Karato (1993) law (mean RMS=254.4835); and (c) a modification of the Karato (1993) relationship on the basis of geodynamic and mineral physics constraints (Karato and Karki, 2001) (mean RMS=281.9063 mgal).

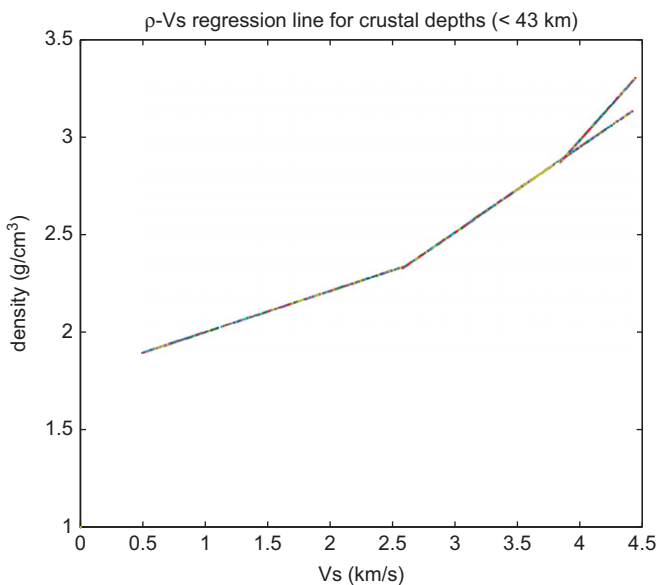


Fig. 10. Piece-wise regression line related to the linear $v_s - \rho$ relationships used to scale the v_s crustal model to a density model. The relationships are derived from $\rho - v_p$, $v_p - v_s$ Brocher (2005) regression fit used to derive the EPCrust crustal model (Molinari and Morelli, 2011).

efficiency with 506 processors (distributed in $22 P_{ROWS} \times 23 P_{COLUMNS}$) that is especially evident for the SII solution. The problem is not correlated with the ratio of P_{ROWS} , $P_{COLUMNS}$, to the N , M parameters, and hence it is probably due to communication issues of the parallel environment. For the small order matrix (Fig. 7a), the efficiency related to the computation of partial derivatives is better than that to solve the optimization problem; for the medium (Fig. 7b) and big order (Fig. 7c) matrices, the results are exchanged.

4.2. Correctness

With the aim to verify the accuracy of our parallel solutions, we first take care to exclude problems associated with ill-conditioning. Secondly, the results obtained with a different number of processors have been compared, and the standard deviation (std) with respect to the mean has been estimated for each parameter.

The condition number determines the loss in precision due to random errors in Gaussian elimination, and it can be used to estimate the accuracy of results obtained from matrix inversion and linear equation solutions. Generally speaking, if the condition number of the matrix is $\geq 1/\sqrt{\epsilon_{ps}}$, where ϵ_{ps} is the spacing of the floating point numbers, then caution is advised for subsequent computations. With the ScaLAPACK routine PDPOCON, it was possible to compute the reciprocal of the condition number, which for these three matrices resulted respectively in 0.1045 (i.e. condition number=9.5694), $0.286089E-6$ (i.e. condition number=3.4954E6), and $0.312052E-3$ (i.e. condition number=3.2046E3). The worst conditioning is given by the medium sized matrix, but in any case, these are below the reciprocal of the $\sqrt{\epsilon_{ps}}$ of the machine, which is in the order of $10E7$.

For the second test, Fig. 8 shows that the order of the std increases with the order of the matrix; however, the effect on the solution is negligible, considering that the estimated a priori density error has been set to 10 kg/m^3 .

5. The 3D seismic and density structure beneath the European continent

The European continent is a geophysical laboratory that has long been the subject of many scientific studies (Piromallo and Morelli, 2003; Pilidou et al., 2004; Boschi et al., 2004; Schivardi and Morelli, 2009). Large scale structural heterogeneities have been found here, together with local tectonic structures, some of which remains to be understood. They are also often localized where population and infrastructure are concentrated. In this scenario, to discriminate among rival models, it is almost essential to jointly analyse the information provided by different geophysical datasets and to reconstruct models that are consistent with each measurement.

Following the flow chart of our procedure in Fig. 1, our starting model is a velocity model. It is the superposition of the 3D crustal model EPCrust that was recently assembled for the whole European region by Molinari and Morelli (2011), and the portion of the 3D isotropic global shear-wave model S20RTS (Ritsema et al., 1999) that represents the upper mantle below the study area. It covers the geographical frame between 20° and 84°N and -40° and 70°E , and it extends down to 500 km in depth. The seismological observations to be fitted within the model framework consist of 1550 Rayleigh and 900 Love fundamental mode group velocity dispersion curves measured on regional paths (600–6000 km) and related to intermediate-to-large-magnitude shallow earthquake events that occurred between 1998 and 2005. The resulting shear-wave velocity model is compared

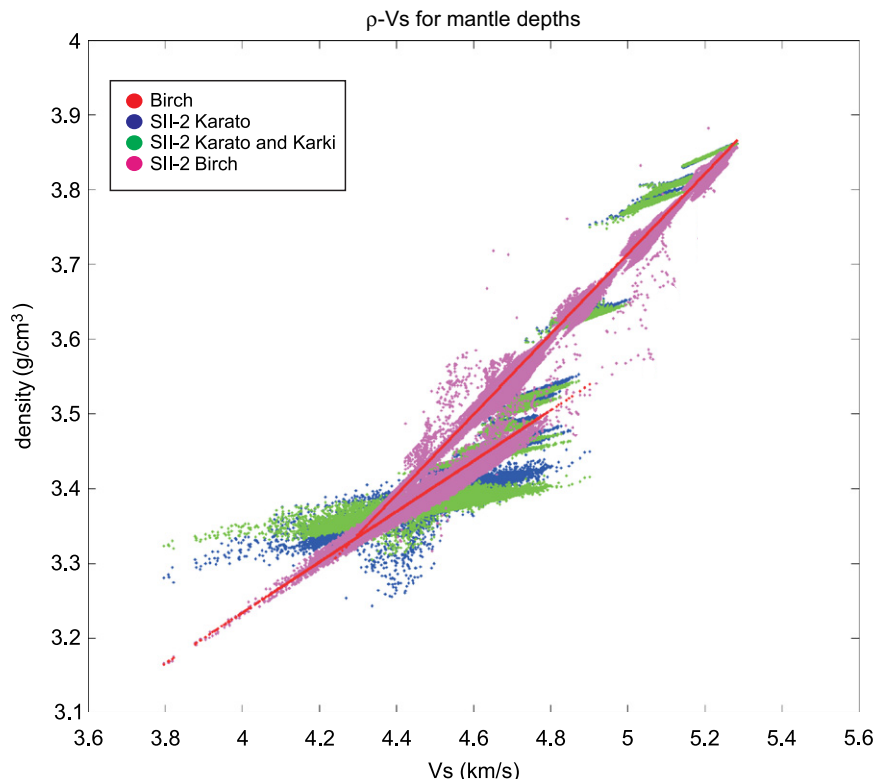


Fig. 12. Comparison of our SII ρ – v_s final models with the a priori interpolated Birch regression lines (solid red lines).

to the starting velocity model (v_0), and Δv^0 , which is needed by Eq. (7), is calculated. Contextually, the seismic data misfit and a posteriori seismic covariance matrix (Eq. (2)) are calculated. Fig. 9 shows the location of the stations and events, and the Rayleigh ray path distribution for the period of 60 s.

The second step of our procedure implies the transformation of the starting velocity model into the starting density model, through a node-dependent linear relationship. We tested three radially symmetrical scaling profiles: the first is as derived from the empirical Birch's (1964) v_s – ρ table, the second is as derived by Karato (1993) for the whole mantle minerals, and the last is a modification of the Karato (1993) relationship on the basis of geodynamic and mineral physics constraints (Karato and Karki, 2001). The widespread success of Birch's law

can be ascribed to its virtual coincidence with, and hence linearization, a power law derived from lattice dynamics over the density range from $\approx 2.5 \text{ g/cm}^3$ to 4.0 g/cm^3 (Chung, 1972), which is typical of rock and minerals in our upper mantle model. Birch's law concerns P-wave velocities; our shear-wave- ρ depth-dependent relationships are the regression lines drawn through the plot of the two variables extracted from his tables. The possibility to experiment with different scaling relationships to find the one that is most appropriate is favoured by the performance of the parallelized algorithm. Additionally, for the parameters that belong to the upper mantle, JPP (Eq. (5)) is maximized also with respect to the α parameter (Eq. (8)), which is inverted independently before each SII inversion.

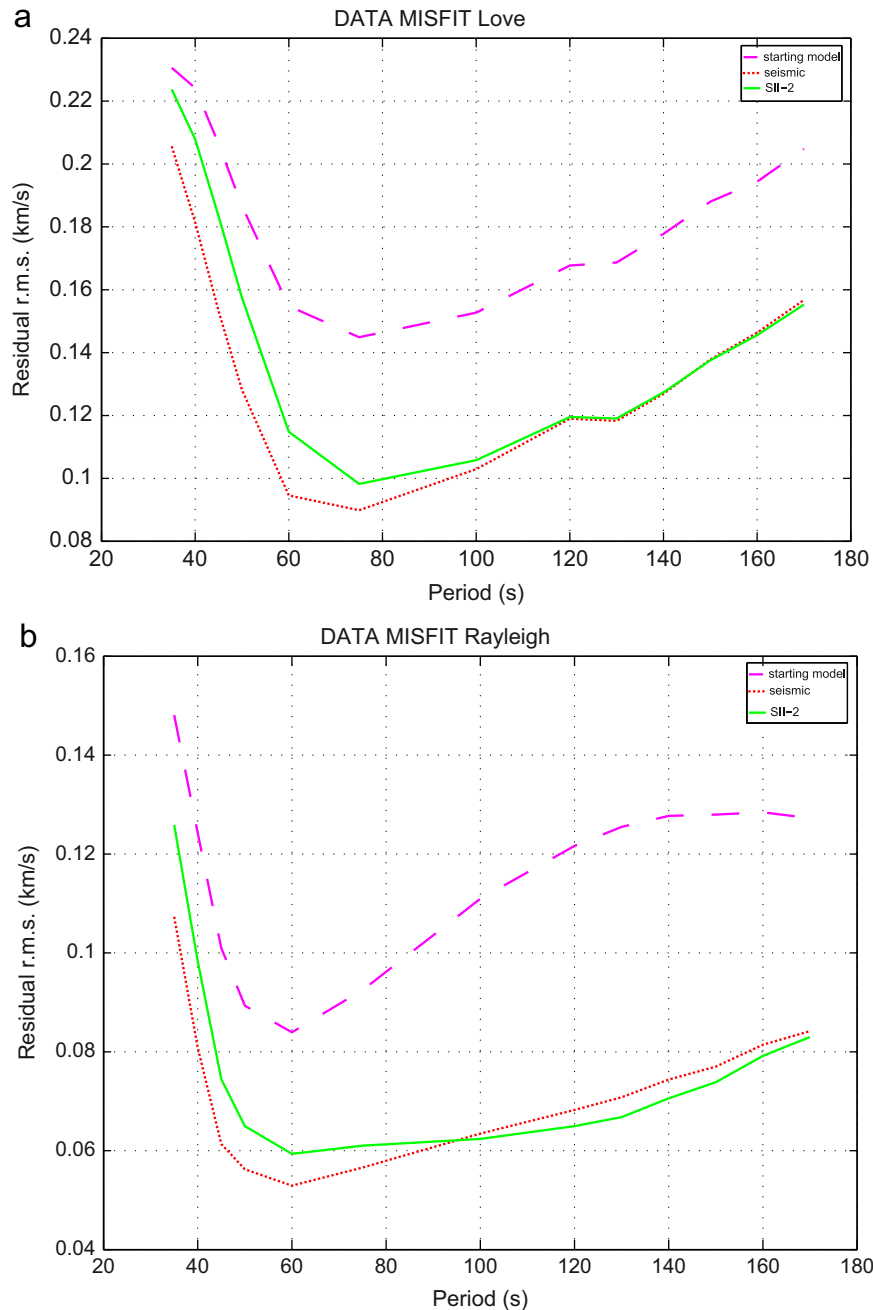


Fig. 13. Comparison of waveform residuals ($t_{\text{OBS}} - t_{\text{CALC}}$ expressed in velocity) between the starting, the seismological, and the SII-2 shear-wave velocity model. Values are calculated for each wave period and separated for the Love (a) and Rayleigh (b) datasets. The wave periods are approximately related to the sampled depth.

The transformed grid of $112 \times 66 \times 23$ density values is represented by a collection of 154 880 internal prisms that are sized $1^\circ \times 1^\circ$ and are of variable thickness, and 7744 external bodies that are sized $6^\circ \times 6^\circ$ and are also of variable thickness, to counteract edge effects. The first $112 \times 66 \times 8$ values represent the crustal model, which is scaled to the density according to relationships (shown in Fig. 10) that remain invariant throughout the whole process of the optimization. The synthetic gravity field and the model partial derivatives are calculated using the parallel procedure explained in Section 3, and they are visualized in Fig. 4. The results are compared to the Bouguer gravity anomalies derived from the satellite GGM02C (Tapley et al., 2004) high-resolution global gravity model. The residuals (Δg) at 170 255 points (Fig. 11a–c), are used as input for the parallel inverse procedure. The a priori gravity data error to be used for the \mathbf{G}_{gg} matrix is set to 1.5 mgal for all of the measurements.

For the values on the main diagonal of \mathbf{C}_{mm} , we consider the resolution analysis performed on seismological data (extensively analysed in Schivardi and Morelli, 2011, Section 3.3), and higher $\sigma(\alpha_m)$ and $\sigma(\beta_m)$ uncertainties are assigned below 250 km in depth. This allows the complementarity of the seismological and gravity datasets to be exploited, and for reliable solutions to be obtained also below the seismic depth resolution. \mathbf{C}_{mm} is a diagonal matrix, and no correlation among the errors for each node is assumed. The density updates ($\Delta \rho$) given by the SII are transformed in the update to the velocity model ($\Delta \mathbf{v}$) through the fixed (for the crust) or optimized $\rho_m - \mathbf{v}_m$ linear relationships. The optimization process is repeated until the fit to the observations deteriorates with respect to the former iteration. Separate optimizations are carried out with the three $v_s - \rho$ scaling profiles, and after SII inversion, the a priori velocity-to-density relationship that yields the most successful fit to the combined dataset is the Birch relationship (Fig. 12). Our preferred model is obtained after two SII iterations (SII-2). A consistent misfit reduction is clearly observable in both the seismological and gravity data. The Rayleigh data (Fig. 13a) are predicted with a misfit of 0.0757 s, and the Love measurements (Fig. 10b) with a misfit of 0.1458 s. As can be seen from Fig. 13, the major seismic improvement is verified for larger wave periods and hence for depths below 100 km, where the gravity information counteracts the loss of seismic resolution. The Bouguer variance reduction is 85%

Table 1

Data fit to seismological and gravity data of the starting model, of the model derived from the inversion of seismic data, and of the SII-2 model (derived after the second iteration of sequential integrated inversion of the seismological and gravity data). The Bouguer variance reduction is calculated as the percentage relationship between gravity data misfit and the amplitude of gravity observations.

Models	Seismic rms (km/s)	Bouguer rms (mgal)	Bouguer variance reduction (%)
Starting	0.1887 Love 0.1160 Rayleigh	484.3982	67.51
Seismic	0.1355 Love 0.0719 Rayleigh	479.0049	68.14
SII-2 (Birch)	0.1458 Love 0.0757 Rayleigh	53.6622	85.47

(Fig. 14). Table 1 shows the data residuals related to the seismological and gravity data, and the comparison between the seismological inversion and the integrated inversion. Taking into account the seismic ray coverage (which is completely missing at the edges of the model) and the observed satellite gravity field resolution (about 1° pixel size), the results are satisfactory. Fig. 15 depicts the SII-2 density adjustment, with respect to the starting density model. The main observation is the decrease in the densities below the continental plate, which are associated with an increase in the same amplitude below the oceanic plate. Iron depletion in the cold cratonic roots explains the general decreases in the densities below the continental shields, which have also been observed in other studies (Artemieva, 2003; Jordan, 1978; Kaban et al., 2003). Further geological and dynamic implications of our models are better explained in a dedicated paper (Tondi et al., in press).

The size of the experimental problem is the same as in the third case analysed for the correctness and performance of the algorithms. Hence, for the computational behaviour of the adopted forward and inverse parallel solvers, we refer to the case of matrix \mathbf{G} sized $170\,255 \times 170\,255$.

6. Concluding remarks

We have successfully tested here the parallel implementation of a method for 3D combined inversion of seismological and gravity data. The use of the MPI library together with some LAPACK and ScaLAPACK routines allows us to use the algorithm in high resolution 3D geophysical analysis when a massive number of model parameters and data information are needed. In particular, we have handled the presence of a large dense matrix, which is common with information coming from a potential field, and with detailed parametrization, relatively easily. Our strategy exploits the block-cyclic decomposition scheme to divide both the computational load of the matrix and the storage among a 2D grid of processors. Three matrices of different sizes are used in several experiments that are aimed at the evaluation of both the performance and the correctness achieved by the forward and inverse solvers. Good scalability is observed for the three matrices up to an asymptotic limit of the execution time, which is in direct proportion with the sizes of the matrices. The loss of accuracy of the results increases with the order of the matrices and with the distance of the parameters from the measurement points; however, the error is much smaller than the expected variance of the parameter estimate.

The results of the code as applied to real observations show that the inclusion of information coming from potential fields in

EUROPE Bouguer residual field for SII-2 density model (Birch)

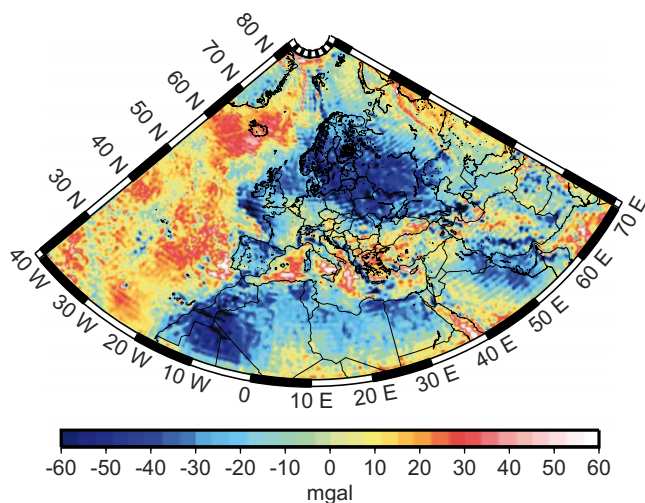


Fig. 14. Gravity data misfit (OBS–CALC) produced by the SII-2 density model recovered with the empirical Birch's (1964) law density–velocity scaling relationships. The values are calculated at 170 255 points that are regularly distributed in the study area.

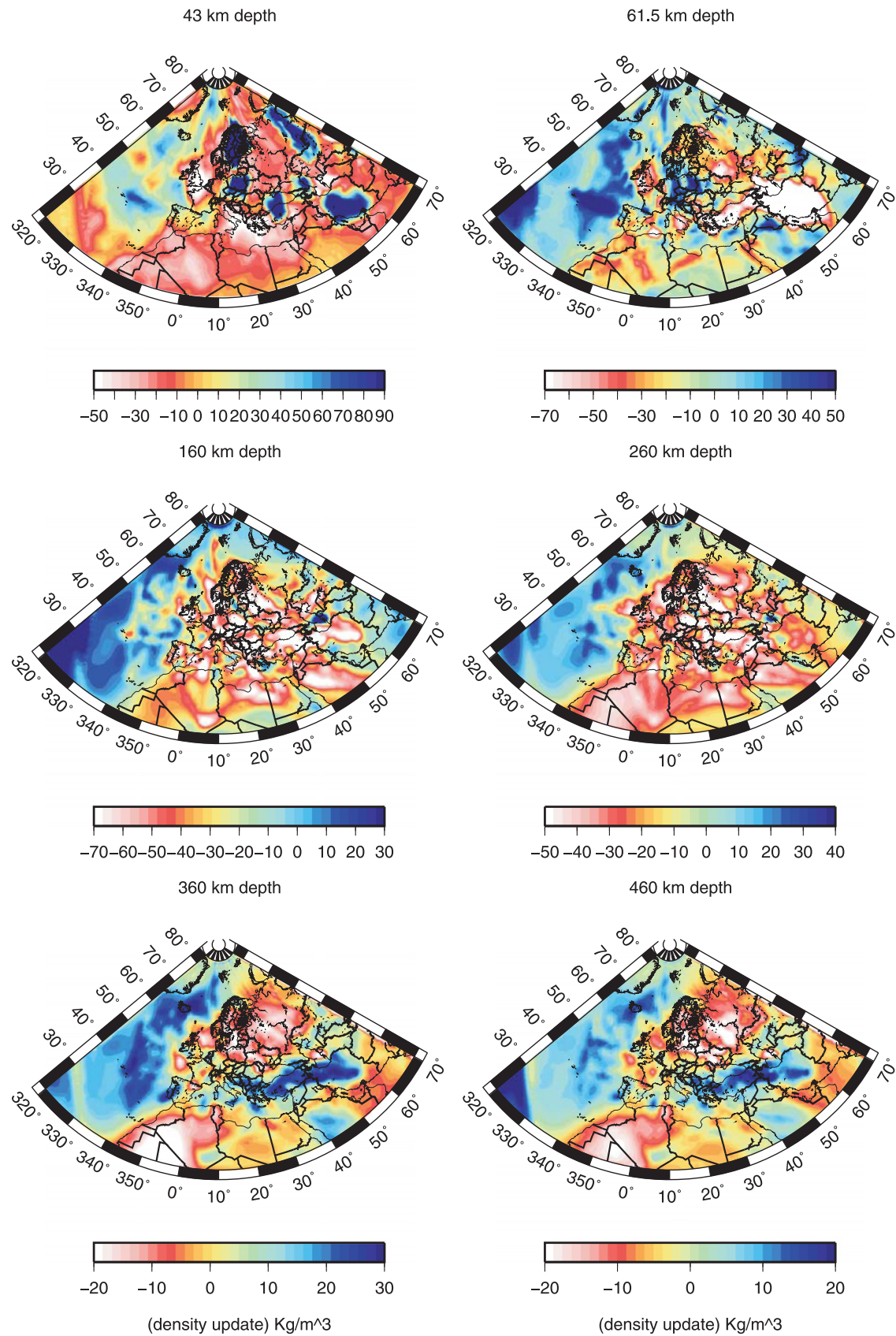


Fig. 15. Depth sections of the density update given by the SII-2 density model with respect to the starting model. The areal model resolution is $1^\circ \times 1^\circ$.

3D high-resolution regional or global studies of earth structures is attainable.

Acknowledgments

This research was supported by NERIES INFRAS-2.1-026130 and by MERG-CT-2007-046522. We thank two anonymous reviewers, who improved the manuscript by their constructive comments.

References

- Anderson, E., Bai, Z., Bischof, C., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., Ostrouchov, S., Sorensen, D., 1992. LAPACK Users's Guide. SIAM, Philadelphia, PA, 235 pp.
- Aki, K., Richards, P.G., 1980. Quantitative Seismology. W.H. Freeman & Co, San Francisco 700 pp.
- Artemieva, I., 2003. Lithospheric structure, composition, and thermal regime of the East European Craton: implications for the subsidence of the Russian platform. *Earth and Planetary Science Letters* 213, 431–446.
- Birch, F., 1964. Density and composition of mantle and core. *Journal of Geophysical Research* 69 (20), 4377–4384.
- Bosch, M., 1999. Lithologic tomography: from plural geophysical data to lithology estimation. *Journal of Geophysical Research* 104 (B1), 749–766.
- Boschi, L., Ekström, G., Kustowski, B., 2004. Multiple resolution surface wave tomography: the Mediterranean basin. *Geophysical Journal International* 157, 293–304.
- Brocher, T.M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America* 95 (6), 2081–2092.
- Camacho, A.G., Montesinos, F.G., Vieira, R., 1997. A three dimensional gravity inversion applied to Sao Miguel Island (Azores). *Journal of Geophysical Research* 102, 7717–7730.
- Choi, J., Demmel, J., Dhillon, I., Dongarra, J., Ostrouchov, S., Petit, A., Stanley, K., Walker, D., Whaley, R.C., 1996. ScaLAPACK: a portable linear algebra library for distributed memory computers—design issues and performance. *Computer Physics Communications* 97, 1–15.
- Choi, J., Dongarra, J., Walker, D., 1994. PB-BLAS: a set of parallel block basic linear algebra subroutines. In: *Proceedings of Scalable High Performance Computing Conference*. IEEE Computer Society Press, pp. 534–541.
- Chung, D.H., 1972. Birch's law: why is it so good? *Science* 177, 261–263.
- Edelman, A., 1993. Large dense numerical linear algebra in 1993, the parallel computing influence. *Journal of Supercomputing Applications* 7, 113–128.
- Gallardo, L.A., Maju, M.A., 2004. Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints. *Journal of Geophysical Research* 109, B03311, <http://dx.doi.org/10.1029/2003JB002716>.
- Jordan, T.H., 1978. Composition and development of the continental tectosphere. *Nature* 274, 544–548.
- Kaban, M.K., Schwintzer, P., Artemieva, I.M., Mooney, W.D., 2003. Density of the continental roots: compositional and thermal contributions. *Earth and Planetary Science Letters* 209, 53–69.
- Karato, S., 1993. Importance of anelasticity in the interpretation of seismic tomography. *Geophysical Research Letters* 20, 1623–1626, <http://dx.doi.org/10.1029/93GL01767>.
- Karato, S., Karki, B.B., 2001. Origin of lateral variation of seismic wave velocities and density in the deep mantle. *Journal of Geophysical Research* 106 (B10), 21,771–21,783.
- Li, Y., Oldenburgh, D.W., 1996. 3-D inversion of magnetic data. *Geophysics* 61, 394–408.
- Lines, L.R., Schultz, A.K., Treitel, S., 1988. Cooperative inversion of geophysical data. *Geophysics* 53, 8–20.
- Molinari, I., Morelli, A., 2011. EPCrust: a reference crustal model for the European plate. *Geophysical Journal International* 185, 352–364.
- Moorkamp, M., Jones, A., Fishwick, S., 2010. Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data. *Journal of Geophysical Research* 115, B04318.
- Newman, G.A., Alumbaugh, D.L., 1997. Three dimensional massively parallel electromagnetic inversion, II, analysis of a crosswell electromagnetic experiment. *Geophysical Journal International* 128, 355–363.
- Pilidou, S., Priestley, K., Gudmundsson, O., Debayle, E., 2004. Upper mantle S-wave speed heterogeneity and anisotropy beneath the North Atlantic from regional surface wave tomography: the Iceland and Azores plumes. *Geophysical Journal International* 159, 1057–1076.
- Piromallo, C., Morelli, A., 2003. P wave tomography of the mantle under the Alpine–Mediterranean area. *Journal of Geophysical Research* 108 (B2), 2065 <http://dx.doi.org/10.1029/2002JB001757>.
- Pohanka, V., 1998. Optimum expression for computation of the gravity field of a polyhedral body with linearly increasing density. *Geophysical Prospecting* 46, 391–404.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W., 1986. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK.
- Ritsema, J., van Heijst, H.J., Woodhouse, J.H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science* 286, 1925–1928.
- Schivardi, R., Morelli, A., 2009. Surface wave tomography in the European and Mediterranean region. *Geophysical Journal International* 177, 1050–1066.
- Schivardi, R., Morelli, A., 2011. EPmantle: a three-dimensional transversely isotropic model of the upper mantle under the European Plate. *Geophysical Journal International* 185, 469–484.
- Tapley, B.D., Bettadpur, S., Watkins, M., Reigber, C., 2004. The gravity recovery and climate experiment: mission overview and early results. *Geophysical Research Letters* 31 (L09607), 555, <http://dx.doi.org/10.1029/2004GL019920>.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia 342 pp.
- Tiberi, C., Diamant, M., Déverchère, J., Mikhailov, V., Tikhotsky, S., Achauer, U., 2003. Deep structure of the Baikal rift zone revealed by joint inversion of gravity and seismology. *Journal of Geophysical Research* 108 (B3), 2133 15pp.
- Tondi, R., Achauer, U., Landes, M., Ravi, R., Besutiu, L., 2009. Unveiling seismic and density structure beneath the Vrancea seismogenic zone, Romania. *Journal of Geophysical Research* 114 (B11307) <http://dx.doi.org/10.1029/2008JB005992>.
- Tondi, R., de Franco, R., 2006. Accurate assessment of 3D crustal velocity and density parameters: application to Vesuvius data sets. *Physics of the Earth and Planetary Interiors* 159, 183–201.
- Tondi, R., de Franco, R., Barzaghi, R., 2000. Sequential integrated inversion of refraction and wide-angle reflection traveltimes and gravity data for two-dimensional velocity structures. *Geophysical Journal International* 141, 679–698.
- Tondi, R., Schivardi, R., Molinari, I., Morelli, A. Upper mantle structure below the European continent: constraints from surface wave tomography and GRACE satellite gravity data. *Journal of Geophysical Research*, <http://dx.doi.org/10.1029/2012JB009149>, in press.