An efficient algorithm for double-difference tomography and location in heterogeneous media, with an application to the Kilauea volcano

Vadim Monteiller and Jean-Luc Got

Laboratoire de Géophysique Interne et Tectonophysique, Université de Savoie, Le Bourget-du-Lac, France

Jean Virieux

Géosciences Azur, Sophia-Antipolis, Valbonne, France

Paul Okubo

U.S. Geological Survey, Hawaiian Volcano Observatory, Hawaii National Park, Hawaii, USA

[1] Improving our understanding of crustal processes requires a better knowledge of the geometry and the position of geological bodies. In this study we have designed a method based upon double-difference relocation and tomography to image, as accurately as possible, a heterogeneous medium containing seismogenic objects. Our approach consisted not only of incorporating double difference in tomography but also partly in revisiting tomographic schemes for choosing accurate and stable numerical strategies, adapted to the use of cross-spectral time delays. We used a finite difference solution to the eikonal equation for travel time computation and a Tarantola-Valette approach for both the classical and double-difference three-dimensional tomographic inversion to find accurate earthquake locations and seismic velocity estimates. We estimated efficiently the square root of the inverse model's covariance matrix in the case of a Gaussian correlation function. It allows the use of correlation length and a priori model variance criteria to determine the optimal solution. Double-difference relocation of similar earthquakes is performed in the optimal velocity model, making absolute and relative locations less biased by the velocity model. Double-difference tomography is achieved by using high-accuracy time delay measurements. These algorithms have been applied to earthquake data recorded in the vicinity of Kilauea and Mauna Loa volcanoes for imaging the volcanic structures. Stable and detailed velocity models are obtained: the regional tomography unambiguously highlights the structure of the island of Hawaii and the double-difference tomography shows a detailed image of the southern Kilauea caldera−upper east rift zone magmatic complex.

Citation: Monteiller, V., J.-L. Got, J. Virieux, and P. Okubo (2005), An efficient algorithm for double-difference tomography and location in heterogeneous media, with an application to the Kilauea volcano, *J. Geophys. Res.*, *110*, B12306, doi:10.1029/2004JB003466.

1. Introduction

[2] Using earthquake data for describing Earth structure with the best possible resolution is a long-term quest in seismological research. After the pioneering work of *Geiger* [1910] and *Aki and Lee* [1976], earthquake location and tomography have become widely used means to infer active and passive structures of the Earth interior from the available seismological data. A large effort has since been devoted by seismologists to improving the knowledge of crustal or lithospheric structures through the imaging of seismic velocities and earthquake locations. Whereas on a global scale, earthquake location makes the imaging of the major geodynamic features of the Earth possible, its accuracy remains too low to allow the description of many seismotectonic features on smaller scales. Similarly, earthquake tomography is often revealed to be unable to provide high-resolution imaging of Earth structures. On small scales, the unfavorable geometry of both the earthquake distribution and seismic networks combined with the nonlinear character of the tomographic problem generates trade-offs between model parameters and induces uncertainties in earthquake location and seismic velocity estimates.

[3] Double-difference relocation methods [e.g., *Jordan and Sverdrup*, 1981; *Poupinet et al.*, 1984; *Ito*, 1985; *Got et al.*, 1994; *Slunga et al.*, 1995; *Shearer*, 1997; *Rubin et al.*, 1999; *Waldhauser and Ellsworth*, 2000] (see *Wolfe* [2002] for a discussion of double-difference algorithms) have in the

last decade shown that seismogenic region knowledge may be deeply changed by the accurate relocation of earthquakes. However, we have little experience with the limitations of such methods and are sometimes unreasonably confident in their results. These relocations are often performed in approximate one-dimensional (1-D) velocity models, although little is known about the effect of such an approximation on relocation results. There is therefore some interest in improving the reliability of double-difference earthquake relocation in heterogeneous media by the use of more realistic seismic velocity models [*Michelini and Lomax*, 2004]. There also may be some interest in profiting from the accuracy provided by the double-difference approach to improve the tomography. *Zhang and Thurber* [2003], *Zhang et al.* [2004], and *Thurber et al.* [2004] made applications of double-difference tomography in various contexts. The present work aims at exploring the possibility of improving the tomography by the use of spatially extended sets of similar events allowing accurate cross-spectral time delay measurements. Travel time difference residuals are second-order quantities which may be small and should be accurately estimated to be interpretable. Accuracy at each step of the computation is therefore a key question to really succeed in this task. It is therefore required to design a stable and accurate method for the classical and double-difference tomography in very heterogeneous media, that is, to choose appropriate schemes for both the direct and the inverse problem. Travel time computation in heterogeneous media may be performed by the use of finite difference methods [see, e.g., *Vidale*, 1988]. We chose *Podvin and Lecomte*'s [1991] robust eikonal solver for this computation. A robust solution to the inverse problem is achieved by using the LSQR solver [*Paige and Saunders*, 1982]: from the strict point of view of the use of these tools, the classical part of our tomographic procedure is close to the one published by *Benz et al.* [1996]. In this work, however, we follow the approach of *Tarantola and Valette* [1982] based on the probabilistic optimization theory for estimating the model parameters. It provides a unique and coherent framework for both the tomographic and double-difference relocation processes. Double-difference equations are therefore solved to find earthquake positions and seismic velocities. This method is applied for processing earthquake data of Kilauea and Mauna Loa volcanoes recorded by the permanent seismic network of the U.S. Geological Survey (USGS) Hawaiian Volcano Observatory (HVO). These volcanoes are indeed very good targets for tomographic studies, due to the strong velocity contrasts already evidenced, to the relatively simple geological setting, and to the quantity and very good quality of the data recorded by the HVO permanent seismic network. Is it possible to infer the fine structure of the Kilauea magmatic system by double-difference tomography? If so, what may be the geometry of the magmatic system around Kilauea caldera and east rift zone? Answers to these questions are useful to set up mechanical models needed to deeply understand the Hawaiian volcanoes long-term dynamics.

2. Method

2.1. Principle of the Algorithm

[4] The conventional presentation for finding the location of an earthquake from arrival times observed at many receivers is due to *Geiger* [1910]. This nonlinear problem is solved using a linearized approach. This method has been extended to the determination of both hypocenters and velocity field parameters by *Aki and Lee* [1976]. Travel time residual $r_i^k = t_{\mathrm{obs}_i}^k - t_i^k$, (where $t_{\mathrm{obs}_i}^k$ and t_i^k are the observed and theoretical travel time, respectively, for an event i recorded at station k) is related to perturbations $\delta \mathbf{m}_i$ of the model parameter vector $\mathbf{m}$ by a linear equation:

$$\frac{\partial t_i^k}{\partial \mathbf{m}} \delta \mathbf{m}_i = r_i^k \tag{1}$$

or, in matrix notation,

$$\mathbf{G}\, \delta \mathbf{m} = \mathbf{r} \tag{2}$$

where $\mathbf{G}$ is the matrix of the partial derivatives of the theoretical travel times with respect to the model parameters. The term $\delta \mathbf{m}$ is the model perturbation vector and $\mathbf{r}$ is the travel time residual vector. Model parameters are the hypocentral parameters for the earthquake location, hypocentral parameters and velocity parameters in the earthquake tomography.

[5] This linearized approach has been followed for double-difference location [*Waldhauser and Ellsworth*, 2000]. Following *Waldhauser and Ellsworth* [2000], we will name

$$\frac{\partial t_i^k}{\partial \mathbf{m}} \delta \mathbf{m}_i - \frac{\partial t_j^k}{\partial \mathbf{m}} \delta \mathbf{m}_j = r_{ij}^k \tag{3}$$

the double-difference equation for the events i and j, recorded at the station k; r_{ij}^k is the travel time difference residual ("double difference"). In the double-difference location, the model parameters are hypocentral parameters. In double-difference tomography [*Zhang and Thurber*, 2003; *Monteiller et al.*, 2003], they comprise both hypocentral and velocity parameters. The double-difference equation system may therefore be written

$$\mathbf{G}\, \delta \mathbf{m} = \mathbf{r} \tag{4}$$

where $\mathbf{G}$ is the matrix of the partial derivatives of the theoretical travel time differences with respect to the model parameters, $\delta \mathbf{m}$ is the model perturbation vector, $\mathbf{r}$ is the travel time difference residual vector.

[6] The system expressed by equation (4) being generally ill-conditioned, *Waldhauser and Ellsworth* [2000] and *Zhang and Thurber* [2003] have chosen to solve the system

$$\mathbf{W}\begin{pmatrix} \mathbf{G} \\ \lambda \mathbf{I} \end{pmatrix} \delta \mathbf{m} = \mathbf{W}\begin{pmatrix} \mathbf{r} \\ \mathbf{0} \end{pmatrix} \tag{5}$$

where $\mathbf{W}$ contains a priori quality weights. $\mathbf{G}$ is the matrix containing the partial derivatives of the travel time differences [*Waldhauser and Ellsworth*, 2000] or the partial derivatives of the travel times and travel time differences [*Zhang and Thurber*, 2003] with respect to the model parameters; λ is a damping factor, $\mathbf{I}$ is the identity matrix, and $\mathbf{r}$ is the vector of residuals in travel time differences [*Waldhauser and Ellsworth*, 2000] or travel times and travel time differences [*Zhang and Thurber*, 2003]. *Waldhauser*

and Ellsworth [2000] add the constraint (for a N event cluster)

$$\sum_{i=1}^{N} \delta \mathbf{m}_i = 0 \tag{6}$$

Earthquake location and tomography are, however, simultaneously nonlinear and ill-posed. Determining the hypocenter location and the seismic velocities is an optimization problem whose solution has been discussed by *Tarantola and Valette* [1982] and *Tarantola* [1987].

[7] The tomographic problem should actually be posed

$$g(\mathbf{m}) = \mathbf{d} \tag{7}$$

where *g* represents the functional used to solve the direct problem and **d** is the travel time data.

[8] *Tarantola and Valette* [1982] proposed minimizing a cost function which is a weighted sum of the data misfit and a penalty function chosen to be the a posteriori model variance:

$$[g(\mathbf{m}) - \mathbf{d}]^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1} [g(\mathbf{m}) - \mathbf{d}] + (\mathbf{m} - \mathbf{m}_0)^{\mathrm{T}} \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_0) \tag{8}$$

where $\mathbf{C_d}$, $\mathbf{C_m}$, and $\mathbf{m}_0$ are the data covariance matrix, the a priori model covariance matrix and the a priori model vector, respectively. In the case the Gauss-Newton scheme is used to minimize the cost function (8); the solution at iteration *k* is, in the case of the nonlinear problem:

$$\mathbf{m}_{k+1} - \mathbf{m}_k = \left(\mathbf{G}^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1} \mathbf{G} + \mathbf{C}_{\mathbf{m}}^{-1} \right)^{-1}$$
$$\cdot \left[\mathbf{G}^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1} (\mathbf{d} - g(\mathbf{m}_k)) + \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m}_0 - \mathbf{m}_k) \right] \tag{9}$$

which may be written, in terms of the perturbation vector $\delta \mathbf{m}_k = \mathbf{m}_{k+1} - \mathbf{m}_k$,

$$\left(\mathbf{G}^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1} \mathbf{G} + \mathbf{C}_{\mathbf{m}}^{-1} \right) \delta \mathbf{m}_k = \mathbf{G}^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1} [\mathbf{d} - g(\mathbf{m}_k)] + \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m}_0 - \mathbf{m}_k) \tag{10}$$

an equation which may be understood as

$$\left(\mathbf{G}^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1/2} \quad \mathbf{C}_{\mathbf{m}}^{-1/2} \right) \begin{pmatrix} \mathbf{C}_{\mathbf{d}}^{-1/2} \mathbf{G} \\ \mathbf{C}_{\mathbf{m}}^{-1/2} \end{pmatrix} \delta \mathbf{m}_k$$
$$= \left(\mathbf{G}^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1/2} \quad \mathbf{C}_{\mathbf{m}}^{-1/2} \right) \begin{pmatrix} \mathbf{C}_{\mathbf{d}}^{-1/2} (\mathbf{d} - g(\mathbf{m}_k)) \\ \mathbf{C}_{\mathbf{m}}^{-1/2} (\mathbf{m}_0 - \mathbf{m}_k) \end{pmatrix} \tag{11}$$

Therefore minimizing equation (6) is equivalent to solving iteratively the system

$$\begin{pmatrix} \mathbf{C}_{\mathbf{d}}^{-1/2} \mathbf{G} \\ \mathbf{C}_{\mathbf{m}}^{-1/2} \end{pmatrix} \delta \mathbf{m}_k = \begin{pmatrix} \mathbf{C}_{\mathbf{d}}^{-1/2} (\mathbf{d} - g(\mathbf{m}_k)) \\ \mathbf{C}_{\mathbf{m}}^{-1/2} (\mathbf{m}_0 - \mathbf{m}_k) \end{pmatrix} \tag{12}$$

In the following, we will see that this numerical approach is appropriate for the travel time tomography, the double-difference relocation, and the double-difference tomography.

2.2. Estimating Accurate Model Parameters From Travel Time and Time Delay Tomography

[9] Even well-designed computational schemes may lead to poor results when used with inadequate numerical strategies, unsuited parameters or insufficient data. In this

paragraph, we will focus on the accuracy of the direct problem (travel time computation) and on the estimation of the model parameters from the inversion of travel time data. The conditioning of the inversion is a key notion in our study. Condition number cond is defined as the product of the norm of a square matrix by the norm of its inverse and may be calculated as the ratio between its major and minor eigenvalues. The relation [see, e.g., *Tarantola*, 1987]

$$\left\| \frac{\delta \mathbf{m}}{\mathbf{m}} \right\| \leq \text{cond} \left\| \frac{\delta \mathbf{d}}{\mathbf{d}} \right\| \tag{13}$$

shows the importance of the condition number in the estimation of model parameters. Reaching a high relative accuracy on **m** implies keeping a high relative accuracy on **d** and a low condition number.

[10] When using double differences, the data **d** may become small and sensitive to uncertainty, especially when determined from travel time differences. Alternatively, cross-spectral measurements of the time delays are measurements independent of the arrival time picking error. They are an order of magnitude better than arrival time picking differences. Such data allow us to keep the relative accuracy $\|\delta \mathbf{d/d}\|$ to reasonable values (generally less than 10%), and will be used in this work. In the following we will see which conditions have to be filled to keep a correct conditioning while using cross-spectral time delays of similar events.

2.2.1. Robust and Accurate Travel Time Computation in a Heterogeneous Medium

[11] As mentioned in section 1, this work is mostly motivated by inferring seismogenic structures in fault zones and volcanoes from travel times and time delays. Such structures are strongly heterogeneous and therefore require a robust solution for travel time computation. Finite difference solutions to the eikonal equation have been proven to provide robust travel time computations [e.g., *Vidale*, 1988; *Podvin and Lecomte*, 1991]. In this work we used the *Podvin and Lecomte* [1991] finite difference scheme to compute the entire first-arrival travel time field. This method allows an accurate computation of travel times in complex media, provided that an appropriate finite difference grid is used. A posteriori ray tracing is performed by computing the travel time field gradient, and travel time derivatives as well as more accurate travel times are then calculated. This travel time recomputation allows the accuracy to reach the $O(10^{-4})$ needed to perform double-difference computations at every scale with low computational costs (Figure 1).

2.2.2. Understanding and Parameterizing the Double-Difference Tomographic Problem
2.2.2.1. Constraining Absolute Locations With Double-Difference Equations

[12] Understanding the double-difference relocation problem is necessary to undertake double-difference tomography correctly. In the case of two events in close proximity (a doublet), the analysis of the eigenvalues of the double-difference equation system shows that one eigenvalue is exactly zero and corresponds to the reference time, and three others, corresponding to the geometrical center coordinates, may be very low [see also *Wolfe*, 2002; *Menke and Schaff*, 2004]. Choosing a reference time, using an average origin time constraint (equation (6) limited to the origin time
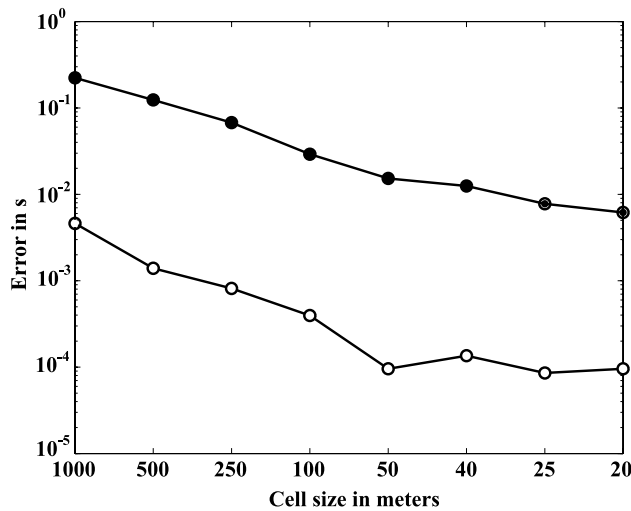
Figure 1. Relative accuracy of a posteriori finite difference travel time recomputation, as a function of the discretization of the finite difference grid. Bold circles indicate error after Podvin-Lecomte finite difference computation of travel times; open circles indicate error after a posteriori ray tracing in the travel time field and recomputation of travel times. The reference travel times are computed using the analytic formulation of *Evjen* [1936] for a gradient model.

perturbations), or choosing differential ΔT_{0ij} parameters for the origin time eliminates the corresponding zero eigenvalue. The three eventually low eigenvalues decrease to zero and conditioning strongly increases (Figure 2) when ρ/Δ tends to zero: slowness vectors for two events i and j tend to be subparallel for each station k and may not allow the determination of the geometric center coordinates. To overcome this problem, an intuitive approach may be to hold one event, or the geometric center of a N event cluster fixed (equation (6)) [e.g., *Waldhauser and Ellsworth*, 2000]. However, this constraint may introduce a bias in the estimate of the absolute position and noise in the relative position computation. In the work by *Waldhauser and Ellsworth* [2000], equation (6) is downweighted to let the geometric center move. The latter weight is therefore a critical parameter for estimating the geometric center position. To help constraining the cluster geometric center location without using equation (6), *Zhang and Thurber* [2003] solve simultaneously the weighted absolute and double-difference equations.

[13] The eigenvalue analysis (Figure 2) shows that a sufficient interevent distance allows the computation of the geometric center location. *Menke and Schaff* [2004] show that in favorable cases, double-difference equations constrain absolute locations better than absolute equations. Conditions that favor this constraint are (1) the existence of extended sets of events for which cross-correlation time delay measurements are possible and the absence of small-extent isolated event subsets and (2) the high quality of the velocity model. A favorable condition to get accuracy in both relative and absolute locations is to constitute dense, continuous, and extended correlated earthquake sets. Combining a sharp distance weighting with damping may be

unfavorable to providing a solution for the cluster geometric center.

[14] The simultaneous resolution of absolute and double-difference equations is therefore not strictly necessary, not only for finding relative positions but also for finding absolute positions. It is the reason for which we will not include the absolute equations nor the fixed geometric center equation (equation (6)) in the resolution of the double-difference equation system. It does not mean that absolute information is not needed in an earlier step of the process. To stabilize the solution in the case of small isolated event subsets, we introduce a priori information (equation (12)) concerning the (absolute) location of each event: the a priori location before double-difference relocation is the absolute location in the 3-D velocity tomographic model. Notice that a priori information is of importance only when the double-difference data do not constrain the solution. Control of the amount of a priori information needed is made through the use of the covariance operator $\mathbf{C_m}$. For the computation of hypocentral parameters, $\mathbf{C_m} = \mathbf{C_h} = \sigma^2 \mathbf{I}$, where σ^2 is the a priori variance on the hypocentral parameters, $\mathbf{I}$ is the identity matrix. It leads to a simple optimization scheme for controlling the numerical quality of the solution and the amount of a priori information in the solution. This approach also allows the accurate computation of both relative and absolute positions without needing to a priori cluster the events (as would be needed if equation (6) was used). This strategy benefits from both the almost linear character of double-difference equations for
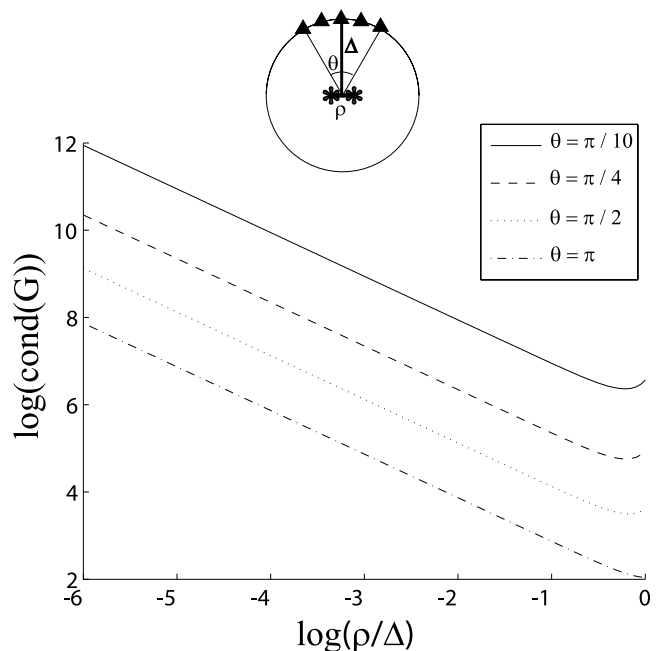

Figure 2. Conditioning of the partial derivative matrix $\mathbf{G}$ as a function of the aperture ρ/Δ of an earthquake doublet, where ρ is the interevent distance and Δ is the hypocentral distance. Conditioning is given for various circular seismic station network geometry, characterized by their aperture θ, in the interval $\rho/\Delta < \theta$. In the top sketch, triangles represent seismic stations and stars represent earthquakes.

closeby events and the well-conditioned nonlinear character of these equations for farther events.

2.2.2.2. Parameterizing the Tomographic Problem

[15] Given the amount of data, we chose to perform the inversion in the model space. In the classical approach, parameterization has to be chosen carefully as it controls the conditioning of the problem, and finally the amount of damping to be used. Condition number strongly increases when cells of the tomographic model are not sampled by seismic rays. However, the minimal solvable wavelength in a tomographic model is controlled by the data, not by the discretization of the model. In the following paragraph we will see that using a Tarantola-Valette approach converts the problem of choosing a parameterization to the problem of choosing a correlation length. We will therefore choose the finest possible discretization for the velocity model.

[16] Improvement in the solution may arise from the improvement in time measurement if conditioning of the double-difference equation system is sufficient (equation (13)). For this reason, and to avoid competition between absolute and cross-spectral data, we will not solve simultaneously the absolute and double-difference equations. Correlated event sets often appear as collections of clusters. Intercluster correlation may be (far) lower than intracluster correlation. This feature means the eigenvalue spectrum of the double-difference equation system may contain low eigenvalues corresponding to the geometric centers of the clusters. Overweighting of well-correlated data [e.g., *Got et al.*, 1994], or distance weighting [*Waldhauser and Ellsworth*, 2000] reinforces clustering and degrades conditioning. The combination of damping and distance weighting tends to break an initial large-extent event set into several smaller-extent subsets by removing weak critical links [see also *Wolfe*, 2002]. This process reduces the spatial extension and resolution of the velocity perturbations. It induces a trade-off between average velocity perturbation along the ray path and geometric center location, and unstable velocity fluctuations may appear.

[17] 1. We use large-extent correlated event sets that are as homogeneous as possible. Notice that weak critical links may be avoided by linking critical pairs by using travel time differences.

[18] 2. We search for the optimal covariance matrix (see sections 2.2.3 and 3.2.2) which provides the a priori information needed to minimize the cost function (equation (8)). Double-difference equations constrain the model when they contain the necessary information. An adequate, optimal amount of a priori information is used, stabilizing the solution when necessary.

2.2.3. An Efficient Preconditioning of the System Prior to Inversion

2.2.3.1. Selection of the Earthquakes as a Preconditioning Step

[19] Preconditioning is generally understood to be matrix handling; it, however, begins with a relevant choice of data. In the regional tomography a significant step is the declustering and selection of the earthquakes in order to reach a homogeneous distribution of rays throughout the tomographic volume. Clustering produces subparallel rays that correspond to almost identical equations but whose replication bears no further information about the sampled inversion cells. The existence of such families of subparallel rays degrades the conditioning of the derivative matrix and thus the stability of the solution. Earthquakes are therefore chosen so as to have a spatial distribution as homogeneous as possible and to be recorded by the largest number of seismic stations.

2.2.3.2. Scaling

[20] The most usual and elementary way to improve the conditioning of a derivative matrix is to scale its columns. We therefore define three groups of parameters (position, origin time and velocity parameters) and we use a primary preconditioning by normalizing each column of the derivative matrix with the maximum value (taken from each group) of its L^2 norm.

2.2.3.3. A Priori Model Covariance Matrix

[21] Equation (12) implies the computation of $\mathbf{C}_m^{-1/2}$. A practical difficulty of this approach in the case of tomography is that the covariance matrix $\mathbf{C}_m$ may be huge, to the point its inverse is not directly computable. *Tarantola* [1987], however, noticed that $\mathbf{C}_m$ acts as a smoothing operator, therefore $\mathbf{C}_m^{-1}$ behaves like a roughening operator. This allows numerous authors to define $\mathbf{C}_m^{-1/2}$ as a differentiating (Laplacian) operator and to add a ("nearest neighbors") constraint:

$$m_i - \sum_{\text{neighbors}} w_j m_j = 0 \qquad (14)$$

(where w_j is a weighting factor) in place of the lower part of equation (12). However, the inverse squared of the matrix corresponding to the Laplacian operator does not exhibit a covariance matrix structure, nor does this approach allow the smoothing width to vary. Optimal ratios between groups of parameters (position, origin time, and velocities) also have to be determined from sensitivity tests [*Le Meur et al.*, 1997; *Latorre et al.*, 2004], which may be a tedious operation.

[22] The covariance matrix is composed of two parts. The first one is related to the hypocentral parameters. It is diagonal (hypocenters are a priori independent) and may be written $\mathbf{C}_h$. The second one is related to the velocity parameters and expresses the relations that may exist between them. In this work, we chose to quantify the correlation between cells located at positions $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$, respectively, by a function

$$\mathbf{C}_v(\boldsymbol{\rho}, \boldsymbol{\rho}') = \sigma_v^2 e^{\frac{-|\boldsymbol{\rho}-\boldsymbol{\rho}'|}{\lambda}} \qquad (15)$$

where σ_v^2 is the variance of the velocity for the cell located at $\boldsymbol{\rho}$, and λ is a correlation length.

[23] In this study we will assume, as a first approximation, that there is no a priori correlation between the hypocentral and velocity parameters. The a priori covariance matrix may therefore be expressed as

$$\mathbf{C}_m = \begin{pmatrix} \mathbf{C}_h & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_v \end{pmatrix} \qquad (16)$$

Coefficients of $\mathbf{C}_v^{-1/2}$ are computed via a Lanczos decomposition:

$$\mathbf{C}_v = \mathbf{U}\mathbf{S}\mathbf{U}^T \qquad (17)$$

where $\mathbf{C}_v$ is a symmetric positive definite matrix, $\mathbf{U}$ is an orthogonal matrix, $\mathbf{U}^T = \mathbf{U}^{-1}$, and $\mathbf{S}$ is a diagonal matrix containing the eigenvalues of $\mathbf{C}_v$.

Therefore

$$\mathbf{C}_v^{-1} = \left(\mathbf{USU}^T\right)^{-1} = \mathbf{US}^{-1}\mathbf{U}^T = \left(\mathbf{US}^{-1/2}\mathbf{U}^T\right)\left(\mathbf{US}^{-1/2}\mathbf{U}^T\right) \quad (18)$$

$$\mathbf{C}_v^{-1/2} = \mathbf{US}^{-1/2}\mathbf{U}^T \quad (19)$$

$\mathbf{C}_v^{-1/2}$ is found to exhibit a band-diagonal structure and the value of its coefficients is independent of the number of model parameters. This feature allows the construction of $\mathbf{C}_v^{-1/2}$ with a low computational expense. This approach is very convenient for tomographic purposes as it allows (see section 3.1.5) the choice of an optimal correlation length λ using a fixed cell size rather than using a multigrid approach or an unstructured mesh with a nearest neighbors smoothing. It is easy to use, as only two coefficients (λ and σ_v) control the quality of the tomographic inversion.

2.2.3.4. Data Covariance Matrix

[24] The travel time covariance matrix may be approximated from picking weights. Estimating confidence intervals for arrival times would be a useful improvement. In the case of cross-spectral time delays, an estimate of the time delay uncertainty is made from the computation of the coherency spectrum [see, e.g., *Got et al.*, 1994]. Statistics are finally applied to double-difference residuals and the data covariance matrix is updated. For this purpose, we use *Mosteller and Tukey*'s [1977] bisquare weighting:

$$W_i = \max^2\left[0, 1 - (r_i/\alpha r_{\mathrm{med}})^2\right] \quad (20)$$

where r_i is the residual for the ith data, and r_{med} is the median of residuals; the coefficient α is generally taken between 4 and 6.

2.2.4. Choosing an Adequate Statistical Law for Representing the Travel Time Residual Distribution

[25] The least squares solution presented from equations (8) to (12) and the LSQR algorithm can only be used if the travel time residuals have a Gaussian distribution. Arrival time picks and travel time residuals are, however, known to have a non-Gaussian long-tailed distribution due to the presence of very strong outliers, most time originating from confusing P and S first-arrival phases in the vicinity of nodal planes. Such outliers may strongly bias the L^2 norm estimation of the parameters. In that case, even downweighting large residuals [e.g., *Mosteller and Tukey*, 1977] may not prevent the solution from showing large biases because the initial solution and therefore the residuals may be strongly biased themselves.

[26] Alternatively, the hyperbolic secant function

$$f(x) = \frac{1}{\pi\sigma}\operatorname{sech}\left(\frac{x - x_0}{\sigma}\right) \quad (21)$$

where x is an independent variable, may represent a probability density function with mean x_0 and standard deviation $\pi\sigma/2$. It behaves like the L^1 distribution for large values of the residual, and like a Gaussian distribution for small values. It is therefore well suited for representing the

long-tailed arrival time pick and the travel time residual distribution [see, e.g., *Crase et al.*, 1990].

[27] Let us consider a datum x with the probability density $f(x)$ and mean x_0 and let us define an auxiliary datum

$$x' = \mathrm{Erf}^{-1}\left(2\int_{x_0}^{x} f(u)\,du\right) \quad (22)$$

where erf (x) is the error function. x' is necessarily a centered Gaussian variable with variance $^1/_2$. When $f(x)$ is the sech function (B. Valette and P. Lesage, Inferring mean Earth mechanical models from normal modes, mass and inertia: 1. Theoretical developments, submitted to *Geophysical Journal International*, 2005),

$$x' = \mathrm{erf}^{-1}\left[\frac{2}{\pi}\arctan\left(\sinh\frac{x - x_0}{\sigma}\right)\right] \quad (23)$$

$$\frac{\partial x'}{\partial x} = \frac{e^{x'^2}}{\sigma\sqrt{\pi}\cosh\left(\frac{x - x_0}{\sigma}\right)} \quad (24)$$

This choice for x' allows the rewriting of the inverse problem for the data x (with the sech probability density function) in terms of least squares inversion of x.

[28] This approach has proven to be more adapted to the L^2 norm tomographic inversion than the simple or weighted least squares inversion of the non-Gaussian travel time residuals.

2.2.5. General Algorithm

[29] As a summary, the general algorithm used in this work may be decomposed as follows: Step 1a is earthquake selection/declustering on a regional scale. Step 1b is the location of the selected earthquakes in the a priori 1-D velocity model, using hand-picked arrival times; it provides a priori earthquake locations for the regional tomography. Step 2 is earthquake travel time tomography on a regional scale; search for the optimal covariance matrix and the optimal solution (3-D velocity model and earthquake locations) by exploring a wide range of a priori model variances and correlation lengths. This solution will be used as a priori information in the double-difference relocation and tomography. Step 3a is a similar earthquake selection on a local scale. Step 3b is double-difference relocation on a local scale in the 3-D velocity model; building spatially extended correlated event sets. These relocations will serve as a priori information for the double-difference tomographic process. Step 4a is earthquake selection/declustering on a local scale. Step 4b is double-difference tomography on a local scale; search for the optimal covariance matrix and the optimal solution: final 3-D local velocity model and earthquake locations.

3. Application to the Kilauea Volcano, Hawaii

3.1. Regional-Scale Tomography: Example of the Kilauea and Mauna Loa Volcanoes, Hawaii

3.1.1. Volume of Interest

[30] This work aims at improving the seismic velocity tomography in a limited crustal volume illuminated by earthquake sources. However, such a study needs a correct
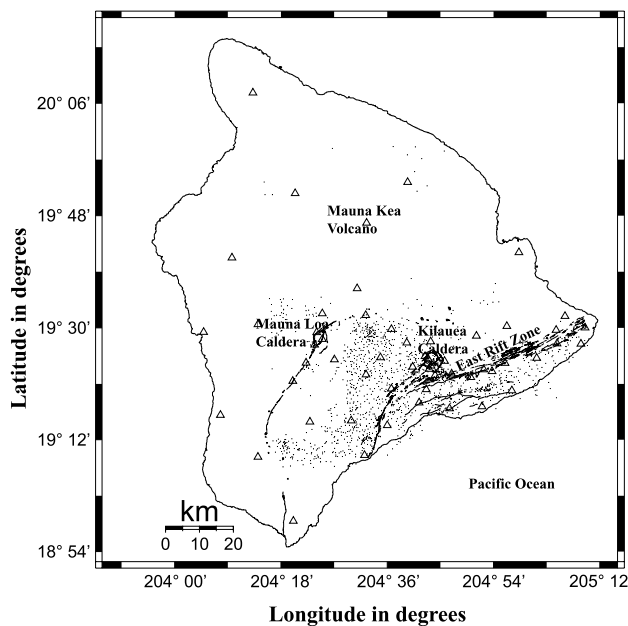
Figure 3. Map of island of Hawaii showing the main topographic features, the epicenters of the earthquake used for the regional-scale tomography (black dots), and the location of the HVO short-period seismic stations (open triangles).

knowledge of ray propagation in a large volume encompassing the sources and therefore requires performing larger-scale tomography. This large-scale tomography requires a specific data set. Clustering is generally avoided in tomographic studies, which require the optimization of the ray coverage in the tomographic volume. Quality of the tomographic results is first controlled by the geometry of the network and the distribution of earthquakes. Note that whereas in most tomographic studies a dense portable network is installed to record earthquakes during several months, we have chosen to use earthquake data issued from a permanent network. Density and homogeneity of the ray distribution is therefore reached not because of the seismic network density but rather because of the use of a very large set of earthquakes, that is, by using data spanning a long time period (10 years in our case). Given (1) the scale of the network (average interstation distance around Kilauea and Mauna Loa volcanoes is 15 km), (2) the spatial distribution of the earthquakes and especially their depth (typically 10 km), and (3) the need for a correct takeoff angle coverage, comprising enough refracted rays to accurately compute the earthquake depth, the adequate scale for travel time tomography is the regional scale. We therefore designed a volume of 160 km × 160 km × 36 km, which comprises the whole island of Hawaii (Figure 3). Finally, we choose to use a 1 km × 1 km × 1 km discretization of the 160 km × 160 km × 36 km inversion volume.

3.1.2. Data

[31] The data are arrival times and time delays from ~30,000 microearthquakes recorded from 1988 to 1999 by the USGS Hawaiian Volcano Observatory (HVO) seismic network, comprising up to 50 short-period seismic stations. Seismograms were digitized at a rate of 100

samples per second. Arrival times were manually picked with care at the observatory. From our initial 30,000 event data set, we extracted 1358 declustered earthquakes, recorded by at least 30 stations and spanning most of the seismogenic regions of island of Hawaii. The actual data contain an average of ~31 well-picked arrival times per event. These events are located using rays having a large range of takeoff angles. About two thirds of the rays are going up, one third going down. To perform an efficient declustering, the whole inversion volume has been discretized in 1 km^3 cubic cells, and earthquakes have been ranked by using the number of P readings in each cell. Only the first (best recorded) event of each list was selected. We therefore obtained a homogeneous initial distribution of well-located earthquakes throughout the seismogenic regions of island of Hawaii. Tomographic inversion was therefore performed using more than 41,886 high-quality P arrival times from 1358 well-located earthquakes and 959,077 inversion cells.

3.1.3. Initial Model

[32] The tomographic problem being nonlinear and possibly ill-conditioned, it is necessary to start conjugate gradient iterations in the vicinity of the true minimum of the cost function by using an appropriate initial model. Notice, however, that preconditioning correctly the model tends to smooth the cost function and to remove secondary minima. This is the reason for which convergence is fast (see section 3.1.5). The choice of the initial model and the notion of "vicinity" are less critical when the data constrain the model. *Kissling et al.* [1994] proposed to construct the best 1-D velocity model and station correction set from the inversion of earthquake arrival times. The average 1-D velocity model for Hawaii has been extensively studied by *Eaton* [1962], *Ryall and Bennett* [1968], *Hill* [1969], *Ward and Gregersen* [1973], *Crosson and Koyanagi* [1979], and *Klein* [1981]. Gradient models proposed by *Klein* [1981], *Okubo et al.* [1997], and *Benz et al.* [2003] seem to be adapted to the 1-D velocity modeling of this volcanic edifice: heterogeneity in rock composition is expected to exist both horizontally and vertically, and a large part of the mechanical heterogeneity is due to different degrees of compaction and is expected to create a vertical velocity gradient. In this work we will adopt the initial model already used in the same crustal volume by *Okubo et al.* [1997] and *Benz et al.* [2003] (Figure 4). Experiments with a unique gradient model extending from the surface (3.8 km/s) up to 14 km depth (7.6 km/s), and a constant 7.6 km/s half-space below lead to results identical to the ones presented here.

3.1.4. Sensitivity Tests

[33] The size of the model space leads us to estimate the (geometric) resolution of the tomographic inversion by representing the inversion sensitivity from checkerboard tests. Numerical experiments show that a checkerboard velocity model may be reconstructed for very favorable source-receiver geometries only. We therefore think that if the source-receiver geometry we use allows the reconstruction of a checkerboard velocity model, it will allow the reconstruction of the actual propagation medium from long wavelengths up to the checkerboard wavelength. Checkerboard tests, however, cannot be used to find the optimal correlation length to use in the actual travel time tomo-
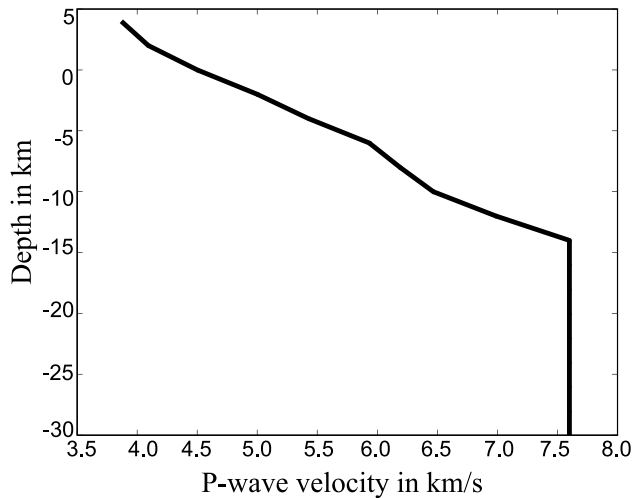
Figure 4. Initial velocity model used for the regional-scale tomographic study.

graphic inversion, as this parameter is also determined by the medium's propagation velocity spectrum.

[34] Checkerboard velocity model was built into the tomographic volume by applying a 200 m/s sinusoidal perturbation with a 12-km wavelength to the final velocity model. Theoretical travel times were computed with the actual source–station geometry and the ray distribution used in the regional tomography. The velocity model parameters were computed with a 1-km grid size. Earthquakes where located using a 1-km a priori variance.

[35] Figure 5 displays the result of this checkerboard test. It shows that providing an adequate spatial distribution of rays and an adapted correlation length are chosen, the tomographic inversion algorithm is able to reconstruct the propagating medium. Figure 5 suggests that the ray distribution mostly controls the spatial extent of the valid tomographic reconstruction. It shows that the velocity model may be correctly estimated in the central Mauna Loa–Kilauea caldera area at depths between 5 and 10 km.

3.1.5. P Travel Time Tomography Results

[36] Using the Tarantola-Valette approach with our approximation of the covariance matrix (section 2.2.4) enables us to explore the effect of changing the correlation length λ and the a priori model standard deviation σ_v. Exploring a wide range of values for λ and for σ_v allows us to find a family of solutions parameterized by λ and σ_v, which gives stable and reasonable velocity models. The variation of the cost function as a function of the penalty function (a posteriori model variance), for various values of the correlation length (Figure 6a) presents interesting features. The cost and penalty functions are indeed high when λ is small, due to the poor conditioning induced by the elementary cell size and the size of the model (whose parts are poorly sampled by rays). When λ increases (from 1 to 5 km), both functions decrease (the conditioning is improved, the solution is more accurate, the data misfit decreases; the model is smoother and the a posteriori model variance decreases) down to a minimum in the cost function (reached for $\lambda = 5$ km). Increasing λ (from 5 to 10 km) leads to an increase in the cost function (and data misfit). For longer λ, the

model variance also increases as the model is modified in a larger volume: the correlation function brings (coherent) information to cells that are poorly sampled by seismic rays. The cost function = f(penalty function) curve therefore presents a turning point located at its minimum. A quite similar behavior is shown by the data misfit equals f (penalty function) curve (Figure 6b). We define the optimal correlation length as the one corresponding to the minimum of the cost function equals f (penalty function) curve. For this value, limited but necessary information is brought in poorly sampled cells to reach a correct conditioning. The optimal correlation length is therefore adapted to the recoverable velocity spectrum (given the source-receiver geometry) of the propagation medium.

[37] The value of σ_v limits the amplitude of the seismic velocity variation of each cell. The variation of the cost function as a function of σ_v at $\lambda = 5$ km (Figure 6c) shows a minimum for $\sigma_v = 3$ km/s. However, models for $\sigma_v = 4$ km/s and $\sigma_v = 5$ km/s exhibit strong fluctuations when the velocity parameters are poorly constrained by data; $\sigma_v = 3$ km/s appears therefore to be a limiting value. The fact that the optimal "numerical" σ_v is larger than the physical value mainly means that the data correctly constrain the model and there is little need of a priori information. For σ_v values between 1 and 3, the models show little difference. We finally decided to retain a value of $\sigma_v = 1$ km/s, which is farther from numerical instabilities. A similar search of the optimal solution has been performed for the hypocentral parameters (Figure 6d), leading to an optimal value $\sigma_h = 10$ km, which shows they are geometrically well constrained by the data. Notice that this optimization approach gives tools to choose a model among the infinity of solutions proposed by the tomographic inversion.

[38] Tomographic inversion was run using P travel time data from the regional earthquake set. It was performed using 1 km^3 cubic cells, with $\lambda = 5$ km and $\sigma_v = 1$ km/s. Convergence of the inversion is extremely fast: 95% of the decrease in root-mean-square (RMS) is reached after only two iterations (Figure 7), which shows that the problem is not ill-conditioned nor strongly nonlinear in the vicinity of the chosen initial model. The RMS after two iterations drops to about 0.07s (initial RMS with the initial model and catalog locations: 0.3 s; RMS at the first tomographic iteration: 0.15 to 0.20 s). The velocity model is correctly resolved (Figure 5) between 5 and 10 km depth and is remarkably stable. Relative velocity variations range from −20% to 20%, with the strongest variations concentrated around the Kilauea and Mauna Loa calderas and rifts (Figure 8). In the regions that are not correctly resolved (Figure 5), the model is not subject to random fluctuations. It remains close to the initial model, excepted for volumes where some data are available (Figure 8).

3.2. Double-Difference Tomography on a Local Scale: Example of Kilauea's Caldera–Upper East Rift Zone

3.2.1. Data and Volume of Interest

3.2.1.1. Data

[39] In this section we will address the problem of imaging the magmatic system and its surrounding area by using double-difference tomography with a large set of well-located similar microearthquakes and accurate cross-spectral time delays. We decided to investigate the upper
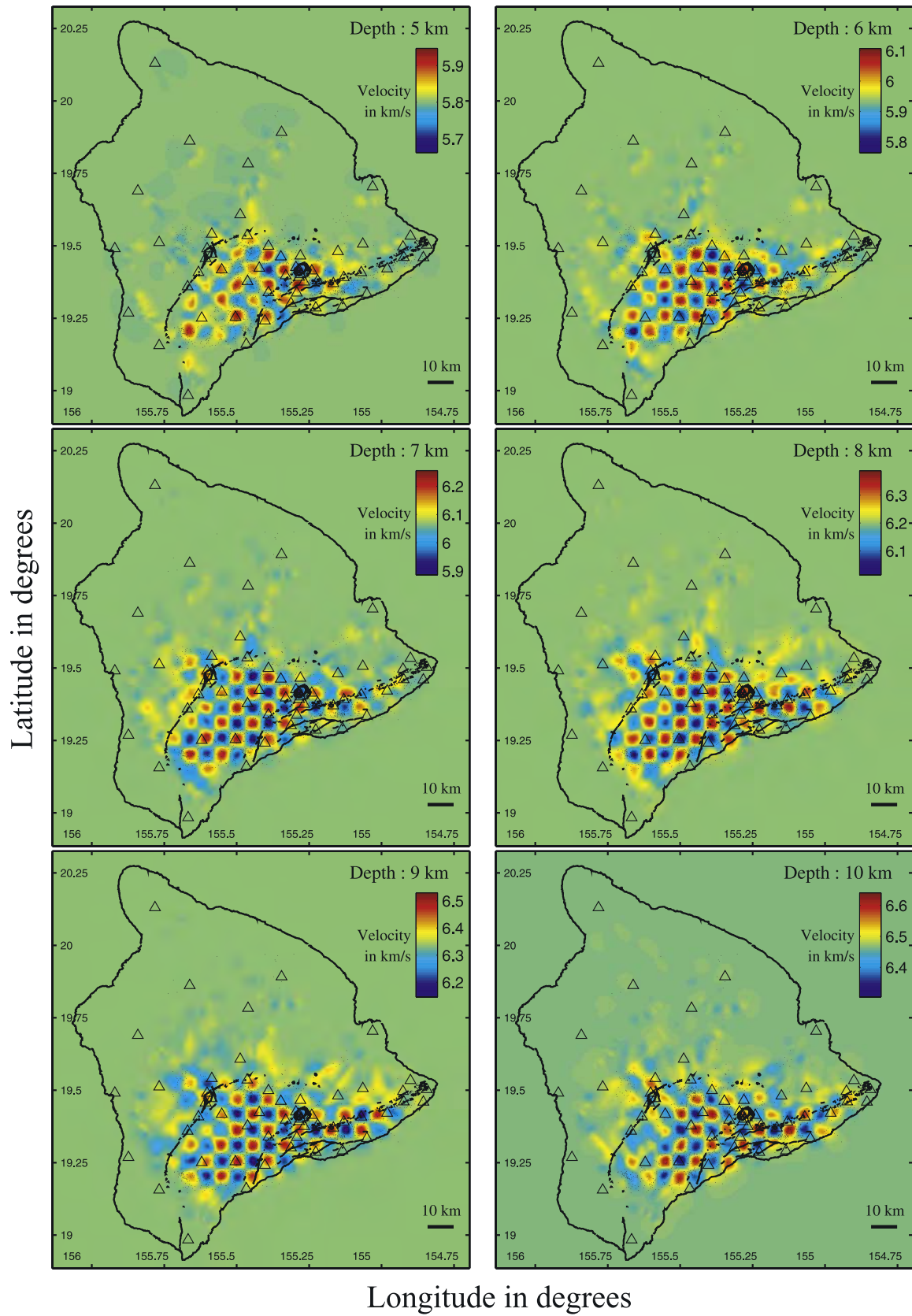
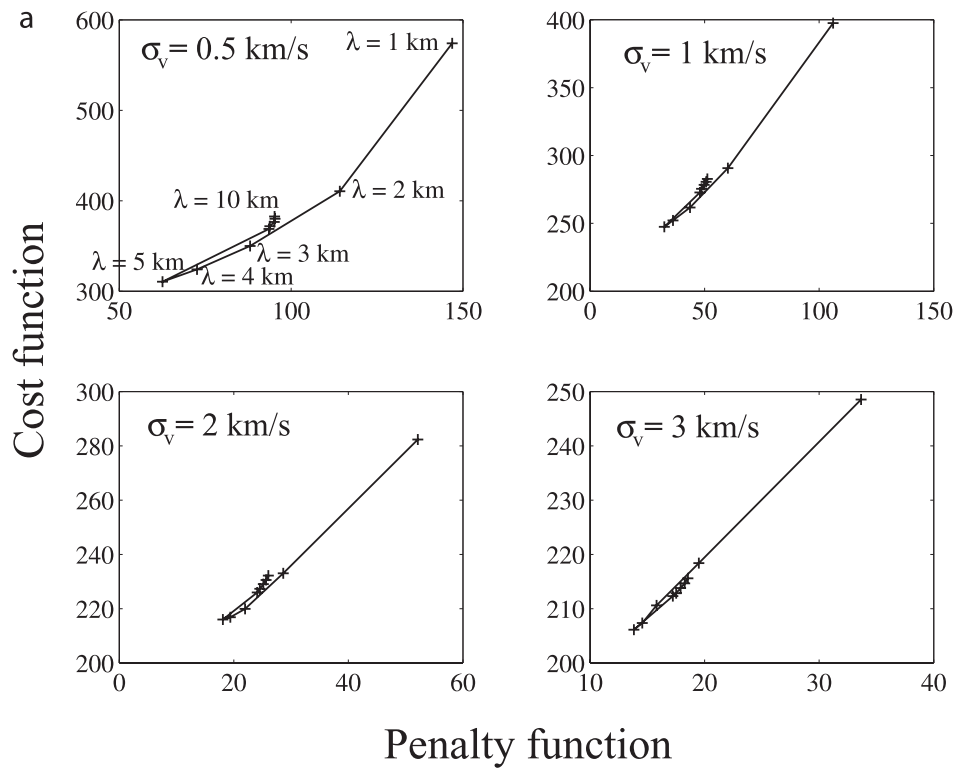**Figure 5.** Checkerboard test represented at regional scale and various depths.

Figure 6. (a) Cost function as a function of the penalty function (a posteriori model variance) for various correlation lengths and various a priori velocity model standard deviations (regional-scale tomography). (b) Data misfit as a function of the penalty function (a posteriori model variance) for various correlation lengths and various a priori velocity model standard deviations (regional-scale tomography). (c) Cost function as a function of the a priori velocity model standard deviation, for the correlation length $\lambda = 5$ km. (d) Cost function (solid line) and RMS (dashed line) as a function of the a priori standard deviation on spatial hypocentral parameters. Hypocentral parameters are computed in the initial 1-D velocity model and RMS is not weighted. Cost, penalty and data misfit functions (see equation (8)) are adimensional.

east rift zone, where the geometry of the rift zone is quite complex and adequately covered by seismic rays coming from well-located similar events. To that end, we chose to use seismic events occurring in the southeast of Kilauea caldera (southern caldera, upper east rift zone, south flank of Kilauea). In the south flank of Kilauea, the friction over geometrical heterogeneities of a subhorizontal plane provides a high rate of similar and well-recorded microearthquakes. In a preliminary study [*Got and Okubo*, 2003], about 2000 microearthquakes occurring between 1988 and 1999 showed enough similarity to many of their neighbors to be relocated with a typical average accuracy in relative relocations of about 50 meters horizontally and 100 meters vertically. This large database of similar earthquakes is used to build the extended set of correlated events needed to avoid the trade-off between velocity and hypocentral parameters (see section 2.2.3).

[40] We used similar events occurring in four extended correlated event sets of the upper east rift zone and the south flank of Kilauea volcano. The four event sets cover volumes between 7 km × 10 km × 4 km for the largest and 4 km × 1 km × 1 km for the smallest. All clusters were first adequately declustered, keeping enough similar and well-recorded events to have a direct and accurate estimation of cross-spectral time delays. Each event set provides a unique set of cross-spectral time delays. To build the sets of time

delays without favoring close-by events, we selected event pairs comprising the best recorded events that either (1) are highly coherent (>95%) for a large number (>25) of coherent arrivals and located with an interevent distance greater than a "low" threshold (50 m) or (2) have at least a minimum coherency (80%), and located with an interevent distance greater than a "high" threshold (larger than 250 m), decreasing from 2000 to 250 m with the number of coherent arrivals. The farther events are therefore accepted with less coherent arrivals than the closer. Class 1 ensures that data for close-by events are of high quality. Class 2 ensures that long-distance correlation will exist. We finally computed a histogram for the events satisfying these criteria and kept the 614 most frequent. This selection ensures the continuity of the event set over large distances. This data set presents good features for double-difference tomography: the events are widely correlated across the data set and well recorded. Time delays were computed by using a cross-spectral analysis performed on 2.56 s windows of the 100-sample per second digitized signal.

3.2.1.2. Volume of Interest

[41] The double-difference equation system (equation (4)) often generates very large sparse matrices. A compact set of n similar events recorded by p stations giving time delays for every event pair would indeed produce $n(n - 1)p/2$
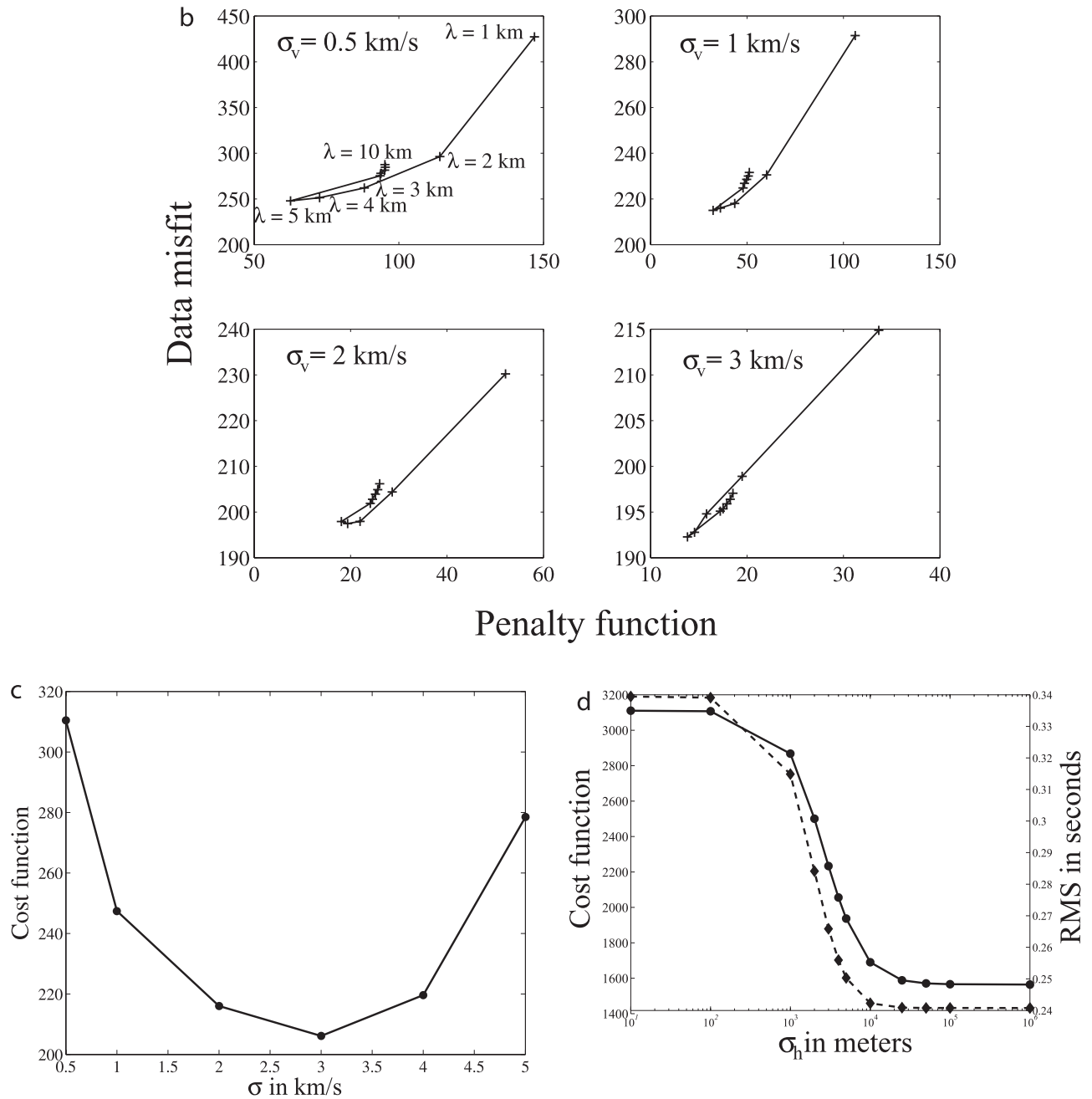
Figure 6. (continued)

equations relating the parameters of the problem. Even though each event is not correlated to the $n - 1$ others, and that each equation does not link the whole set of parameters, the large number of equations limits the volume of investigation. We therefore define three different grids. The first grid is defined over the regional (160 km × 160 km × 36 km) volume where the velocity model has been computed by travel time tomography. It is discretized with a 1 km sampling interval. The second grid is defined over the (local) volume of investigation. Geometrical constraints due to the event distribution lead us to define a 12 km × 16 km × 12 km volume. It is sampled with a 500-m interval and comprises 20,625 inversion nodes. The nodes located outside the local volume and

inside the regional volume are held fixed during the inversion. The third grid is the finite difference grid used for travel time computation, which is defined over subvolumes comprising the station and the set of sources, and sampled with a 250-m interval. The velocity field is interpolated at these nodes.

3.2.2. Double-Difference Tomography Results

[42] The checkerboard tests show that owing to the distribution of the seismic sources and receivers, the model can be estimated correctly in a limited volume (Figure 9). In this application, our interest will be focused on the southern caldera–upper east rift zone of the Kilauea volcano, where the checkerboard test indicates a correct resolution of the tomographic inversion.
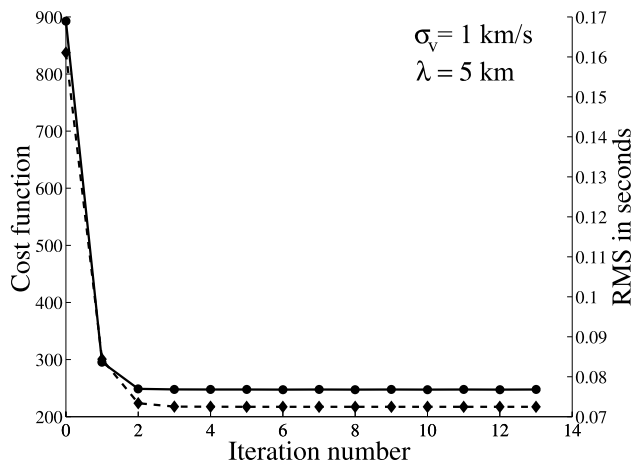
Figure 7. Cost function and RMS as a function of the iteration number during the tomographic process. Solid line indicates cost function; dashed line indicates weighted (*Mosteller and Tukey*'s [1977] weight with $\alpha = 5$, see text) RMS. The fast convergence indicates good conditioning and optimal a priori information.

[43] The selection of similar events allows an accurate computation of time delays by cross-spectral analysis. These time delays are used as data in the double-difference earthquake relocation and tomography. Earthquake relocations are first performed by double-difference relocation in the regional tomographic model, to be used as a priori information in the double-difference tomographic problem. A priori information for these double-difference relocations are the absolute locations obtained in the regional tomographic model from hand-picked arrival times, with large covariance estimates (10 km, Figure 10a). This relocation process provides stable and accurate locations, even though the events are declustered. Double-difference tomographic inversion is therefore performed by using the regional tomographic model (Figure 8) and the double-difference relocations as a priori information. A priori model standard deviation is chosen from the optimization procedure to be $\sigma_h = 1000$ m for the hypocenters, and $\sigma_v = 500$ m/s for the velocity parameters; correlation length is $\lambda = 1$ km. A detailed parametric study of the cost function equals f (penalty function) shows that $\lambda = 1$ km is the optimal value (Figure 10b). Recall that choosing a shorter correlation length does not ensure that the details of the model will be more accurately reconstructed: where the data constrain a small wavelength of the model, the optimal a priori information does not constrain the result. Optimal a priori information only stabilizes the result where the data do not constrain the model.

[44] The double-difference tomographic inversion significantly modifies the velocity image of the volume (Figures 11b and Figure S1 in the auxiliary material[1]), introducing small-wavelength variations in the initially smooth model (Figure 8). The change is stable, and the double-difference tomographic image exhibits a few random fluctuations. This image has to be compared with the results of the tomographic inversion of the travel times of the same set

------

[1]Auxiliary material is available at ftp://ftp.agu.org/apend/jb/2004JB003466.

of 614 events (Figure 11a) and the corresponding checkerboard test (Figure 9). In the area where the model is well estimated, double-difference tomographic results are strongly correlated to the surface geological structures of the Kilauea volcano. There is no systematic positive or negative change in the velocity model directly related to the distribution of earthquakes. RMS smoothly decreases with iteration number and the final RMS is of the same order of magnitude as the uncertainty on the time delay measurements (0.01 s).

4. Discussion

4.1. Regional-Scale Tomography

[45] Checkerboard tests tends to indicate that the velocity model is well reconstructed in most active areas of island of Hawaii, where stable and strong velocity contrasts are evidenced. Results of the regional-scale tomography very clearly reveal the main geological features of the island of Hawaii, namely the magmatic complexes forming the calderas and the rifts. These results are comparable to those obtained by *Rowan and Clayton* [1993], although the methods used are very different (Figure 9a, top, of *Rowan and Clayton* [1993] is, however, quite similar to our Figure 8, at the depth of 5 km). They confirm the results obtained by *Okubo et al.* [1997] and *Benz et al.* [2003]. These results may also be compared to the profiles obtained by *Haslinger et al.* [2001] and *Hansen et al.* [2004] in the east rift zone.

[46] The magmatic complexes appear as fast cores, surrounded by slow materials. This feature was already evidenced by all former authors and appears to be a stable tomographic characteristic of the island of Hawaii volcanism. The east rift zone of Kilauea volcano is unambiguously imaged from 3 km to 10 km depth, through a 30-km-long horizontal distance. Another striking feature is the well-defined slow body embedded between the faster magmatic complexes forming the Mauna Loa and the Kilauea calderas, which may correspond to the one identified by *Thurber* [1984]. A simple interpretation concerning these velocity contrasts and spatial distribution may be that fast cores correspond to the dense magma set in place at high pressure by dikes progressively expanding with time, as the Mauna Loa and Kilauea flanks move seaward. Low-velocity materials may be understood as being the result of the progressive superficial filling of volcano flanks by the erupted, decompressed lava. The well-defined slow body existing between Mauna Loa and Kilauea volcanoes may therefore be interpreted as a topographic valley filled by lava (which is in accordance with what can be seen at the surface), though the magma ridges progressively rose while lava was filling the valley. We also notice the strong horizontal velocity gradient in the south flank of Kilauea, which seems to coincide with the Hilina fault system up to ~9 km depth. This feature is also noticeable, more or less clearly, in the results of all the authors cited at the beginning of this paragraph.

4.2. Double-Difference Tomography

[47] Comparing the results of travel time and double-difference tomographic inversion (Figure 11) with the same
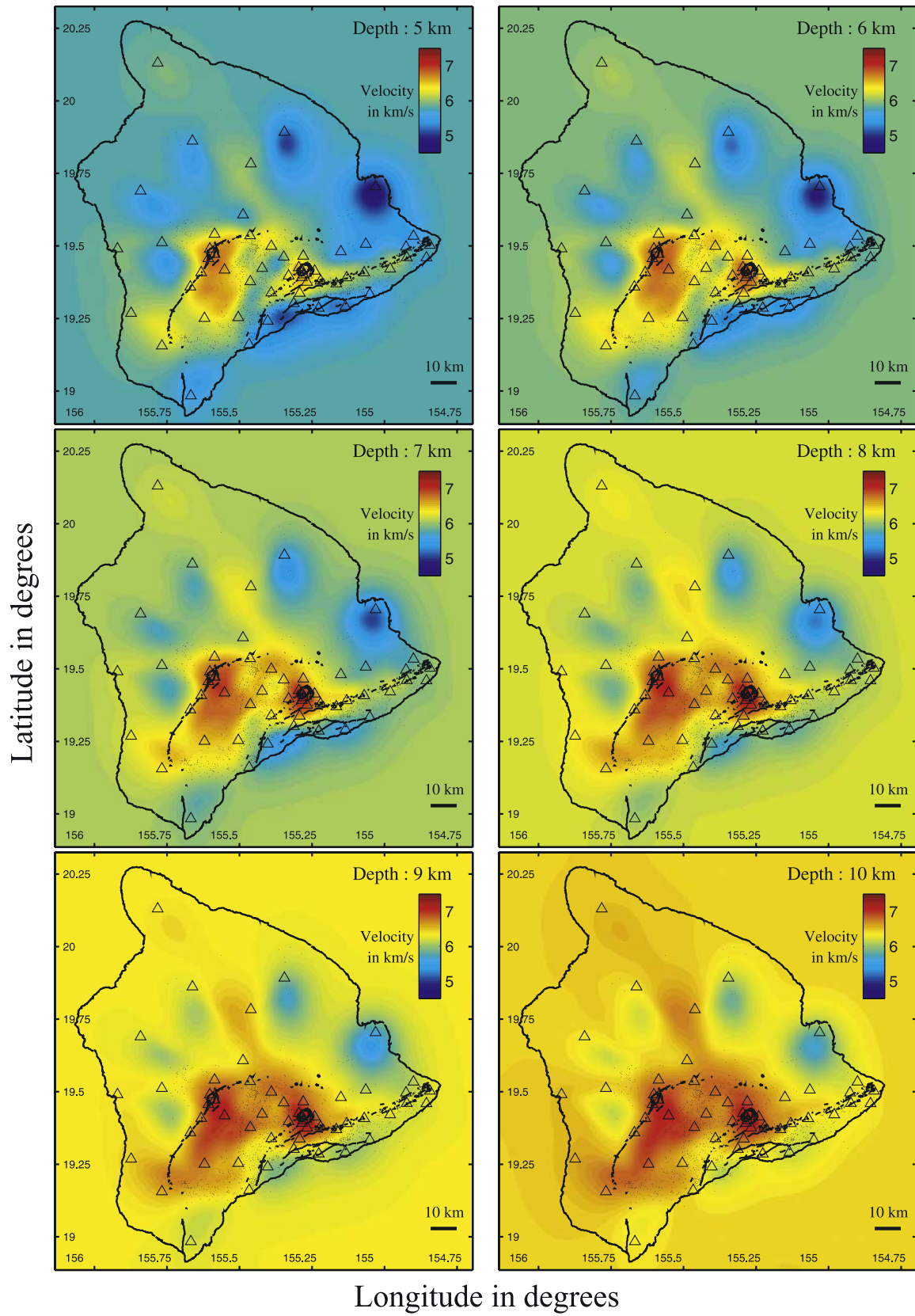
Figure 8. Maps at various depths of the *P* wave seismic velocity after tomographic inversion at regional scale ($\lambda = 5$ km, $\sigma_v = 1$ km/s).
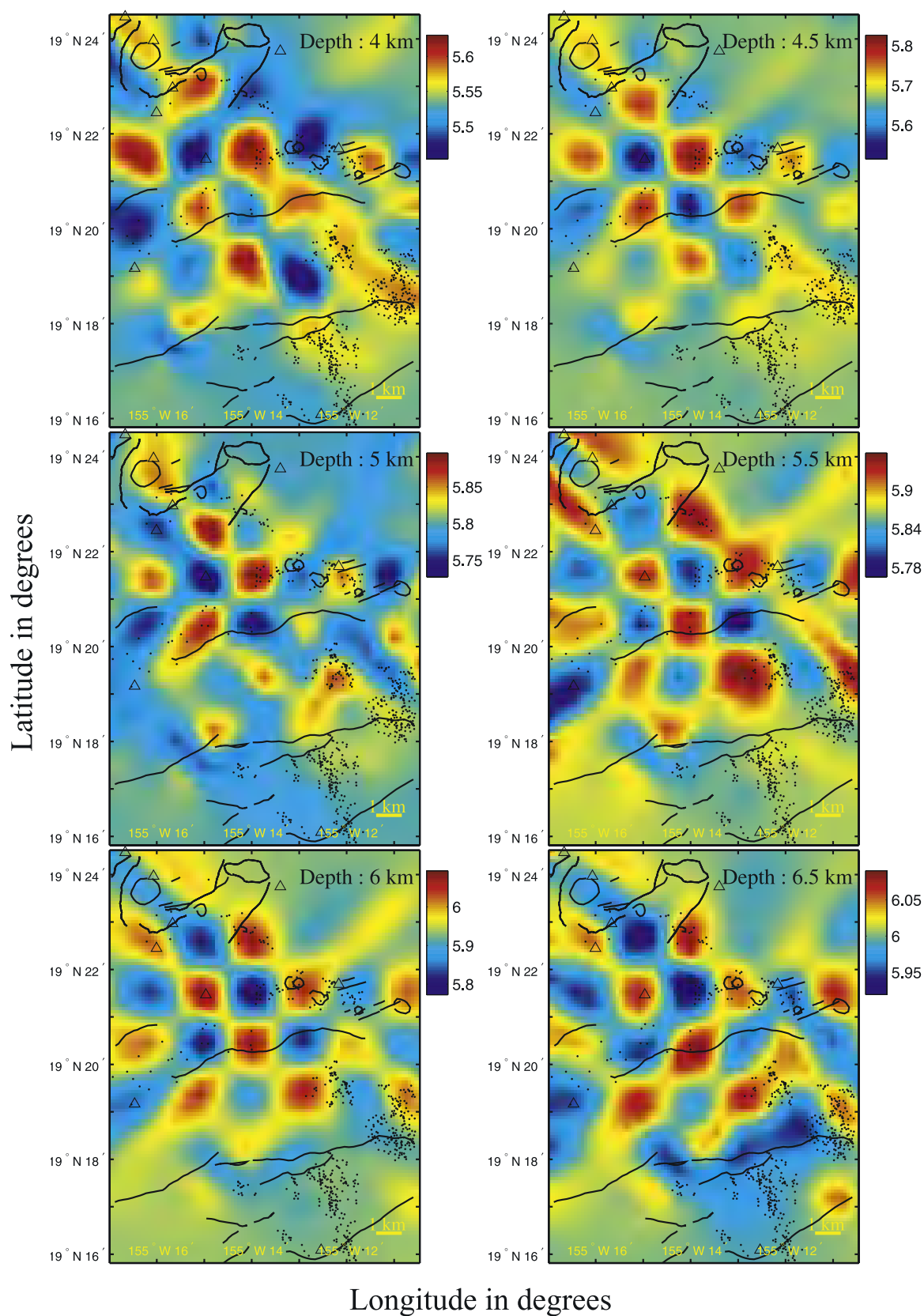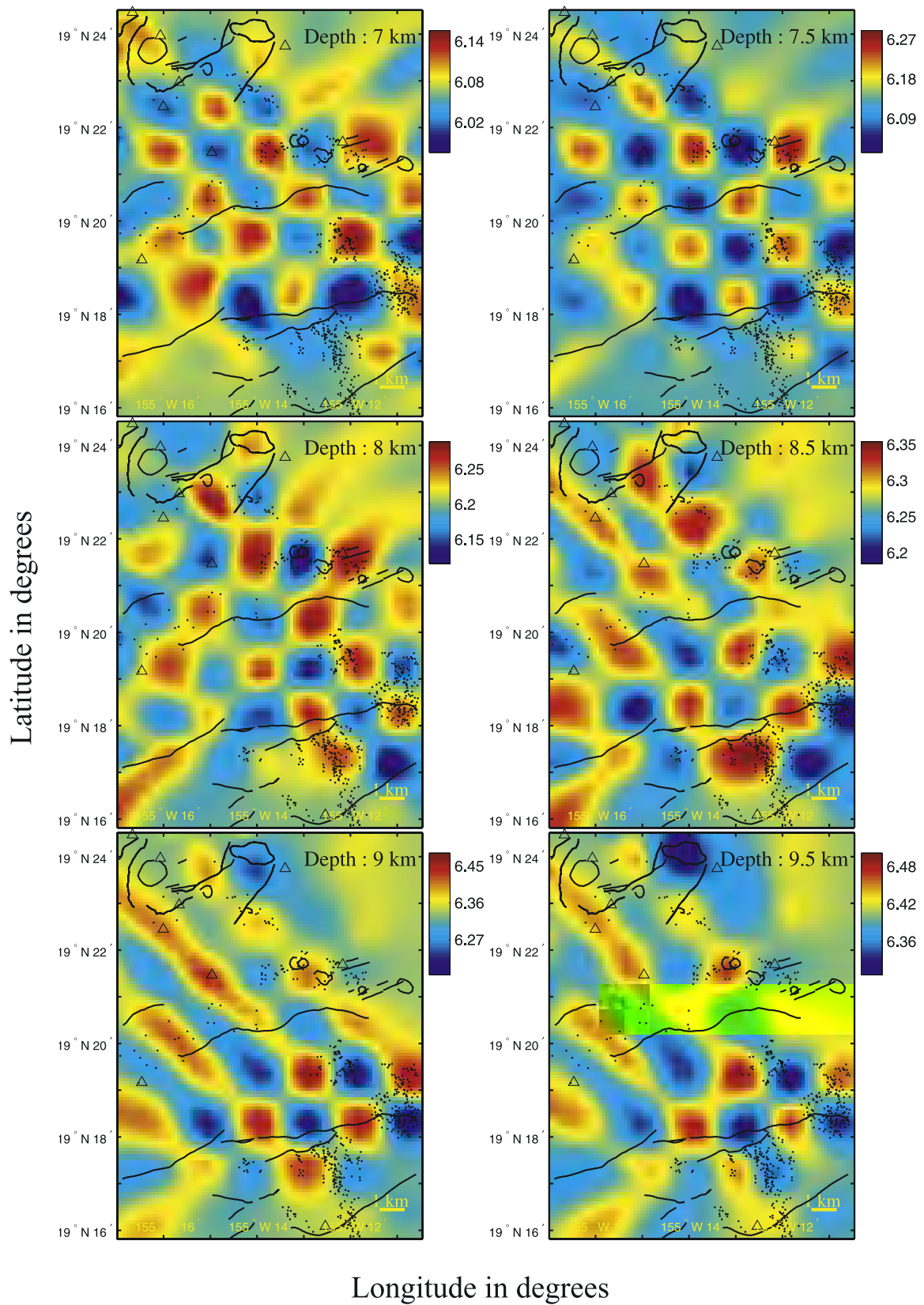
Figure 9. Checkerboard test (double-difference tomography) represented in the southeast of Kilauea caldera at various depths. Color scale indicates *P* wave velocity in km/s.
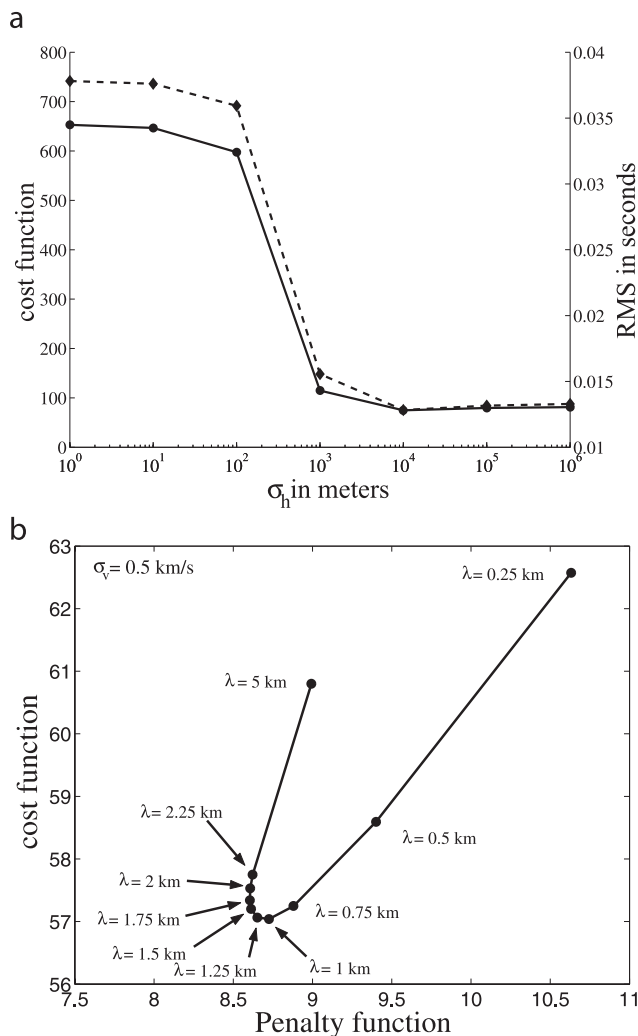
Figure 9. (continued)

a

b

Figure 10. (a) Cost function (solid line) and weighted RMS (dashed line) as a function of the a priori standard deviation on spatial hypocentral parameters (double-difference location in the regional velocity model). (b) Cost function represented as a function of the penalty function (a posteriori model variance), for various correlation lengths λ, and $\sigma_v = 0.5$ km/s (local double-difference tomography). Cost and penalty functions (see equation (8)) are adimensional.

events in the same volume shows significant changes in the tomographic image of that region, especially at shallow depths where the southern caldera–east rift zone region is well reconstructed (Figure 9). First notice that the double-difference tomographic image corresponds to coherent structures and does not display random high-frequency velocity changes. A striking feature of the double-difference tomographic inversion results is the presence, between 4 and 9 km depth, of a fast (dense) magmatic complex cylindrical body beneath the east rift zone (Figures 11b–12) which is not imaged in the travel time tomography (Figure 11a). This occurs in a well-defined volume where the model is expected to be correctly estimated (Figure 9). A fast (dense) magmatic cylindric body is the signature of the presence of magma conduits in which magma is cooling without

decompressing. The dense bodies may therefore outline the magma transport system of Kilauea volcano. An interesting feature of the double-difference tomographic results is the evidence of a change in direction of the magma transport system between 4 and 6 km. Magma coming from the depths below the caldera and southward, takes a N120° direction corresponding closely to the east rift zone root near the caldera, with an approximate 25°NW dip.

[48] At greater depths, the velocity model in the southern caldera–upper east rift zone region may be correctly reconstructed up to ~8.5 km depth (Figure 9). The results tend to indicate that the magma transport system is composed of two different feeding zones between 8.5 and 7 km depth (Figure 11b–12). Another striking feature is that the northern zone is approximately cylindrical from 8.5 to 7 km depth before joining the southern zone. It parallels the upper east rift zone and approximately dips 25°SE between 8.5 and 7 km depth. It turns to be subvertical between 6 and 4 km depth. The southern zone is approximately vertical. Both structures intersect just beneath the caldera between 6 and 7 km depth. Between 6 and 4 km depth, the most striking feature is the development of the upper east rift zone branch. The travel time tomography results (Figure 11a) do not display these features, although the images are not incompatible: travel time tomographic results appear as an "averaged" image of the double-difference tomographic results.

[49] There is little work directly comparable to these results in that small region of the Kilauea volcano. *Dawson et al.* [1999] used a 67-station network with an average 650-m station spacing within 5 km of the center of Kilauea caldera, during a 1-month experiment. They record 206 events of the caldera, most of them shallower than 3 km in depth. The 3-D velocity model was computable within a 100-km^2 square surface centered on the Kilauea caldera, up to an average depth of ~5 km, using $0.5 \times 0.5 \times 0.5$ km velocity cells. Results are published up to 2.5 km depth, with an announced resolution of 0.5 km. They show a 27 km^3, 10% low-velocity anomaly in the southeastern edge of the caldera, between the surface and 2.5 km depth. In that volume, *Thurber* [1984] detected a low-velocity anomaly. Our study shows high P wave velocities between 4.5 and 9 km depth beneath the caldera, although the contrast between the velocities beneath and around the caldera tends to disappear at 4 km depth. Studies by *Thurber* [1984] and *Okubo et al.* [1997] also found high P wave velocities at 5–8 km depth beneath the caldera. The lack of a dense station network near the surface in our study, the lack of deep earthquakes (>3 km), and the size of the experiment in the study by *Dawson et al.* [1999] explain that the two studies do not overlap, leaving a gap of about 2 km between the top of our results and the bottom of *Dawson et al.*'s [1999]. Both approaches (dense station network/dense earthquake array) are finally complementary and have the potential to reach the same resolution. Another conclusion that may be attempted is that *Dawson et al.*'s [1999] results image an active magma storage zone, where fluid magma occupies a significant part of the volume. In contrast, our results image a magma transfer zone, where the fast anomaly has been built by the progressive intrusion of nondecompressed magma, and where fluid magma only occupies a very limited volume. Figure 9 shows that a 20–
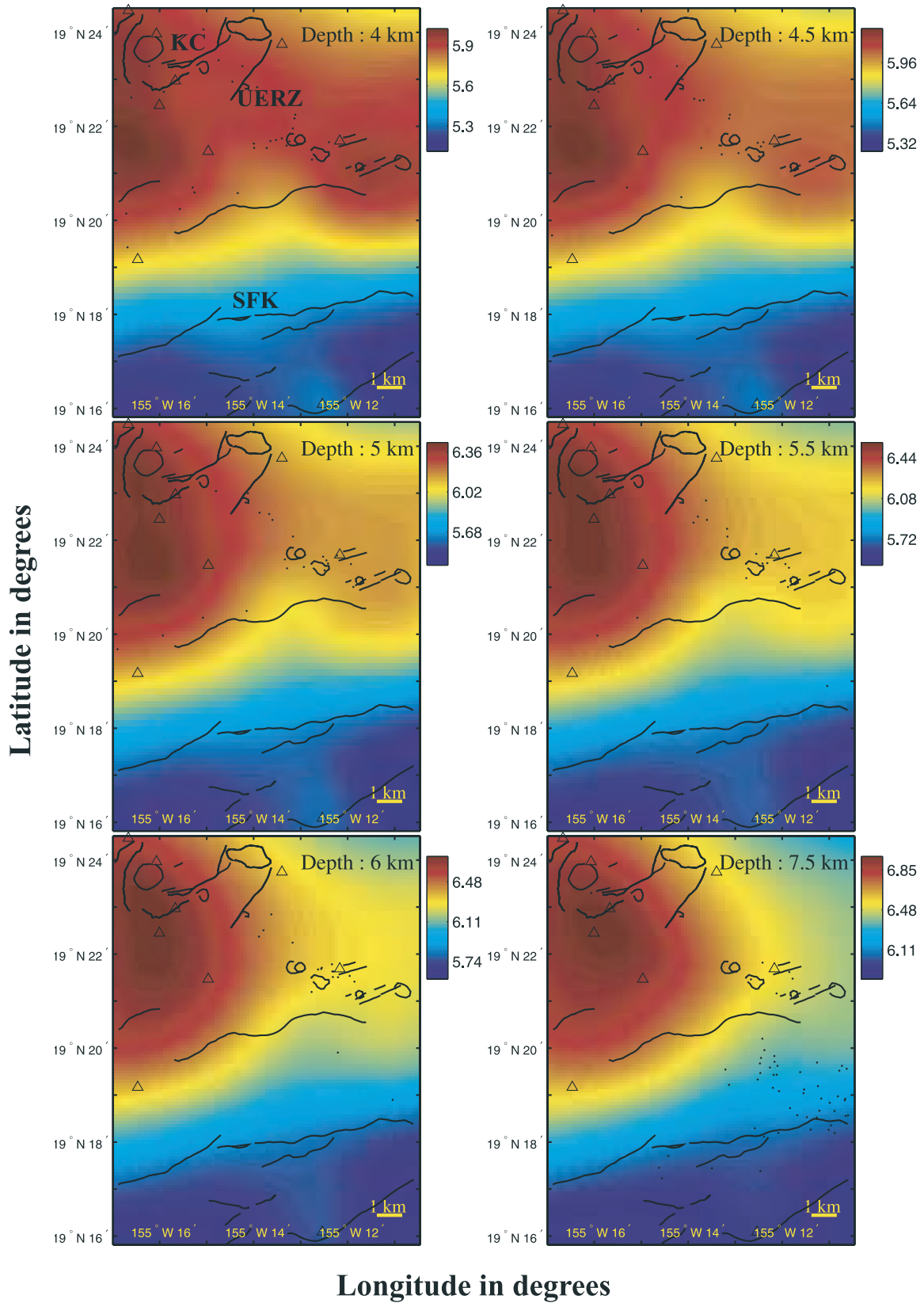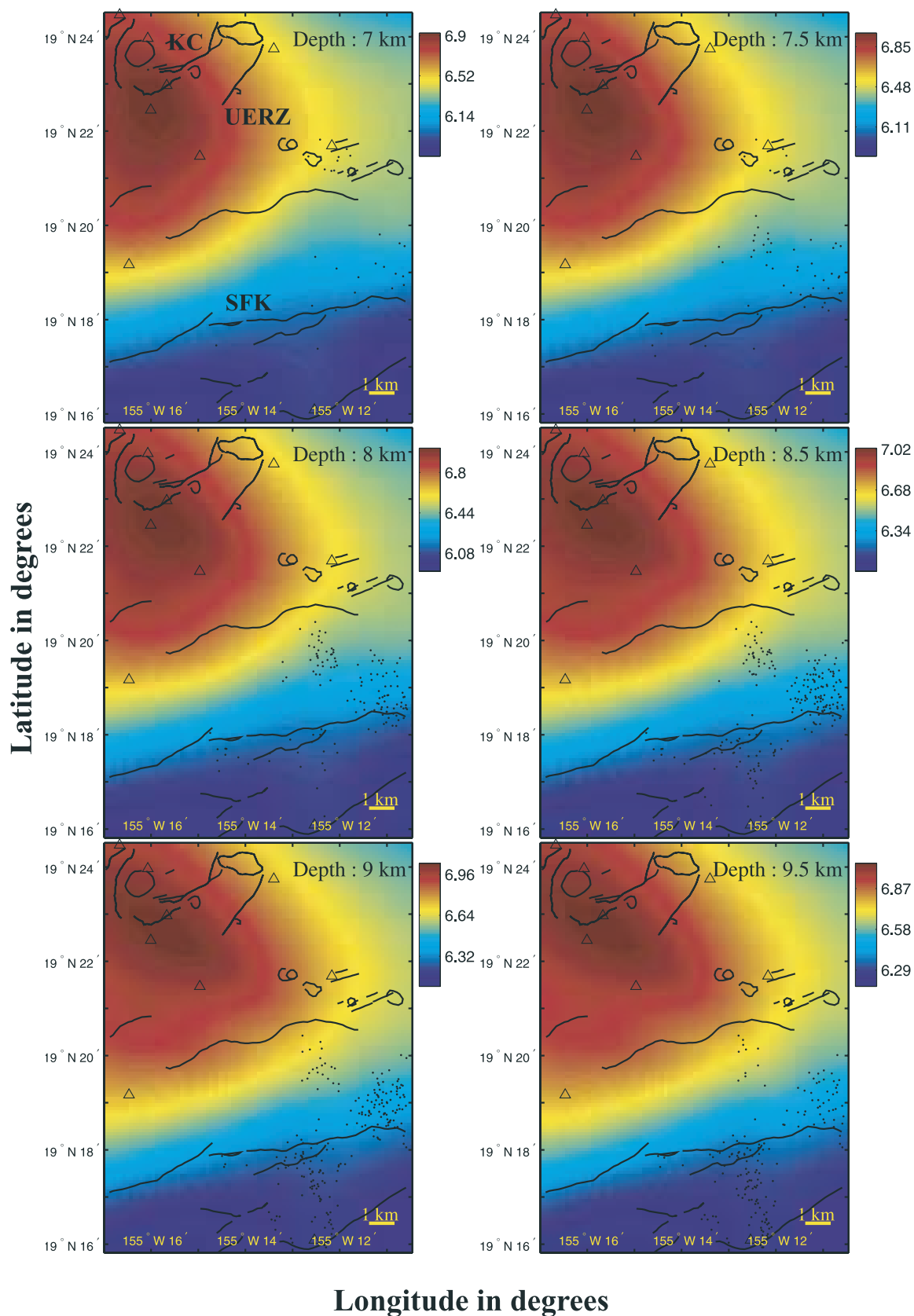
Figure 11a. Maps of the *P* wave seismic velocity in the south flank of Kilauea at various depths after local travel time tomography (λ = 1 km, σ$_v$ = 0.5 km/s). Geographical labels are in the upper left subplot. KC, Kilauea caldera; UERZ, upper east rift zone; KSF, Kilauea south flank.

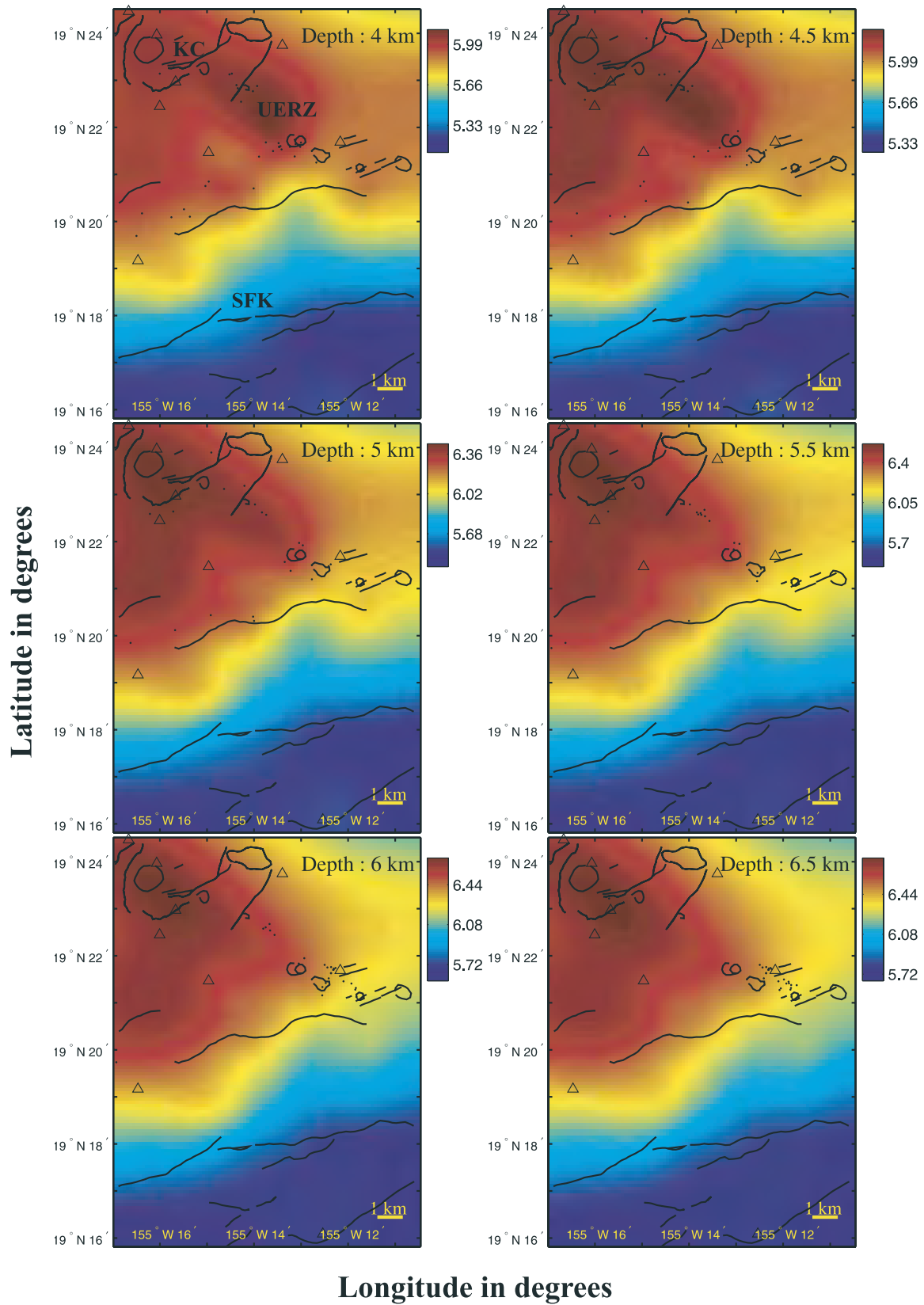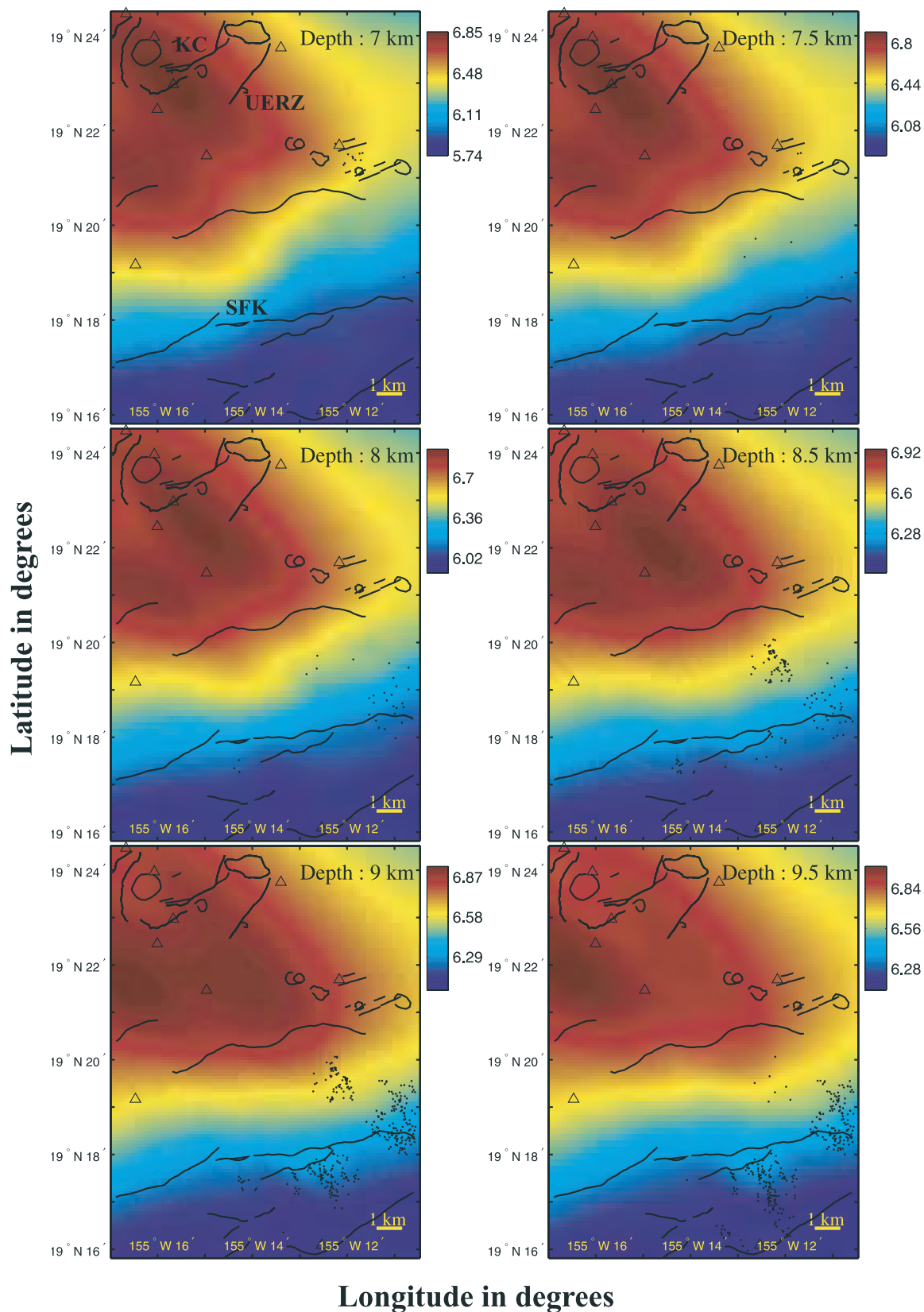**Figure 11a.** (continued)

Figure 11b. Maps of the *P* wave seismic velocity in the south flank of Kilauea at various depths after local double-difference tomography ($\lambda = 1$ km, $\sigma_v = 0.5$ km/s). Geographical labels are in the upper left subplot. KC, Kilauea caldera; UERZ, upper east rift zone; KSF, Kilauea south flank.
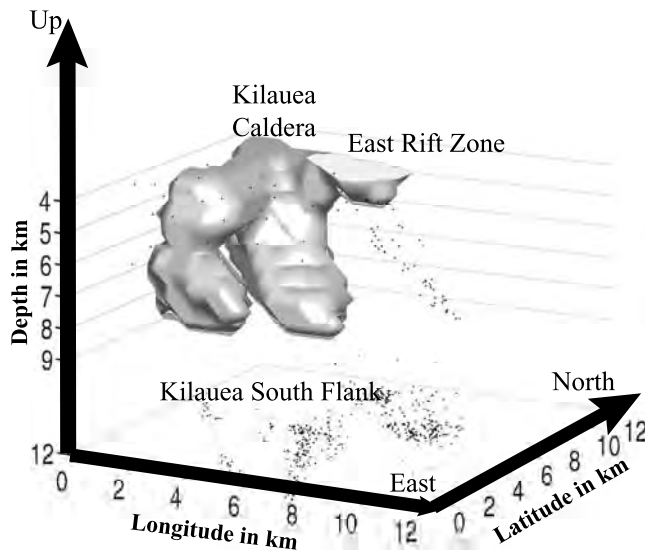
Figure 11b. (continued)

Figure 12. A 3-D representation of the high-velocity anomaly inferred from double-difference tomography in the southern caldera–upper east rift zone between 4 and 9 km depth. The volume is built from the 7% contour surfaces of the relative variation between the double-difference and the initial *Okubo et al.*'s [1997] velocity model. Black dots represent earthquakes. Distances are in km.

30 km^3 10% low-velocity anomaly would not remain undetected in the well-resolved volumes of our results.

5. Conclusion

[50]　In this study, we have explored the numerical methods in seismic tomography and set up appropriate tools, based on the optimization theory, for estimating as accurately as possible P wave velocities in a heterogeneous medium. We have shown that these tools allow the retrieval of a stable velocity model on a regional scale. This model is well correlated with the geological structures of the island of Hawaii. A double-difference tomography algorithm has been set up using the Tarantola-Valette approach. This study allows us to evidence conditions necessary for the correct use of double-difference tomography: (1) accurate independent time delay measurements, (2) a stable, accurate 3-D initial velocity model and travel time computation for estimating reliable residuals, (3) extended correlated event sets to limit trade-off between velocity and hypocentral parameters and to reach good geometric conditioning from data, (4) optimal a priori information to optimize the conditioning and to give a criterion which allows the choice of the final model. Double-difference tomography has been performed using accurate time delays computed by the cross-spectral method, from a well designed set of microearthquakes recorded by the USGS HVO permanent seismic network. In the volume where the medium is correctly estimated, travel time and double-difference tomographic results show significant differences. Double-difference tomographic results exhibit a detailed and coherent image of the magmatic system beneath Kilauea caldera and upper east rift. They show a complex, but coherent trajectory for the magma in the upper east rift zone.

References

Aki, K., and W. H. K. Lee (1976), Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model, *J. Geophys. Res.*, *81*, 4381–4399.

Benz, H. M., B. 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.

Benz, H. M., P. Okubo, and A. Villasenor (2003), Three-dimensional crustal P-wave imaging of Mauna Loa and Kilauea volcanoes, Hawaii, in *International Handbook of Earthquake Engineering and Seismology, Int. Geophys. Ser.*, vol. 81A, edited by W. H. K. Lee et al., pp. 407–421, Elsevier, New York.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola (1990), Robust elastic nonlinear waveform inversion: Application to real data, *Geophysics*, *55*, 527–538.

Crosson, R., and R. Koyanagi (1979), Seismic velocity structure below the island of Hawaii from local earthquake data, *J. Geophys. Res.*, *84*, 2331–2342.

Dawson, P. B., B. A. Chouet, P. G. Okubo, A. Villasenor, and H. M. Benz (1999), Three-dimensional velocity structure of the Kilauea caldera, Hawaii, *Geophys. Res. Lett.*, *26*, 2805–2808.

Eaton, J. P. (1962), Crustal structure and volcanism in Hawaii, in *The Crust of the Pacific Basin*, edited by G. A. Macdonald and H. Kuno, vol. 6, pp. 13–29, AGU, Washington, D. C.

Evjen, H. M. (1936), The place of the vertical gradient in gravitational interpretation, *Geophysics*, *1*, 127–136.

Geiger, L. (1910), Herdbestimming bei Erdbeben aus den Ankunftszeiten, *K. Ges. Wiss. Gött.*, *4*, 331–349.

Got, J., and P. Okubo (2003), New insights into Kilauea's volcano dynamics brought by large-scale relative relocation of microearthquakes, *J. Geophys. Res.*, *108*(B7), 2337, doi:10.1029/2002JB002060.

Got, J.-L., J. Fréchet, and F. W. Klein (1994), Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.*, *99*, 15,375–15,386.

Hansen, S., C. H. Thurber, M. Mandernach, F. Haslinger, and C. Doran (2004), Seismic velocity and attenuation in the East Rift Zone and South Flank of Kilauea volcano, Hawaii, *Bull. Seismol. Soc. Am.*, *94*, 1430–1440.

Haslinger, F., C. H. Thurber, M. Mandernach, and P. G. Okubo (2001), Tomographic imaging of the P-velocity structure beneath Kilauea's East Rift Zone and South Flank: Seismic evidence for the presence of a deep magma body, *Geophys. Res. Lett.*, *28*, 375–378.

Hill, D. (1969), Crustal structure of the island of Hawaii, from seismic refraction measurements, *Bull. Seismol. Soc. Am.*, *59*, 101–130.

Ito, A. (1985), High resolution relative hypocenters of similar earthquakes by cross-spectral analysis method, *J. Phys. Earth.*, *33*, 279–294.

Jordan, T. H., and K. A. Sverdrup (1981), Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seismol. Soc. Am.*, *71*, 1105–1130.

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

Klein, F. W. (1981), A linear gradient crustal model for south Hawaii, *Bull. Seismol. Soc. Am.*, *71*, 1503–1510.

Latorre, D., J. Virieux, T. Monfret, V. Monteiller, T. Vanorio, J.-L. Got, and H. Lyon-Caen (2004), A new seismic tomography of Aigion area (Gulf of Corinth, Greece) from the 1991 data set, *Geophys. J. Int.*, *159*, doi:10.1111/j.1365-246X.2004.02412.x.

Le Meur, H., J. Virieux, and P. Podvin (1997), Seismic tomography of the Gulf of Corinth: A comparison of methods, *Ann. Geophys.*, *40*, 1–25.

Menke, W., and D. Schaff (2004), Absolute earthquake locations with differential data, *Bull. Seismol. Soc. Am.*, *94*, 2254–2264.

Michelini, A., and A. Lomax (2004), The effect of velocity structure errors on double-difference earthquake location, *Geophys. Res. Lett.*, *31*, L09602, doi:10.1029/2004GL019682.

Monteiller, V., J.-L. Got, J. Virieux, and P. G. Okubo (2003), Double-difference tomography of Kilauea volcano, Hawaii, paper presented at EGS-AGU-EUG Joint Assembly, Nice, France, April.

Mosteller, F., and J. W. Tukey (1977), *Data Analysis and Regression*, Addison-Wesley, Boston, Mass.

Okubo, P., H. M. Benz, and B. A. Chouet (1997), Imaging the crustal magma sources beneath Mauna Loa and Kilauea volcanoes, Hawaii, *Geology*, *25*, 867–870.

Paige, C. C., and M. A. Saunders (1982), LSQR: Sparse linear equations and least squares problems, *Trans. Math. Software*, *8*, 195–209.

Podvin, P., and I. Lecomte (1991), Finite difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools, *Geophys. J. Int.*, *105*, 271–295.

Poupinet, G., W. L. Ellsworth, and J. Fréchet (1984), Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California, *J. Geophys. Res.*, *89*, 5719–5731.

Rowan, L. R., and R. W. Clayton (1993), The three-dimensional structure of Kilauea volcano, Hawaii, from travel time tomography, *J. Geophys. Res.*, *98*, 4355–4375.

Rubin, A. M., D. Gillard, and J.-L. Got (1999), Streaks of microearthquakes along creeping faults, *Nature*, *400*, 635–641.

Ryall, A., and D. L. Bennett (1968), Crustal structure of southern Hawaii related to volcanic processes in the upper mantle, *J. Geophys. Res.*, *73*, 4561–4582.

Shearer, P. M. (1997), Improving local earthquake locations using the L1 norm and waveform cross-correlation: Application to the Whittier Narrows, California, aftershock sequence, *J. Geophys. Res.*, *102*, 8269–8283.

Slunga, R., S. Rögnvaldsson, and R. Bödvarsson (1995), Absolute and relative locations of similar events with application to microearthquakes in southern Iceland, *Geophys. J. Int.*, *123*, 409–419.

Tarantola, A. (1987), *Inverse Problem Theory*, 613 pp., Elsevier, New York.

Tarantola, A., and B. Valette (1982), Generalized nonlinear inverse problems solved using the least-squares criterion, *Rev. Geophys.*, *20*, 219–232.

Thurber, C. H. (1984), Seismic detection of the summit magma complex of Kilauea volcano, Hawaii, *Science*, *223*, 165–167.

Thurber, C., S. Roecker, H. Zhang, S. Baher, and W. Ellsworth (2004), Fine-scale structure of the San Andreas fault zone and location of the SAFOD target earthquakes, *Geophys. Res. Lett.*, *31*, L12S02, doi:10.1029/2003GL019398.

Vidale, J. E. (1988), Finite-difference calculation of traveltimes, *Bull. Seismol. Soc. Am.*, *78*, 2062–2076.

Waldhauser, F., and W. L. Ellsworth (2000), A double difference earthquake location algorithm: Method and application to the northern Hayward fault, CA, *Bull. Seismol. Soc. Am.*, *90*, 1353–1368.

Ward, P. L., and S. Gregersen (1973), Comparison of earthquake locations determined with data from a network of stations and small tripartite arrays on Kilauea volcano, Hawaii, *Bull. Seismol. Soc. Am.*, *63*, 679–711.

Wolfe, C. J. (2002), On the mathematics of using difference operators to relocate earthquakes, *Bull. Seismol. Soc. Am.*, *92*, 2879–2892.

Zhang, H., and C. H. Thurber (2003), Double difference tomography: The method and its application to the Hayward fault, California, *Bull. Seismol. Soc. Am.*, *93*, 1875–1889.

Zhang, H., C. H. Thurber, D. Shelly, S. Ide, G. C. Beroza, and A. Hasegawa (2004), High-resolution subducting slab structure beneath northern Honshu, Japan, revealed by double-difference tomography, *Geology*, *32*, 361–364, doi:10.1130/G20261.

————————————

J.-L. Got and V. Monteiller, Laboratoire de Géophysique Interne et Tectonophysique, Université de Savoie, F-73376 Le Bourget-du-Lac, France. (got@univ-savoie.fr)

P. Okubo, U.S. Geological Survey, Hawaiian Volcano Observatory, P.O. Box 51, Hawaii National Park, HI 97560, USA.

J. Virieux, Géosciences Azur, Sophia-Antipolis, F-06560 Valbonne, France.