JI-3D
A new approach to high resolution
regional seismic tomography:
Theory and applications

Dissertation
zur Erlangung des Doktorgrades der
mathematisch-naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Michael Jordan
aus Miltenberg

Göttingen 2003

Ferne Kunde
bringt Dir der schwankende Fels -
Deute die Zeichen!

Emil Wiechert, 1902

Contents

1 Introduction

Seismic waves provide the best method to investigate the Earth's deep interior. Deduced from medical applications, seismic tomography was first introduced by (Aki and Lee, 1976) to determine the 3D variations of velocity and attenuation properties inside the Earth from traveltimes and amplitudes of seismic waves. Regional (teleseismic) tomography can be used to image ongoing geodynamical processes like the subduction of tectonic plates (slabs) or the rise of mantle plumes. Since slabs are usually colder and denser than their surroundings, they appear as regions of increased seismic wavespeed in the tomographic models. Plumes, which are considered as localized upwelling of hot, less dense and therefore buoyant mantle material, are imaged as regions of reduced seismic velocity.

The objective of this thesis is the development of a new method (JI-3D) of joint inversion of teleseismic delaytimes and Bouguer gravity data to derive high resolution tomographic images. The resulting models consist of 3D velocity and density variations, relative to a 1D standard Earth (Kennett et al. (1995)). For the optimization of the inversion, it is possible to incorporate additional information about known subsurface structure as *a priori* information via a Bayes algorithm (Zeyen and Achauer (1997)). Moreover, the method is applied to image a mantle plume beneath the Eifel volcanic fields, Germany. Since thermal plumes interact with the phase changes which occur at the major discontinuities in the Earth's mantle, the corresponding effect on the results of a regional seismic traveltime inversion is examined.

The composition of this thesis is as follows: Chapter 2 consists of two parts, first the new method, called JI-3D is described in detail, including the principles of the joint inversion, the algorithm, the 3D raytracing and the variable parameterization, which is especially focused on. Second, the applications to synthetic data sets are presented, ranging from the non iterative delaytime inversion to the joint inversion of the synthetic delaytime and Bouguer gravity data set. In addition to that, also the use of additional constraints, that further reduce the space of possible solutions is explained. The synthetic data sets are calculated with respect to a synthetic structure that is designed to reveal the limits of this method. Chapter 3 presents the application of the JI-3D method to real data to examine the deep source of the Eifel volcanism in western Germany. This chapter is practically identical to Ritter et al. (2001) in which M. Jordan is the second author. The contributions of M.J. to this paper are data processing and the full inversion part, including parameterization, inversion of the Eifel data set, and the reconstruction tests, using the JI-3D method as described in chapter 2. Since this work was done using an early version of JI-3D, only the non iterative delaytime inversion was applied to the data. The resulting model of velocity perturbations suggest a mantle plume beneath the Eifel, reaching down to at least 410 km depth. Chapter 4 examines the influence of deflected seismic mantle discontinuities on teleseismic tomography. Due to their positive thermal anomaly, mantle plumes can interact with temperature- and pressure-sensitive phase boundaries (e.g at 410 and 660 km depth), causing an up- or downwelling. Since these so called deflections involve lateral changes of the seismic velocities, the effects on the seismic delaytime data and the corresponding tomographic inversion results are examined. This is done for a realistic experimen-

tal setting, using synthetic three-dimensional tomographic modeling. This chapter is identical to Jordan and Ritter, (in prep.). Finally, the results of this thesis are summarized in chapter 5.

2 Joint inversion

2.1   Introduction

Seismic tomography is one of the most important means to image the internal structure of the Earth. Restricted array teleseismic tomography determines the 3D-distribution of velocity perturbations relative to a given background model on a regional scale, even down into the lower mantle (> 660 km depth). This enables the imaging of ongoing tectonic and geodynamic processes e.g. at subduction zones, hot spots and active volcanoes. The resulting models allow estimations of variations of the density, the elastic constants and the temperature. Hence 3D-seismic tomography provides also input models to other geophysical disciplines (e.g. geodynamics and gravity). Since the first seismic tomographic scheme was developed in 1974 (Aki et al., 1976,1977), the fundamental steps have not changed: Parameterization of an initial velocity model, forward calculation of the delaytimes through the model for a given source-receiver configuration, building of a system of equations, and solving the equations by performing a matrix inversion, resulting in changes to the initial velocity parameters. In doing so, the parameterization (i.e. number, size and position of blocks or number and position of nodes, respectively), has always been decisive for the mathematical and spacial resolution and the stability of the inversion.

The ACH-code (Aki et al., 1977) has been further improved. E.g. Evans and Zucca (1988) presented an updated ACH inversion, using a so called offset and average technique, which causes the inversion result to be more independent of the initial, arbitrary model parameterization. Several inversions are calculated, according to shifted parameterizations and the inversion results are averaged. This leads to an improved spacial resolution, a more stable inversion, since larger blocks can be used, and a result more independent from subjective assumptions. Weiland et al. (1995) introduced a 3D-raytracer (Steck and Prothero, 1991) together with an iterative inversion technique to cope with the nonlinearity caused by the interdependence of velocity model and ray paths. Due to the more realistic raypaths and the possibility of modeling diffractions this method was able to image strongly contrasting low velocity structures reliably, which had been underestimated massively before. Since their raytracer requires the inversion grid to be regular, many nodes remain under- or even non constrained, leaving the inversion problem ill posed. Therefore, Weiland et al. (1995) require their models to be smooth, and the inversion is stabilized by increased damping of the unconstrained model parameters. Later Zeyen and Achauer (1997) presented an algorithm for the joint inversion of teleseismic delaytimes and gravimetric data. Using a Bayesian approach (Jackson and Matsu'ura, 1985) they were able to incorporate both, internal and external a priori information in the inversion. This can be the knowledge of certain model parameters (e.g. from previous studies) or the correlation of the model parameters. For the inversion quasi linear correlated velocity and density parameters are assumed, which may vary in the model space, especially with depth. Therefore, the correlation factor is treated as an additional parameter which also is inverted for. Due to the strong nonlinearity of the inversion problem, the inversion is done iteratively. However, raytracing is performed by a 1D-raytracer prior to the inversion to create a

Frechet matrix which is used as an onetime input in the iterative inversion scheme. Therefore, the inversion result strongly depends on the starting velocity model.

Velocity and density models are parameterized using identical block structures. In this chapter, a method is presented, using the same Bayesian approach as Zeyen and Achauer (1997) to perform a fully iterative joint inversion. However important improvements are introduced which lead to maximum stability and maximum spacial resolution at the same time. This is achieved by avoiding non- or under constrained model parameters by a variable parameterization which is completely adapted to the information density in the model space. Moreover, a 3D-raytracer (Steck and Prothero, 1991) is implemented to obtain realistic raypaths. In the next sections, first the method is described in detail, then the synthetic data set and finally the various inversion results are presented, demonstrating the effects and benefits of this inversion method. The method is called 'JI-3D' (Joint Inversion in 3 Dimensions).

2.2   Method

2.2.1   Principles of joint inversion

One and the same subsurface structure may have different physical properties, causing independent geophysical responses at the Earth's surface. In case the properties are correlated, a joint inversion can mitigate the restrictions inherent to the single data sets. Moreover, the number of independent observations is increased. After the inversion the resulting model should be able to explain both data sets leading to a reduced space of possible solutions and a more unique and stable model. Here, teleseismic delaytime and gravity data are used to obtain models of velocity and density perturbations with respect to the respective background model. This combination is advantageous especially for regional studies, since the data complement each other on this scale. While teleseismic delaytime data provide good horizontal resolution, their inversion is often affected by vertical smearing and insufficient near surface resolution due to the lack of crossing rays. This ideally is compensated by the good near surface resolution of the gravity signal. The correlation between density ρ and velocity v is assumed to be linear (Birch, 1961). This relationship $\rho = A + Bv$ (A and B are empirical constants) is pressure dependent and only valid if temperature and mean atomic weight are constant. Since we only consider variations of velocities and densities, the relation we use reads $\delta\rho = B\delta v$ (Lees and VanDecar, 1991).

However, most structures that are examined using teleseismic tomography, imply a lateral change in temperature or composition or both (e.g. plumes, rifts). Hence, the conditions, provided above, do not apply here. To account for the depth and temperature dependence of the correlation factor we treat B as an additional parameter which we also invert for in the joint inversion. One B-value is determined per layer, comprising the average of the local correlations.

2.2.2   Derivation of the algorithm

Regional teleseismic delaytime inversion

Data   Regional delaytime tomography uses arrival times of teleseismic earthquakes recorded at a 2D-array of seismic stations (figure 2.1). The aim is to determine the 3D velocity structure beneath the station array in terms of deviations relative to an 1D standard Earth model (e.g. IASP91) (Kennett and Engdahl, 1991). As suggested in figure 2.1, seismic rays are employed to track the displacement from source to receiver through the Earth.
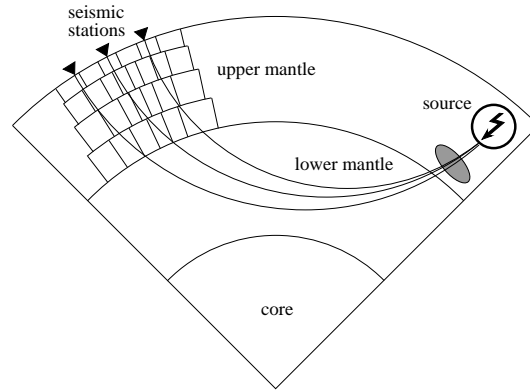

Figure 2.1: *Schematic setting of a teleseismic delaytime inversion consisting of an array of seismic stations, seismic sources with epicentral distances larger than 30° and the target volume (indicated by blocks). Note, that upper and lower mantle and core are not at scale.*

Since we are merely interested in the velocity perturbations inside the target volume, it is important that the traveltimes of the rays are not affected by heterogeneities outside. Hence, the aperture of the array (usually some hundred kilometers) has to be small compared to the epicentral distance ($\Delta > 25°$). Rays enter the target volume at the bottom, mostly traveling through the lower mantle and the core, which are much more homogeneous than the upper mantle and the crust. Therefore, the expected distortion of the travel times by heterogeneities outside the target volume is small. Moreover, the effects of source-side, small scale structure are removed from the data by wavefront healing due to the long distance to the target volume.

The traveltime (t) of ray i is expressed by the path integral 2.1 and depends on both, the velocity distribution ($v(x, y, z)$) and the raypath (S).

$$t_i = \int_S \frac{1}{v(x, y, z)} ds \tag{2.1}$$

Since S and v are interdependent, the inverse problem, to determine the velocity structure from the recorded traveltime t_i is nonlinear.

Since one is interested in deviations from the standard Earth model, the theoretical traveltimes for a 1D standard Earth model are removed from the observed ones, resulting in so called travel time residuals.

$$t_{res,i} = t_i - t_{0,i} = \int_S \frac{1}{v(x, y, z)} ds - \int_{S_0} \frac{1}{v_0(x, y, z)} ds \tag{2.2}$$

S_0 and $v_0(x, y, z)$ are the background velocity model and the corresponding raypath. The traveltime residuals often are affected by large errors (up to several seconds), caused by the misslocation of source, source time errors and travel time effects by large scale heterogeneities along the ray paths (see figure 2.1). As all rays of the same event are affected in the same way, these errors can be eliminated easily by forming relative travel time residuals, also simply referred to as relative residuals.

$$t_{res,i,rel} = t_{res,i} - t_{ref} \qquad (2.3)$$

For each event j, a reference residual traveltime t_{ref} is subtracted from the residuals of all stations i. As reference residual often the mean of all residuals i, determined for event j is used. Alternatively also single stations as well as clusters of single stations may be considered as references (Barth, 2002).

It is assumed, that the relative residual only comprises the effect of velocity perturbations inside the target volume.

Inversion   The relative traveltime residuals contain the effect of the 3D-deviations from the 1D background velocity model in the target volume integrated along the raypaths.
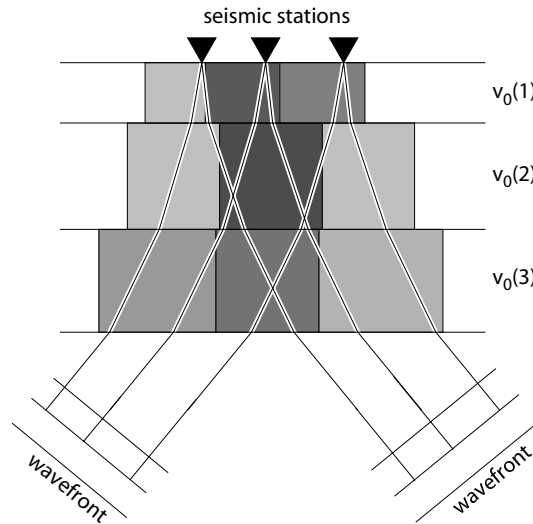

Figure 2.2: *Sample velocity model with background velocities $v_0(n)$, model parameters given according to blocks. The data have to be projected into the model volume along the seismic rays from the seismic stations to the respective wavefront, resulting in perturbations of the initial background model (indicated by the different shading of the blocks).*

To determine the velocity structure, the target volume has to parameterized. This means, that it is divided into layers with corresponding background velocities (see figure 2.2). The layers are subdivided into blocks , which correspond to the velocity model parameters, i.e. changes to the background velocity model. The entity of all velocity model parameters comprise the velocity model.

Supposed, the data and model parameters are represented by vectors $\vec{d}$ and $\vec{m}$ respectively, then the basic inversion problem reads

$$\vec{d} = G\vec{m} + \vec{\varepsilon}. \qquad (2.4)$$

G is the so called Frechet matrix which relates observations $\vec{d}$ and model $\vec{m}$. G_{ij} contains the partial derivative of the i-th data with respect to the j-th model parameter. $\vec{\varepsilon}$ is a vector containing the data errors.

Since usually there are merely estimates of the error vector and the model vector is not known at all, we assume a hypothetical model vector $\vec{m}_h$ and the corresponding data set is calculated forward from $G \cdot \vec{m}_h$.

$$\vec{e} = \vec{d} - G\vec{m}_h \qquad (2.5)$$

In case the hypothetical model vector is the true one, the difference between the observed and forward calculated data, $\vec{e}$, is equal to the data error vector $\vec{\varepsilon}$. This means, that in order to find the true model vector the difference between observed and forward calculated data has to be minimized.

Since the Frechet matrix strongly depends on the raypaths, true raypaths are vital for the inversion. Incorrect raypaths lead to wrong models, contradictions between model and data and instability.

Equation 2.4 represents a set of equations that can be fully solved only if there are sufficient linearly independent equations. Otherwise, the model parameters cannot be resolved independently. Hence, in contrast to current practice (e.g. Aki et al., Spakman and Bijwaard, Zeyen and Achauer, Tiberi et al.) who completely rely on the number of rays per block, the total number of passing rays is meaningless for the resolution of a velocity model parameter.
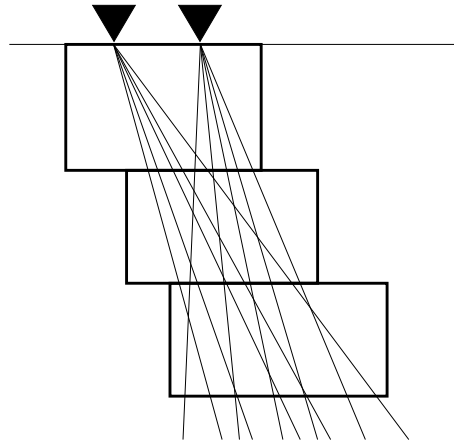

Figure 2.3: *Minimum raynumber or raydensity approaches do not necessarily lead to stable, well resolved inversion results. This example shows that even though ray crossings exist, the three blocks merely can be resolved commonly, resulting in an average velocity perturbation for all three blocks.*

As an example, figure 2.3 illustrates a ray bundle, with even crossing rays, that passes the same three blocks. Hence, the corresponding set of equations is completely linearly dependent and one can merely determine an average of the model parameters of the three blocks. Consequently, for an inversion this ray bundle is of no more use than one single ray in terms of resolution.

It is obvious, that the independent resolution of the velocity model parameters and the stability and uniqueness of the inversion problem strongly depends on the realistic raypaths and an optimized parameterization.

Bouguer gravity inversion

Data According to Newton, the force of gravitation between two masses m_1 and m_2 is

$$\vec{F} = \gamma \cdot \left(\frac{m_1 m_2}{r^2}\right) \hat{r_{12}}. \tag{2.6}$$

It is direct proportional to the product of m_1 and m_2 and inversely proportional to the square of their distance r. The direction of the force $\vec{F}$ is given by the unit vector $\hat{r_{12}}$, directed from m_1 towards m_2. The universal gravitational constant γ is approximately $6.672 \cdot 10^{-11} \frac{m^3}{kgs^2}$.

The corresponding acceleration $\vec{a}$ is the gravitational force per unit mass (Menke and Abbott, 1990).

$$\vec{a} = \gamma \cdot \left(\frac{m_1}{r^2}\right) \hat{r_1}. \tag{2.7}$$

In this case, the unit vector $\hat{r_1}$ is directed from the sample (unit) mass towards m_1. If we consider the Earth, m_1 will be the Earth's mass, r will be the distance between the center of the Earth and the observational point and $\hat{r_1}$ will point towards the center of the Earth.

To obtain the gravity vector $\vec{g}$, we also have to consider the centrifugal acceleration $\vec{c}$.

$$\vec{g} = \vec{a} + \vec{c} \tag{2.8}$$

The unit of gravity is Gal after Galileo Galilei ($1\text{Gal} = 1\frac{cm}{s^2}$).
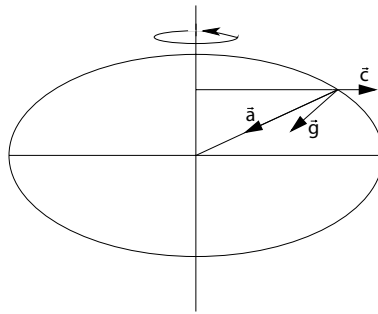Since gravity depends on the latitude, values range from 978.0 Gal at the equator and 983.3 Gal at the poles.

Figure 2.4: *The gravity vector $\vec{g}$ is composed of the vector of gravitational and centrifugal acceleration $\vec{a}$ and $\vec{c}$, respectively (after Berckhemer, 1990).*

Both, gravitational acceleration and gravity itself can be expressed as potential fields.
The gravity signal contains the subsurface mass and density distribution in an integrative manner. Due to the rotation, the shape of the earth, Earth tides and topography, the observed gravity values are not directly comparable, so that the observed gravity has to be reduced prior to interpretation. The observed gravity is reduced to the so called Bouguer anomaly g_{boug} (see equation 2.9) which merely contains the gravity effects of mass inhomogeneities beneath a reference level.

$$g_{boug} = g_{obs} - \gamma_0 - \delta g_{free} - \delta g_{topo} + \delta g_{boug} \tag{2.9}$$

As illustrated in figure 2.5 gravity is observed at some point P at the surface. In a first step the theoretical gravity γ_0 is subtracted from the measured gravity g_{obs}. γ_0 is the gravity that was

observed at point P for a standard Earth model. This correction is latitude dependent and also involves Earth tides. The resulting gravity signal contains the heterogeneous subsurface structure, the effects of elevation and topography.

To correct the elevation effect δg_{free}, the observational point P is projected onto the reference level and the gravity data are corrected due to the mere height difference. The space between point P and the reference level is regarded massless. The resulting data set is called free air data. In a next step, the effect of the mass of the so called Bouguer plateau which was neglected in the previous step is removed δg_{boug}. To do so, an average density of usually $2.67 \frac{kg}{dm^3}$ is assumed. The resulting data set is called Bouguer data.

If strong topography is present, an additional topographic correction δg_{topo} is applied prior to the Bouguer correction. Since the effects of mass excess (mountains) or mass deficit (valleys) also bias the data, the corresponding signals also have to be removed from the data. This requires an exact knowledge of the topography of the surface.
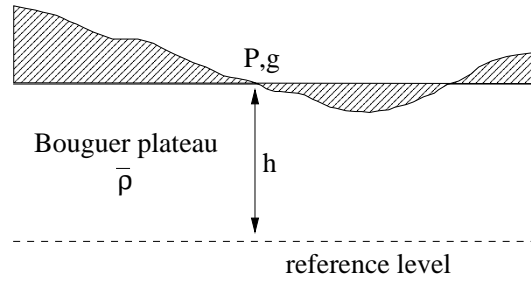

Figure 2.5: *Gravity $\vec{g}$ is observed at point P. To make gravity data comparable, corrections are applied for location, elevation and topography. This results in Bouguer gravity data given at a reference level.*

Inversion   The Bouguer gravity data contain the effects of all mass inhomogeneities beneath a reference level in an integrative manner.

To determine a 3D-model of density variations, the proceedings are analogous to the ones applied to determine the velocity model in case of the teleseismic delaytimes. For the formal inversion the relation between the vectors of Bouguer gravity $\vec{g}_{boug}$ and density variations $\delta \vec{\rho}$ is

$$\vec{g}_{boug} = \mathbf{G} \cdot \delta \vec{\rho} + \vec{\epsilon}. \tag{2.10}$$

The elements G_{ij} of matrix $\mathbf{G}$ are the partial derivatives of the data i with respect to the model parameter j and $\vec{\epsilon}$ is the error vector. Similar to the delay time case, the model space is parameterized by dividing it into layers which, in turn, are subdivided into blocks. Starting values for the density perturbations are attributed to the blocks which correspond to the density model parameters. The Bouguer gravity response to the starting model is calculated according to Zeyen and Pous (1993). The gravimetric anomaly of N blocks with density ρ_n (n=1,..,N) and limits $x_{1n}, x_{2n}, y_{1n}, y_{2n}, z_{1n}, z_{2n}$ at an observational point r_i is given by

$$g_i(r_i) = \sum_{n=1}^{N} \{ -\gamma \rho_n \sum_{j=1}^{2} \sum_{k=1}^{2} \sum_{l=1}^{2} s_{jkl} \cdot \left[\Delta x_{jni} ln(\Delta y_{kni} + R) \right. \tag{2.11}$$

$$\left. + \Delta y_{kni} ln(\Delta x_{jni} + R) + \Delta z_{lni} \arctan \frac{\Delta z_{lni} R}{\Delta x_{jni} \Delta y_{kni}} \right] \}.$$

Where γ again is the gravitational constant, $s_{jkl} = s_j s_{ksl}$, with $s_1 = -1$ and $s_2 = +1$ and $R = \sqrt{\Delta x_{jni}^2 + \Delta y_{kni}^2 + \Delta z_{lni}^2}$.

Since the resulting 3D-distribution of density contrasts should be able to explain the data, the inversion tries to minimize the difference between observed and forward calculated data.
In contrast to the teleseismic delaytime case, the inversion is a linear problem, and the elements of the Frechet matrix G_{ij} can be calculated analytically: $\frac{\delta g_i}{\delta \rho_j} = \frac{g_i}{\rho_j}$ (Zeyen and Pous, 1993). Here, i is the index of the data point and j is the index of the density model parameter.

However, due to the principle of equivalence no unique model of density heterogeneities can be derived from Bouguer gravity data (Menke and Abbott, 1990). This is illustrated in figure 2.6, showing several density bodies causing exactly the same gravity anomaly.
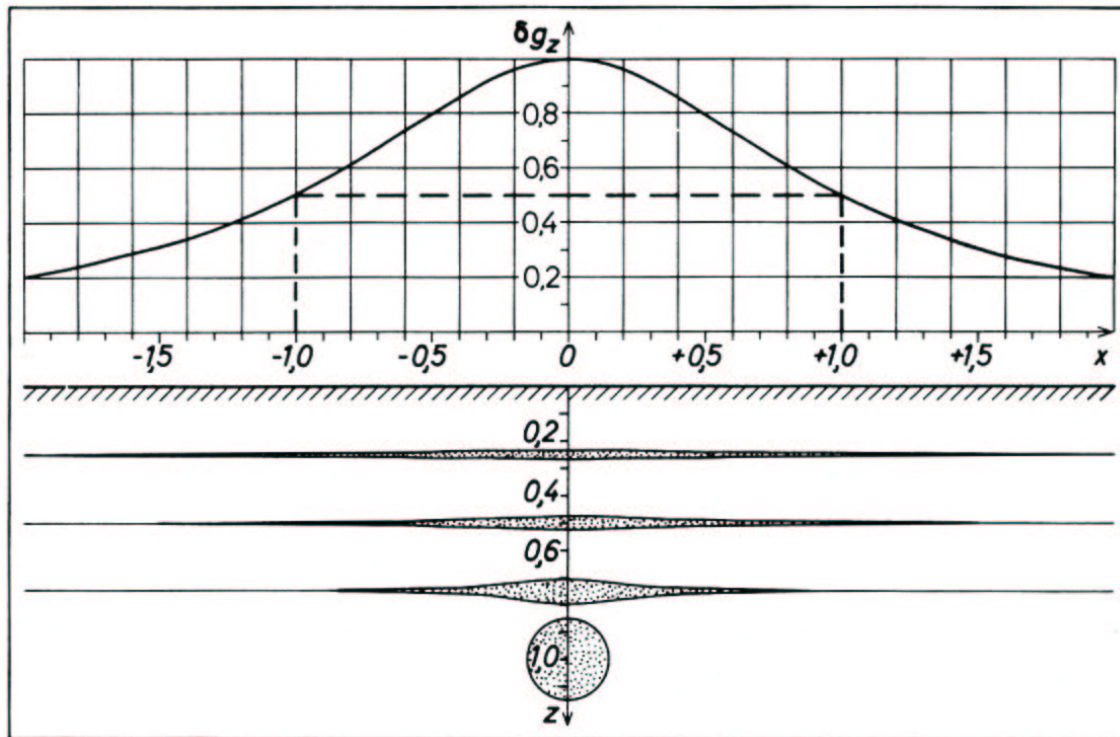

Figure 2.6: *Gravity anomaly above an infinite cylindric body. In the top section the gravity is given in relative units. The dashed lines mark the halfwidth of the gravity signal. In the lower section equivalent bodies are shown, all causing the same gravity signal, shown at the top. Note that the relative densities of the bodies are given as annotations (0.2 to 1.0) (taken from Kertz, 1995)*
.

The maximum depths of the anomaly's center of mass can be estimated from the halfwidth of the gravity signal by assuming that the possible density contrasts inside the Earth are limited. According to Jung (1960) the maximum depth of the center of mass is approximately 2/3 of the halfwidth of the corresponding gravity signal.
In contrast, it is not possible to derive a minimum depth from the gravity signal. Generally *a priori* information is applied to restrain the results to reasonable models. The information can

be the surface structure, geology, the assumption of reasonable density contrasts and constraints from other geophysical methods.

To obtain reasonable results unbiased by the strong distance dependence of the gravity signal the influence of the density blocks on the observable result should be constant for all blocks. This means that the size of the blocks has to increase with distance from the points of observation.

Velocity-Density relation

Since teleseismic delaytimes and Bouguer gravity data complement each other in terms of resolution power in the shallow and deep regions of the model space, the two data types are inverted simultaneously. To do so, a relation between density and velocity model parameters is required. From laboratory measurements of seismic velocities and densities, Birch (1961) concluded that a linear velocity-density relation exists, provided that temperature and mean atomic weight are constant.

$$\frac{dv}{d\rho} = B \qquad B \approx 3\frac{\frac{km}{s}}{\frac{g}{cm^3}} \qquad\qquad (2.12)$$

Formula 2.12 only considers velocity- and density contrasts, dv and $d\rho$, respectively. B is a constant which is valid for all materials if the mean atomic weight does not change. Usually these conditions are not valid since most teleseismic tomography studies focus on thermally or compositionally heterogeneous structures (e.g. plumes, magma chambers, subducting slabs). Hence, B is used as mainly depths dependent mean of the velocity-density relations of the individual parameters and one mean B-value is determined for each layer.

The algorithm

In both, single and joint inversions of delaytime and gravity data sets, the result is not unique and the inversion problem may be ill posed and unstable, especially in the case of delaytime inversion and joint inversion. Since nonuniqueness and instability both are caused by over- and underdetermined parameters, it is straightforward to directly cope with these conditions. Overdetermination usually is dealt with a least squares approach (Menke, 1989) whereas underdetermination here is approached by two different methods. The first one relieves underdetermination by using additional constraints and to incorporate these as *a priori* information into the inversion (e.g. Tarantola and Valette, 1982). The resulting models have to be consistent with these constraints. The second approach is to change the formulation of the inversion problem, avoiding underdetermined model parameters. This is done by the variable parameterization, presented later in this chapter.

Both data and model parameters are treated in the same way in this approach (Tarantola and Valette (1982)). The *a priori* information consists of measured data, initial estimates of model parameters or their relations and the respective reliabilities. This means, that data with a small observational error have a stronger influence on the inversion result than those with large uncertainties. Analogously, model parameters, which are well known, serve as constraints, the rest of the inversion result has to be compatible with. For the non specific *a priori* information realistic values for the expected velocity-density relation and the degree of linearity are assumed

or reasonable smoothing of the density model is demanded.

The derivation of the algorithm starts at the linear parametric inverse problem:

$$\vec{y} = A\vec{x} + \vec{\varepsilon} \tag{2.13}$$

$\vec{y}$ is a n-vector of observed data, $\vec{x}$ is a m-vector of unknown model parameters and A is an $n \times m$ coefficient matrix relating data and model parameters. $\vec{\varepsilon}$ is a n-vector of random errors with a known probability density function (pdf).

Assuming a hypothetical parameter vector $\vec{\chi}$, a n-vector of residuals $\vec{e}$ can be calculated

$$\vec{e} = \vec{y} - A\vec{\chi}. \tag{2.14}$$

If $\vec{\chi}$ equals the 'true' parameter vector $\vec{x}$, then the residual vector $\vec{e}$ also equals $\vec{\varepsilon}$. Consequently,

$$p(\frac{\vec{x}}{\vec{y}}) = p(\vec{e}) = p(\vec{\varepsilon}) \tag{2.15}$$

which means that the pdf of the parameters $\vec{x}$ for given data $\vec{y}$ equals the pdf of the random errors $\vec{\varepsilon}$ providing that the errors are Gaussian, with zero mean and covariance matrix C.

A priori information about the model are introduced directly in the inversion by formulating the prior estimate $\vec{x}_0$ as data, subject to unknown errors (Jackson and Matsu'ura, 1985).

$$\vec{x}_0 = \vec{x} + \vec{\delta} \tag{2.16}$$

Where $\vec{\delta}$ is a m-vector of unknown errors and $\vec{x}$ is the true model. The further proceedings are analogous to the data.

For the derivation of the model vector, Bayes' rule (Hoel, 1971) is applied:

$$p(\frac{\vec{x}}{\vec{y}}) = p(\frac{\vec{y}}{\vec{x}}) p(\vec{x}) / p(\vec{y}). \tag{2.17}$$

Here, $p(\frac{\vec{x}}{\vec{y}})$ is the *a posteriori* joint pdf which expresses the probability to obtain the parameter vector $\vec{x}$ for a given data vector $\vec{y}$. $p(\frac{\vec{y}}{\vec{x}})$ corresponds to the probability that $\vec{y}$ is observed for a hypothetical, assumed parameter vector $\vec{x}$; p($\vec{x}$) is the *a priori* joint pdf for the parameters and p($\vec{y}$) is a function independent of $\vec{x}$ to normalize the equation.

The Bayes' theorem expresses the *a posteriori* probability of the solution in terms of the probability of the *a priori* information. Thus the *a posteriori* likelihood can be maximized knowing the *a priori* pdfs.

In case the observational errors of $\vec{y}$ are statistically independent of the errors of the parameters $\vec{e}$ the *a posteriori* pdf reads

$$p(\frac{\vec{x}}{\vec{y}}) = a \cdot e^{\frac{-T^2}{2}} \tag{2.18}$$

where a is a constant and

$$T = \vec{d}^T C_d^{-1} \vec{d} + \vec{p}^T C_p^{-1} \vec{p} \tag{2.19}$$

In equation 2.19 $\vec{d}$ and its transposed $\vec{d}^T$ are n-vectors containing the residual data and C_d is an $n \times n$ data covariance matrix. Similarly, $\vec{p}$ and $\vec{p}^T$ are m-vectors of the residual model parameters and C_p is an $m \times m$ model covariance matrix.

To maximize the likelihood $p(\frac{\vec{x}}{\vec{y}})$ the least squares expression 2.6 has to be minimized.

These two constraints are realized in the inversion as follows:

$$(\vec{d}_{obs} - \vec{d}_{cal})^T C_d^{-1} (\vec{d}_{obs} - \vec{d}_{cal}) \tag{2.20}$$

The n-vectors $\vec{d}_{obs}$ and $\vec{d}_{cal}$ contain the observed data and the data calculated for a hypothetical model vector, respectively. The n entries consist of n_{dt} values for the teleseismic relative delaytime residuals and n_g values for the Bouguer gravity data. The resulting model should minimize the variance between $\vec{d}_{obs}$ and $\vec{d}_{cal}$. C_d is a $n \times n$ matrix and contains *a priori* data variances along its main diagonal. Since the data are considered to be independent, all off-diagonal elements are zero. The *a priori* data variances take into account variations in data accuracy and scale variations between the data sets.

For the incorporation of the *a priori* information concerning the model parameters, the term to be minimized is:

$$(\vec{p} - \vec{p}_0)^T C_p^{-1} (\vec{p} - \vec{p}_0) \tag{2.21}$$

The m-vectors $\vec{p}_0$ and $\vec{p}$ contain the initial *a priori* estimates of the model parameters and a hypothetical model parameter vector, respectively. Both, m_v velocity variances and m_ρ density contrasts are contained in $\vec{p}_0$ and $\vec{p}$. Since the model parameters are considered to be independent, non-zero *a priori* variances of the model parameters only occur along the main diagonal of the $m \times m$ matrix C_p. These control how far the corresponding model parameters may deviate from the initial estimate. Large variances let the parameters vary freely from the originally assumed values while small ones rather fix the (well known) *a priori* model parameters (Zeyen and Achauer (1997)).

Zeyen and Achauer (1997) add two more constraints, the velocity-density relation and the smoothness of the density model, as *a priori* information.

A quasi linear relation of velocity and density variations is assumed (2.12). Since the correlation value (B) does not only change with depth, but may also change within a layer, it is treated as an unknown model parameter, which may vary. Hence, B has to be determined as mean correlation value for each layer resulting in n_B additional model parameters in the model vector $\vec{x}$. The respective *a priori* information consists of initial estimates about the mean correlation between velocity contrasts and density variations in each layer and reasonable estimates about the allowed deviations from linearity. The corresponding condition to be minimized results in the term:

$$(\Delta\vec{v} - B\Delta\rho)^T (C_b)^{-1} (\Delta\vec{v} - B\Delta\rho) \tag{2.22}$$

$\Delta\vec{v}$ and $\Delta\vec{\rho}$ are m_v and m_ρ-vectors that compose the model vector $\vec{p}$. Matrix B relates the corresponding parameters in both vectors. Hence, the dimension of the covariance matrix C_b is $m_v \times m_v$ and of matrix B it is $m_v \times m_\rho$.

The density model is in large parts not constrained by velocity model parameters. Due to the non-uniqueness of the gravity inversion these model parameters are prone to be attributed unreasonable large amplitudes. Therefore, Zeyen and Achauer (1997) add a smoothness constraint that connects the unconstrained parameters to the neighboring ones and to the rest of the model. The corresponding expression reads:

$$\left(\frac{\Delta\vec{p}}{\Delta R}\right)^T (C_s)^{-1} \frac{\Delta\vec{p}}{\Delta R} \tag{2.23}$$

$\Delta\vec{p}$ is the difference of the density model parameters of adjacent density blocks and ΔR is the distance of the corresponding block centers. Covariance matrix C_s contains non-zero values along the main diagonal for unconstrained blocks and off-diagonal elements for the corresponding neighboring blocks. C_s is also responsible for the relative influence of the smoothness with respect to the other constraints.

These four conditions result in a least squares expression that has to be minimized:

$$
\begin{aligned}
T \quad = \quad & \left[(\vec{d}_{obs} - \vec{d}_{cal})^T C_d^{-1} (\vec{d}_{obs} - \vec{d}_{cal}) \right] + \left[(\vec{p} - \vec{p}_0)^T C_p^{-1} (\vec{p} - \vec{p}_0) \right] \quad (2.24) \\
+ \quad & \left[(\Delta\vec{v} - B\Delta\rho)^T (C_b)^{-1} (\Delta\vec{v} - B\Delta\rho) \right] + \left[(\frac{\Delta\vec{p}}{\Delta R})^T (C_s)^{-1} \frac{\Delta\vec{p}}{\Delta R} \right]
\end{aligned}
$$

After linearization (see Zeyen and Achauer, 1997) the following equation for the iterative solution of the inversion problem results.

$$
\begin{aligned}
\vec{p} = \vec{p}_0 + \quad & (A^T C_d^{-1} A + C_p^{-1} - C_d^{-1} D_b - C_s^{-1} D_s)^{-1} \cdot \quad (2.25) \\
& (A^T C_d^{-1} (\vec{d} - \vec{c}) - C_p^{-1} (\vec{p}_0 - \vec{\mu}) + C_b^{-1} \vec{b} + C_s^{-1} \vec{s})
\end{aligned}
$$

Where $\vec{p}$ is the solution of the current iteration and $\vec{p}_0$ is the respective starting model. The $n \times m$ matrix A contains the partial derivatives of the forward calculated data with respect to the model parameters. D_b and $\vec{b}$ are a matrix and a vector related to the velocity-density relationship (2.22), respectively. D_b is a symmetric matrix with different blocks corresponding to velocity parameters and constrained density parameters, their cross products and the B-values. Vector $\vec{b}$ basically corresponds to $\Delta\vec{v} - B\Delta\rho$. The detailed composition of D_b is given in Zeyen and Achauer (1997). D_s and $\vec{s}$ are a matrix and a vector corresponding to the smoothness constraint. The elements of D_s are composed of the summed squared distances between adjacent density blocks, while $\vec{s}$ is composed of the differences of density variations of the neighboring blocks, weighted by their distance.

Since the inversion part is performed by a full matrix inversion, the parameter resolution matrix, connecting the 'true' Earth and the obtained solution, can be calculated as follows:

$$
R = (A^T C_d^{-1} A + C_p^{-1} - C_d^{-1} D_b - C_s^{-1} D_s)^{-1} \cdot A^T A \quad (2.26)
$$

In contrast to ordinary inversion schemes the $m \times m$ resolution matrix is not mainly used to determine where the inversion result is close to the 'true' Earth, but it is used prior to the actual inversion to determine a model in which each model parameter can be resolved equally well, leading to stable results.

The flowchart of the realization of the JI-3D method is given in figure 6.1.

2.2.3 3D-raytracing

The raytracer, based on Fermat's principle that raypaths are stationary with time, was originally developed by Steck and Prothero (1991). The minimum traveltime raypaths from a planar wave front at depths to a fixed point at the surface (the seismic station) is determined through a continuous heterogeneous 3D velocity structure. The traveltime is directly minimized by systematically
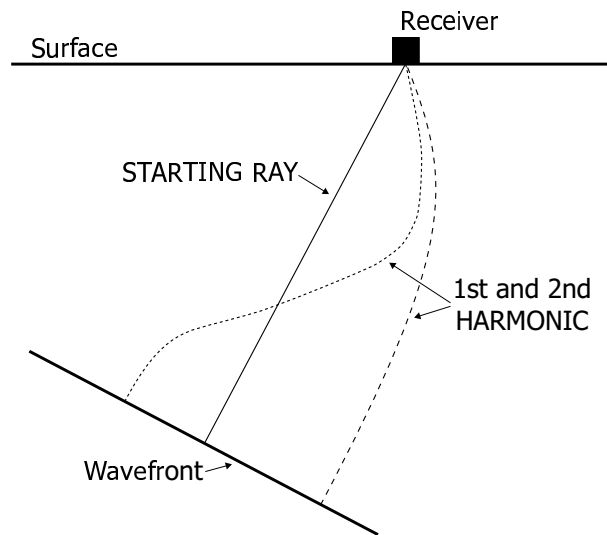

Figure 2.7: *Raytracing starts with a straight ray connecting the planar wavefront and the station. This starting ray is successively perturbed by sinusoids and first and second distortions are indicated by dashed lines. The harmonic functions always have zero displacement at the station and always remain perpendicular to the wavefront.*

distorting an initial straight ray until the minimum traveltime path is found.

To determine the 3D-raypaths, successively higher harmonics are added to the starting ray in both vertical and horizontal direction during an iterative process (see figure 2.7). Odd quarter cycles ($\theta = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, ...$) of sinusoids are used to ensure, that the perturbations are always 0 at the station and that the ray is always perpendicular to the wave front. The amplitudes of the sinusoids are varied, until a minimum time path is found which is used as starting ray for the next iteration. The optimum amplitudes are determined via the simplex algorithm (Nelder and Mead, 1985), which is explained in detail in Prothero et al. (1988). To take into account that the amplitude of a particular harmonic may not be the optimum one after the next few harmonics are added, multiple sweeps through the suites of harmonics are performed, adapting the amplitudes.

The accuracy of the raytracer was examined extensively by Portmann (2000) who determined, that the parameters that are vital for correct traveltimes and raypaths are the steplength along the raypaths (i.e. the sampling rate), the number of harmonics and the so called cut-off parameter that determines if the iterative calculations have converged. The respective values (for a standard teleseismic raytracing problem) were 1 km as sampling rate, 10 to 15 harmonics and a cutoff parameter of at least $2 \cdot 10^{-7}$. The use of wrong raytracing parameters leads to at least partially wrong traveltimes and raypaths, and therfore, wrong inversion results since the data then are projected to the wrong places in the model space.

The extent of the influence of the anomaly on the raypath found by the raytracer depends strongly on two parameters: First, on the spacial extent of the anomaly, as a ray will travel right through a

low-velocity zone, if the possible bypass is too time consuming. Second, the velocity gradient at the edges of the anomaly has a strong influence on the raypath. The stronger the velocity gradient, the stronger is the need for bending the raypath compared to the straight one. A gridding chosen too coarse, for instance, will lead to an artificially weak gradient, even if the velocity contrasts between adjacent nodes are large. In such a case hardly any difference from the undisturbed ray is observed. This is important to evaluate the forward calculations and inversion results with respect of comparability to real applications. In figure 2.8 the influence of amplitudes and gradients of heterogeneities on the raypaths is displayed. The amplitudes examined are $-10\%, -5\%, +5\%$ and $+10\%$. On the left hand side a 2 km gradient (i.e. the velocity jump occurs within 2 km) and on the right hand side the 10 km case is examined. It is obvious, that both, the positive and negative 10% cases change the raypaths to a large extend. As one would expect, steeper velocity gradients cause a more abrupt change of the raypaths, but the general behavior remains the same. In the 5% cases, the differences between hard and soft-bound anomalies is more significant, since the change of the raypaths due to the anomalous structure is much stronger in the steep gradient case. In addition to that, strong diffractions are obvious to occur for the 2 km gradient and -5% and -10% anomalies. For the weaker gradient, diffractions can only be seen for the latter case.

This shows, that even anomalies with moderate amplitudes may have a strong influence on the raypaths. In the case of negative anomalies, also strong diffractions may occur, which may be a problem to real applications if not the first arrivals of a seismic wave are used for the determination of the travel time residuals, but certain seismic phases. In addition to that, figure 2.8 shows, that 3D raytracing is vital for the iterative teleseismic delaytime inversion, since the raypaths depend strongly on the velocity structure and vice versa.

Table 2.1 gives the details of the model used for the calculations:

layer	depth range (km)	v_p background (km/s)	v_p anomaly (%)
1	0-15	6.0	
2	15-30	6.0	-10 to +10
3	30-50	8.0	-10 to +10
4	50-70	8.0	
5	70-120	8.3	

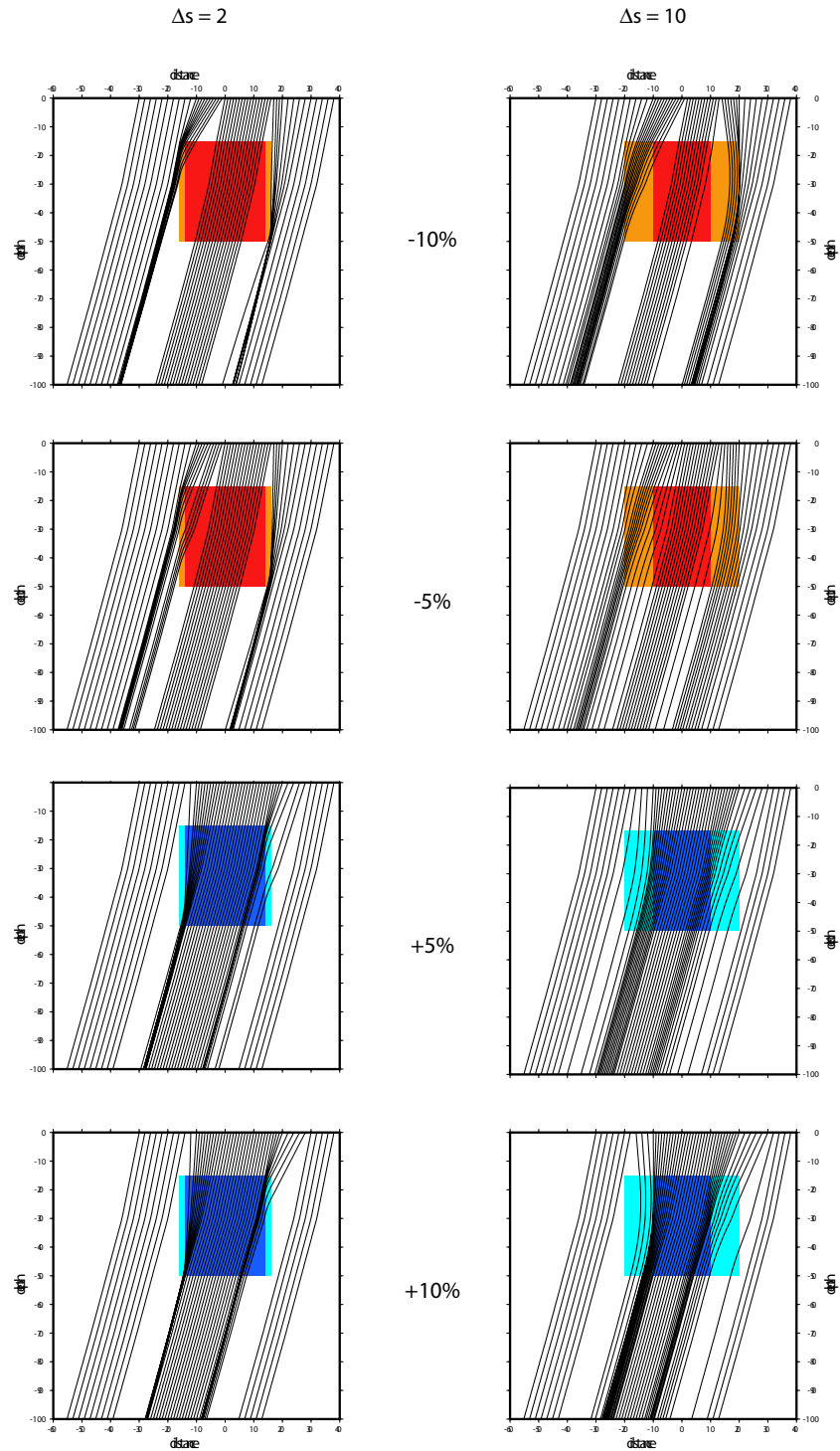Table 2.1: *Summary of the input model for the raytracing.*

Figure 2.8: *The influence of different velocity contrasts and gradients on the raypaths calculated by the 3D-raytracer is illustrated. The depth scale is negative. In the left row velocity contrasts of $-10\%, -5\%, +5\%$ and $+10\%$ are examined using a large gradient. Δs is the distance of nodes, defining the width of the gradient zone. In the right row the same velocity contrasts are used for a smaller gradient. Reddish colors indicate low velocities, bluish fast ones. The brighter regions at the edges of the anomalies indicate the gradient zones.*

2.2.4   Variable parameterization

The aim of the inversion method is to achieve inversion results with maximum stability and (spacial) resolution at the same time. Many inversion problems are ill posed i.e. the model parameters are over- and/or underdetermined. This means, that some parameters are not constrained independently while others are constrained by many data which, however, are erroneous and, therefore, contradict each other. The resulting models often are neither well resolved, nor unique. To avoid ill posed problems we demand that all model parameters have to be equally well constrained and that each model parameter should have the same influence on the inversion result. The resolution of a model parameter depends on the distribution of information in the model space which strongly varies between different types of geophysical responses. E.g. the information density in the velocity model depends on the ray distribution, whereas in the density model it only depends on the distance from the observational points. Hence, we perform a variable parameterization, exactly according to the information density in the model space, separately for the velocity-, density- and correlation model parameters. This means, that this concept rather is to optimize the problem itself, instead of trying to keep control by overly damping and smoothing the result.

Vertical parameterization of the model space

As the correlation between velocity and density model parameters is done layerwise, an identical layering for the velocity, density and B-model is used. Since the teleseismic delay time inversion requires, that the time, any ray spends in each layer is constant, the layer thicknesses are controlled by the background velocity model. Each layer in the velocity model contains a nodal plane, where the velocity parameters are defined at discrete points (nodes). Since a nodal plane can be located anywhere within the respective layer (see figure 2.9), it is possible to incorporate e.g. 1st order velocity discontinuities by placing the nodal planes of two adjacent layers next to each other, around the common layer boundaries. This leads to more realistic raypaths and better inversion results.

Horizontal parameterization of the model space

The velocity model   The velocity model is parameterized exactly according to the ray distribution in the model space. In order to achieve maximum stability and uniqueness at the same time, the model parameters have to be equally well constrained and independently resolved, which requires a well determined, optimized inversion problem. To meet these requirements, we use a variable parameterization, consisting of nodes and surrounding blocks. The size and position of a block are chosen so that sufficient crossing rays pass the block. Sufficient does not necessarily mean the absolute number of rays, but rather a wide range of ray directions in terms of backazimuths and ray parameters. In addition to that, the quality of the rays' delaytime picks also have an influence on the parameterization. As a measure for the degree of optimization of the parameterization we use the resolution matrix, where the diagonal elements give a good impression if the model parameters are constrained equally. The off-diagonal elements show, if the individual model parameters are resolved independently from each other. Even parameters at the edges of the model can be resolved without oscillation. Obviously, it is difficult, to determine an appropriate parameterization for the edge blocks of the layers, since the ray distribution there is almost unidirectional. However, as the application shows, if the diagonal elements of the resolution matrix are all of approximately the same size, the inversion is stable and no oscillations occur. In addition to that, the model space below the station network is hardly affected by vertical smearing. Optimized parameterizations result in small blocks and high spacial resolution, where the ray distribution is
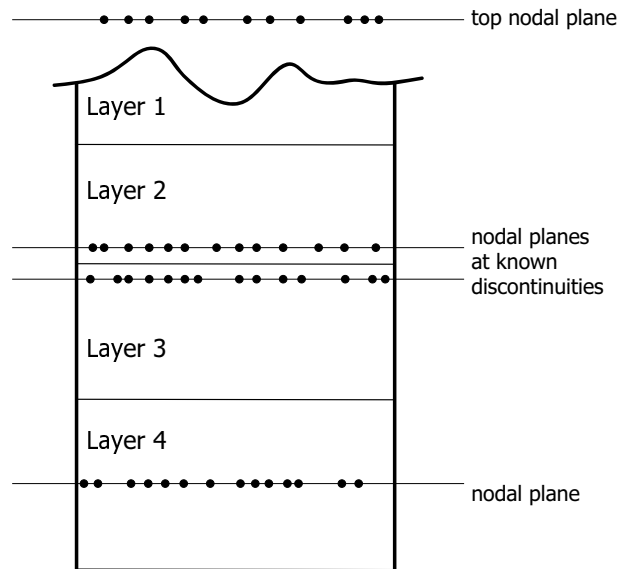
Figure 2.9: *Example of a vertical parameterization of the velocity model. Both nodal planes and layers are displayed. Note, that the nodal plane of the top layer is located above the topography.*

good, whereas in regions with poor ray coverage, the model parameters rather represent a stable average over a larger area. Hence it is important to consider the parameterization in the interpretation of the inversion results. As stated above, the velocity model does not only consist of blocks, but also of nodes, that the velocity perturbations are assigned to. This enables the model to be fully adapted to the ray coverage. While the blocks define the region which contributes to the determination of a certain model parameter, the corresponding node indicates the block's center of information. This means, that the inversion result is projected to the position, where the information comes from. If the inversion result was simply assigned to the geometrical center of the respective block the result could appear skewed, since the inversion result could be projected to places where no rays pass. This effect even would aggravate in an iterative inversion or in cases with a heterogeneous input model, since also the raypaths would be influenced by the wrong locations of the nodes. The velocity in between the nodes is interpolated linearly, so that no artificial velocity jumps can occur. The node concept also enables to model known velocity discontinuities, both vertically and horizontally.

Optimization of the model   The most favorable model is defined as being completely adapted to the ray distribution, so that all model parameters are equally well resolved. Hence, to control the quality of the inversion, the diagonal elements of the resolution matrix are used as a measure of optimization, which all should have the same value. The optimized model is obtained during an iterative process: First, the parameterization is done according to the ray distribution. Then, the resolution matrix is calculated and the model is reparameterized to level its diagonal elements. This is repeated until the diagonal elements have all about the same value. At the moment this is done manually, which definitively is tedious, but very instructive and important for the stability of the inversion result. Controlling both raypaths and resolution matrix enables to avoid strong vertical smearing even in areas where the ray distribution is poor. Since the inversion is done iteratively, the raypaths may change from iteration to iteration. This change is likely to be largest after

the first inversion in case of a homogeneous starting model. Hence it is strongly recommended to recheck the parameterization during each iteration and if necessary to adapt it.

Offset and average   Although the optimized parameterization is likely to produce stable results with high spacial resolution (in terms of wavelength) it is still arbitrary with respect to the unknown structure inside the model volume. As a consequence some anomalies may be only poorly recovered if they are 'smeared' over 2,4 or even 8 parameters (see Evans and Achauer, 1993). To mitigate this effect an offset and average technique, very similar to the one of Evans and Zucca (1988) can be applied (see figure 2.10). Shifting is done layerwise according to the smallest block-size in the respective layer. In contrast to the method of Evans and Zucca (1988), shifting is centered on the original parameterization to keep the node and block distribution adapted to the information density in the model space.
Of course, this cannot compensate the effect of very large blocks but leads to excellent results in the well sampled and densely parameterized areas of the model space. It has to be mentioned, that the velocity model has to be discretized in a slightly different way, taking into account each layer's shift for the offset and average procedure. The single resulting models usually are not all optimum parameterized with respect to the unshifted ray distribution. However, after averaging the results, usually reasonable models are obtained.

The density model   In case of the inversion of gravity data the corresponding density model is parameterized in terms of blocks. In contrast to teleseismic delaytime inversion the information density in the model space merely depends on the distance from the points of observations in case of gravity inversion. In order to have equal weights and influence of all density model parameters on the inversion, the blocksizes increase with distance from the observational points.
The size of these blocks is chosen to have the minimum resolvable size, which, according to Zeyen and Pous (1993), generally equals the distance, the particular block has from the observational points. In this case the block sizes are chosen so that the gravitational attraction of the undisturbed blocks have the same influence on the data. The minimum possible block sizes are implied by the geometry of the array. It is determined by the Nyquist frequency so that the blocksize cannot be smaller than the spacing of the data points. Again the quality of the parameterization has to be checked using the resolution matrix.

B-Value model   The correlation between velocity and density model parameters $\delta\rho = B\delta v$ (Lees and VanDecar, 1991) is assumed to be quasi linear within a layer so that the individual correlation factors may vary around a mean one. Hence one B-value per layer is determined by the inversion and the parameterization consists of assigning one mean B-value per layer.

The incorporation of a priori information into the inversion   The Bayes approach allows to incorporate both, specific and nonspecific *a priori* information into the inversion. Specific *a priori* information generally refers to some, known model parameters. This, e.g. may be the density and velocity structures along refraction lines, or, more punctuated, from bore holes or even previous tomography studies. In addition to that, also known first order discontinuities can be incorporated into the model by placing the layer boundaries and nodal planes accordingly. Moreover, nonspecific information (like smoothness of the density model) can also be included. This can be applied to all density model parameters or merely to those, that are not constrained by the velocity model.
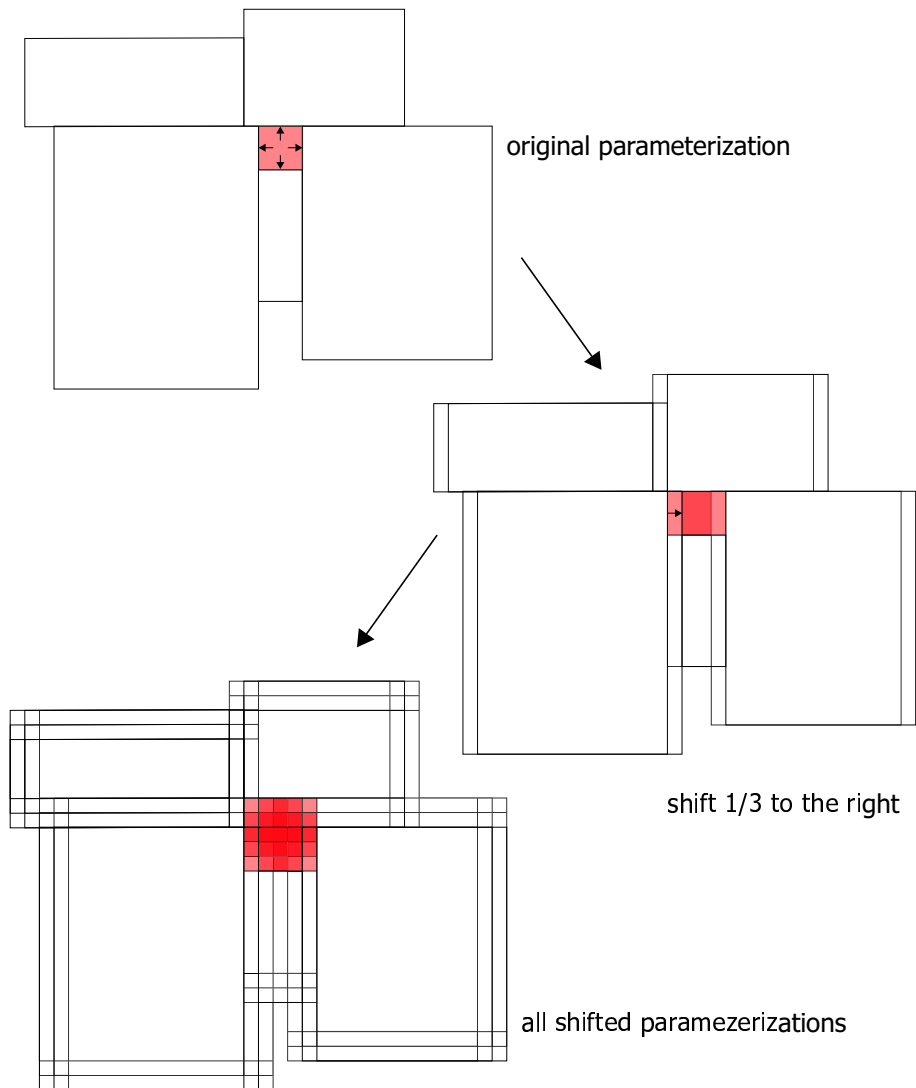
Figure 2.10: *The offset and average technique is shown at a sample layer. The offset is individually determined for each layer, according to the smallest block size. The whole parameterization is shifted by $\frac{1}{3}$ of the respective side length in both positive and negative direction. Inversions are calculated for the 9 parameterizations and the results are averaged leading to an inversion result more unbiased by the initial parameterization.*

2.3 Application to synthetic data

2.3.1 Relative teleseismic traveltime residuals

Teleseismic traveltime residuals and Bouguer gravity data are calculated for a synthetic input model (figure 2.11) which is designed to challenge teleseismic delaytime tomography. A simple background velocity model is used (6.0 km/s and 8.0 km/s) with a first order discontinuity at 30 km depth (see table 2.2). There are two anomalies: The first one is a 20 km wide NS aligned structure with the upper and lower boundaries at 15 and 30 km, respectively. Its lateral relative velocity contrast is +10%. The second one is an EW directed structure, 30 km wide with a negative perturbation of -10%, extending from 50 to 70 km depth. For a summary of the properties of the synthetic input model see table **??**.

layer	depth range (km)	v_{0p} (km/s)	Δv_p (km/s)	ρ_0 kg/m^3	$\Delta\rho$ kg/m^3	B $\frac{km/s}{kg/m^3}$
1	0-15	6.0	0.0	2700	0	$3.0 \cdot 10^{-3}$
2	15-30	6.0	0.6	2700	200	$3.0 \cdot 10^{-3}$
3	30-50	8.0	0.0	3060	0	$3.3 \cdot 10^{-3}$
4	50-70	8.0	-0.8	3060	-240	$3.3 \cdot 10^{-3}$
5	70-120	8.0	0.0	3060	0	$3.3 \cdot 10^{-3}$

Table 2.2: *Summary of the synthetic input model. v_{0p} and ρ_0 are the absolute velocity and density values of the background model, Δv_p and $\Delta\rho$ are the absolute perturbations in the respective velocity and density anomalies. B is the correlation factor between velocity and density model.*

The synthetic input model is difficult to resolve for common teleseismic tomography methods for several reasons. First, the spacing of the two anomalies is small, so that vertical smearing is likely to obstruct their separate reconstruction. Moreover, the upper structure is close to the surface (15-30 km depth). Taking into account the station spacing of 20 km, the maximum ray parameter of 8.6 *s/°* and a crustal velocity of 6.0 km/s, the shallowest depth level, where rays cross, is about 20 km. As a consequence, we expect vertical smearing of the upper anomaly towards the surface. Both problems, the vertical smearing and the missing resolution in the crust are typical problems of teleseismic tomography. Second, due to the opposite signs of the anomalies, their signals in the teleseismic delaytime data cancel out in some places. Third, the amplitude of the anomalies are large ($\pm10\%$) leading to significant effects on the raypaths which demands an iterative inversion scheme. In addition to that, the anomalies are bounded by very large velocity gradients. This leads to strong diffracted phases at the negative velocity anomaly which are difficult to reproduce with the same amplitudes, since the minimum nodal spacing in the inversion model is about 10 km causing significantly weaker gradients. This is a worst case condition, since the edges of true anomalies are mostly blurred. Moreover diffractions are rarely used teleseismic tomography, since they carry only little energy resulting in small amplitudes in the seismogram.

The relative traveltime residuals are calculated using a minimum traveltime 3D-raytracer, based on the work of Steck and Prothero (1991) and described in 2.2.3. Gaussian noise with a standard
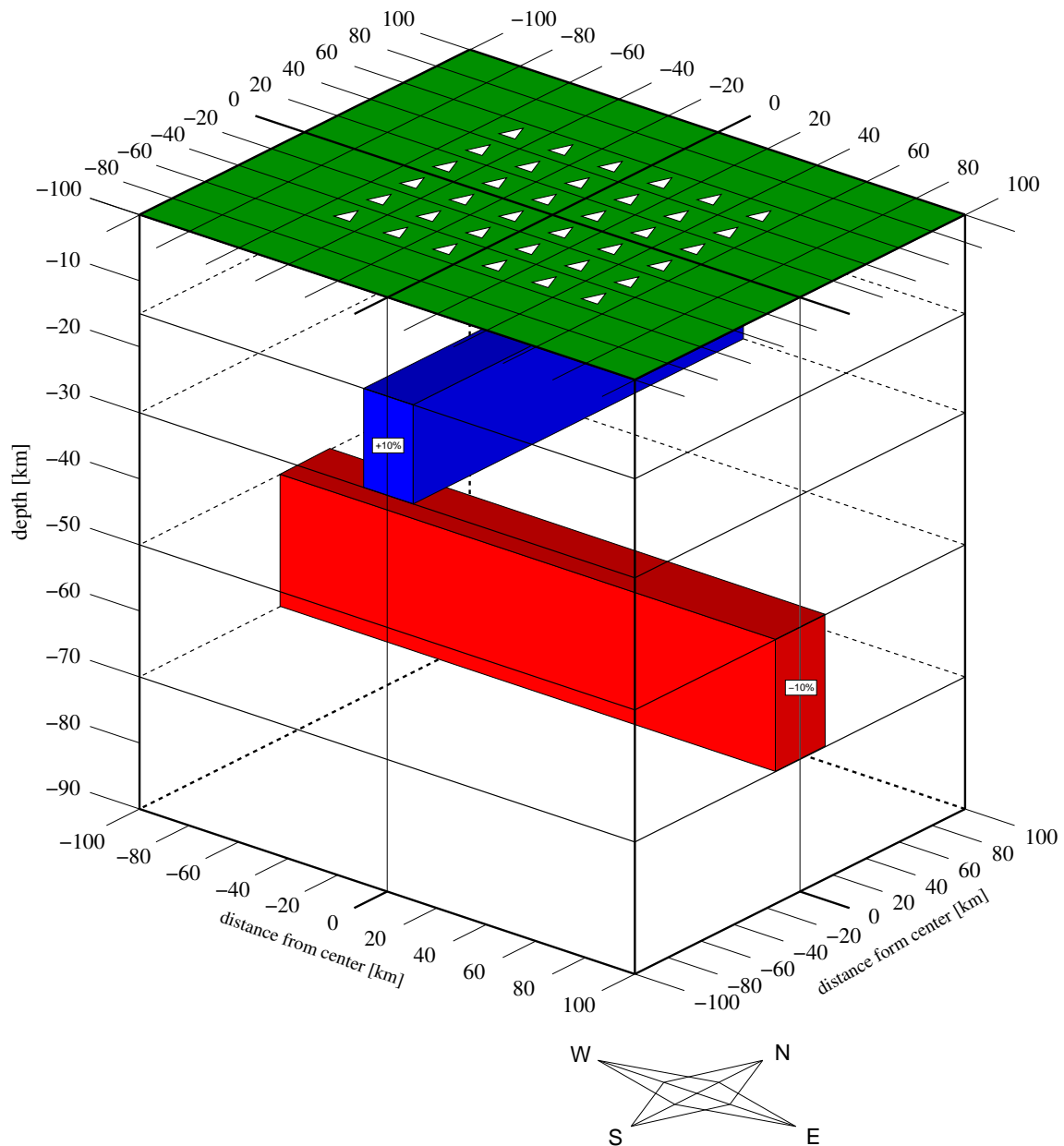
Figure 2.11: *Display of the input model and seismic station distribution for the calcula-*
tion of the synthetic seismic and gravimetric data sets. At the surface, 36
seismic stations are shown, indicated by white inverted triangles. The model
for the forward calculations contains two heterogeneities: First, a +10% ve-
locity anomaly with an increased density of 200 kg/dm³ between 15 and 30
km depth. This structure is NS oriented and 20 km wide in EW-direction.
Second, a −10% velocity anomaly with reduced density of -240 kg/dm³ be-
tween 50 and 70 km depth. This structure is EW-oriented with NS extension
of 30 km. The background velocity model is 6 km/s in between 0 and 30 km
depth and 8 km/s below. Apart from the two heterogeneities mentioned, the
model is laterally homogeneous. Note that the the depth scale is negative.

deviation of 0.04 s is added to the data. The calculations are performed for an array of 36 evenly distributed stations with a spacing of 20 km and 121 ideally distributed events, resulting in 4356 relative traveltime residuals. The ray parameters range from 8.6 $s/°$ to 4.9 $s/°$ in 5 even steps and 20 different, evenly distributed, backazimuths are used. In addition to that also 1 core phase event is contained in the data.

Figure 2.12 displays mean relative traveltime residuals for 9 different directions at the station sites. Negative relative residuals are indicated by bluish, positive ones by reddish colors. The plotted signals can be interpreted as the projections of the anomalous structures on the surface, so that the fast anomaly causes early arrivals and negative residuals in N-S direction, while the low velocity anomaly causes late arrivals and positive residuals at the respective stations. Moreover, one can see clearly some of the difficulties mentioned above: The responses of the opposing signals cancel in some places and the diffractions comprise the largest signals with very early arrivals (for Rays from N and S).
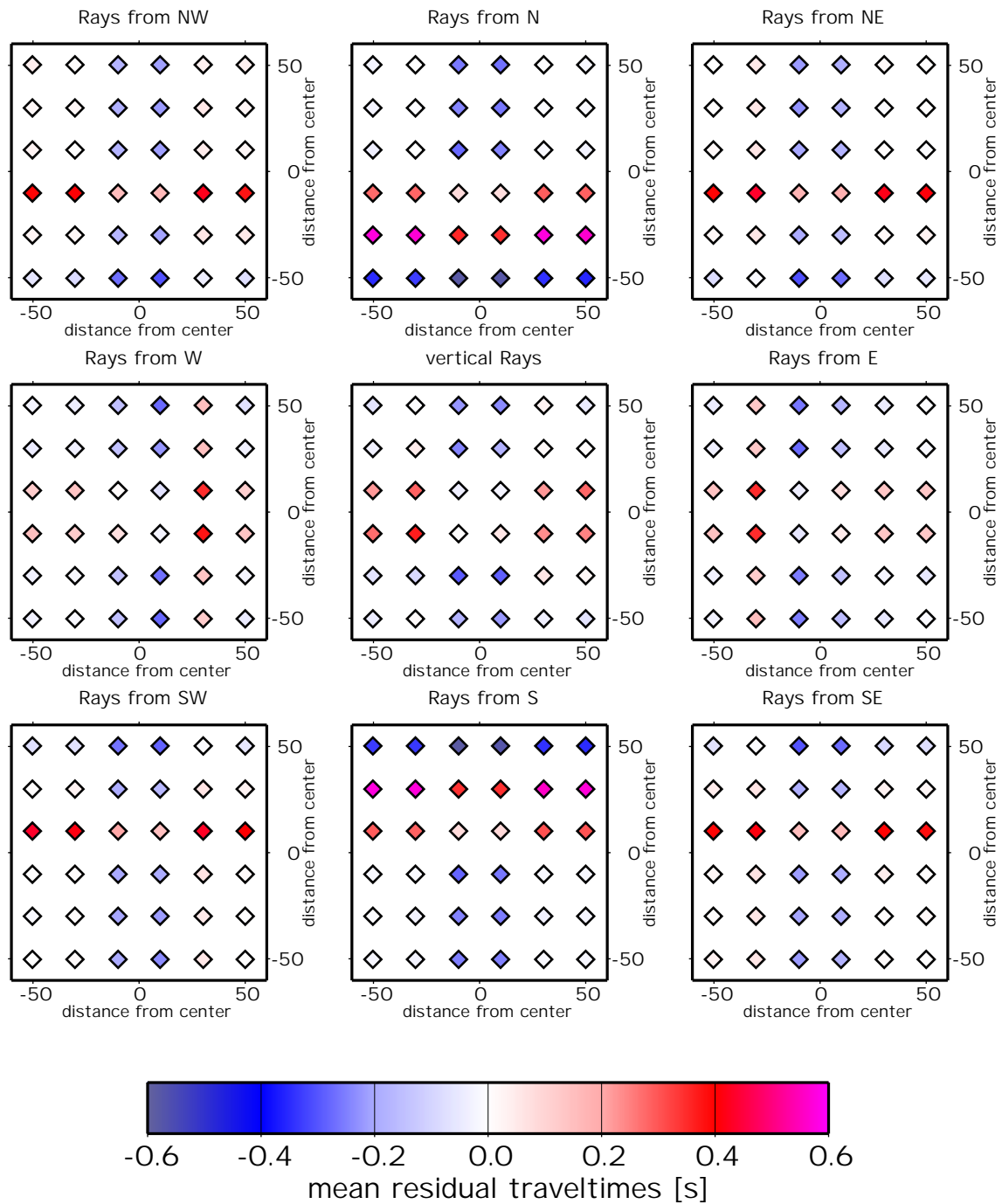
Figure 2.12: *Dataset for the delaytime inversion as calculated from the synthetic model (2.11). The dataset is displayed in terms of mean residuals for 9 different directions at the 36 stations (diamonds). I.e. the relative residual traveltimes are averaged, according to incidence angle and event direction. Reddish colors correspond to positive residuals and late arrivals while bluish colors indicate negative residuals and early arrivals. The data include noise.*

2.3.2 Bouguer gravity data

For the forward calculation of the Bouguer data we also use the model displayed in figure 2.11 with background densities of 2700 and 3060 kg/m^3. The density contrasts of the anomalous structures are calculated from the absolute velocity perturbations using a linear correlation. After Birch (1961) $v_p = A + B\rho$, where v_p is the compressional wave velocity, ρ the absolute density, B the correlation factor between velocity and density and A is a constant. Consequently, the perturbations are related by $\delta v = \delta\rho B$. Figure 2.13 shows the Bouguer gravity data in mGal calculated at the observational points, indicated by open circles. The dense spacing of the observational points (2 km) coincides with the region where the seismic stations are located. Besides, there are more observations far outside that region, since the anomaly cannot be identified from the inner region alone. Hence, the synthetic input and inversion density models have to extend very far into the halfspace in order to avoid edge effects. The maximum amplitudes of the forward calculated Bouguer signal are -29.7 mGal and $+31.8$ mGal. The forward calculation is performed using the algorithm of Zeyen and Pous (1993).
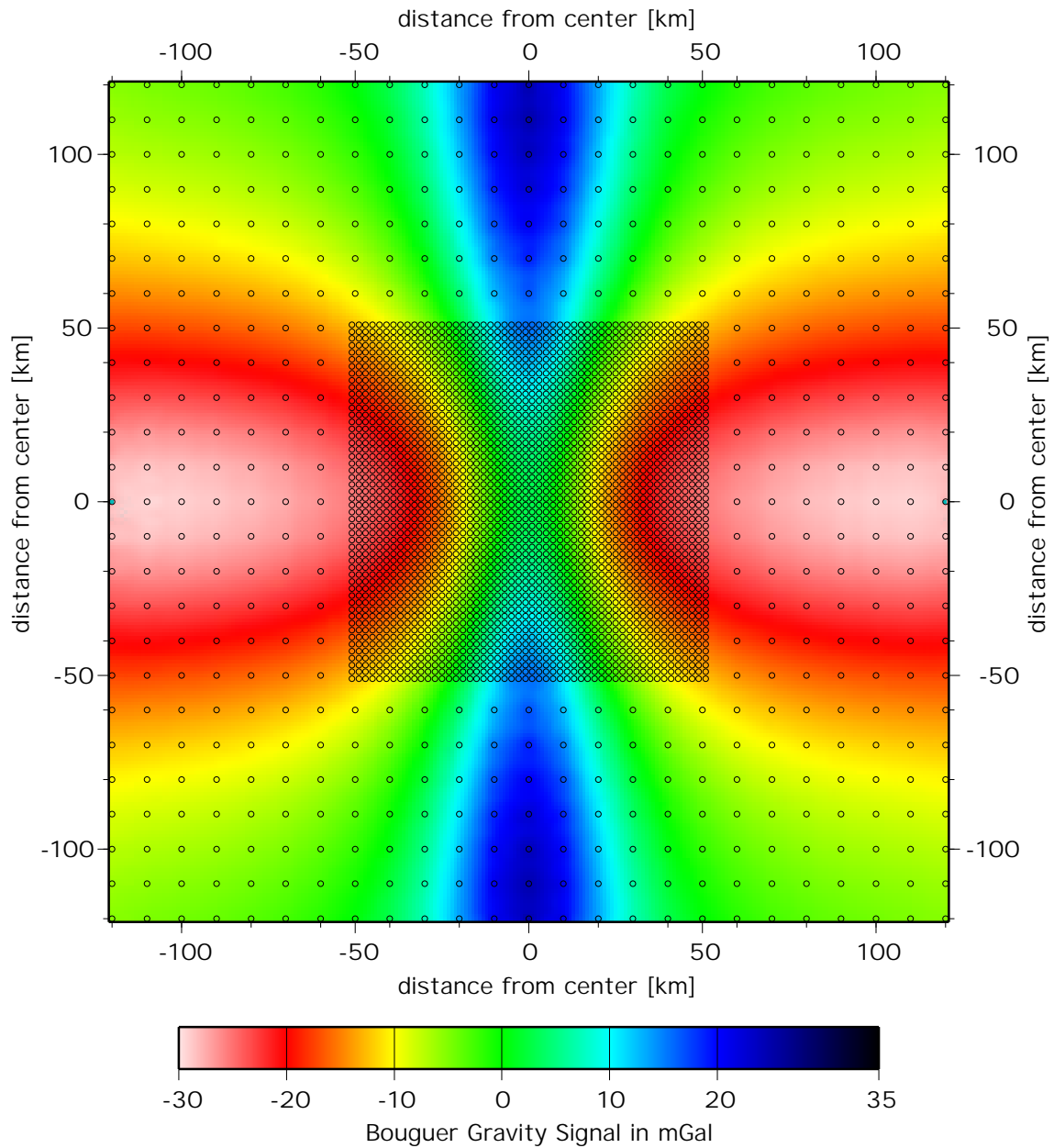
Figure 2.13: *The Bouguer gravity signal is 'observed' at the locations shown as open circles. The inner densely sampled region coincides with the seismic array. Since the complete Bouguer signal cannot be observed from this region alone, gravity observations extend much farer, while the spacing is reduced to keep the number of data within a reasonable limit.*

2.3.3 Parameterization of the model space

In order to invert the synthetic data and recover the input structures, models for the velocity, density and correlation model parameters have to be constructed. The proceedings are as described in chapter 2.4.

Vertical parameterization of the model space

Starting from the aperture of the array of seismic stations (100 km), the maximum depth of the model space is chosen to be 90 km. As standard background velocity model we chose 6.0 km/s for the crust (0-30 km depth) and 8.0 km/s below. In real applications one would apply the velocities from a standard Earth model.

The model space is subdivided into 10 layers, so that a seismic ray spends approximately the same time in each layer. Details of the model are given in table 6.1. In a next step, the individual depths of the nodal layers for the definition of the velocity model parameters are defined: The nodal plane of the first layer is always located above topography, so that all ray segments are inside the defined model. In the deeper layers, the nodal planes are placed more or less in the middle of the respective layer. However, if velocity jumps or large gradients like the Mohorovicic discontinuity (Moho) are known *a priori*, these can be included into the model by placing the nodal planes tightly around the respective depths. In this example, the Moho is assumed to be at 30 km depth, so that the nodal planes are placed at depths of 29 km and 31 km. This results in more realistic raypaths.

The vertical parameterization of the density model is identical to the velocity layer-model determined above. Therefore, the B-model (spacial distribution of the velocity-density correlation parameters) is also fixed. A summary of the parameterization is given in table 6.1.

Horizontal parameterization of the model space

The velocity model   The aim is to obtain a velocity model that is optimized with respect to the ray distribution, so that all model parameters are equally well resolved. First, an initial parameterization is created by subdividing each layer into equally sized blocks. The sidelengths are chosen to be twice the wavelength of the dominant 1 Hz signal in the respective layer. Second, the resolution matrix for this initial parameterization is calculated using only blocks that are constrained by seismic rays (see figure 2.14). The respective diagonal elements of the resolution matrix are shown in figure 2.14. It is obvious, that blocks which are sampled by many rays, are resolved better than others and that, despite having a symmetric ray distribution, the diagonal elements are not completely symmetric. This is caused by the random data errors that influence the inversion and the calculation of the resolution matrix. During an iterative process, the less constrained blocks are merged, according to the ray distribution and the resolution, so that the diagonal elements step by step reach the same niveau. Single blocks, that are much better constrained than the others can also be split and the block borders can be shifted. Sometimes, this may be necessary that few blocks end up being smaller than the dominant wavelength (this is not the case here) to level the resolution. If this is the case, it is extremely important, that the respective inversion results are not interpreted as single blocks, but always in combination with the neighboring blocks. Figure 2.14 shows the resulting optimized parameterization and the diagonal elements of the resolution matrix. The smallest elements are larger than 65%. As stated above, the resolution and consequently the parameterization are not completely symmetric due to random data errors. Apart from one exception, the block sizes of the final model are not smaller than the blocks in the initial model. All blocks are larger than the wavelengths and (mathematically) well resolved. An overview about the initial and final blocksizes of the individual layers is given in table 6.1. In a further step, the

position of the nodes have to be determined according to the ray distribution. The distribution of the velocity nodes is given in figure 2.14 as dots in the blocks of the model.

The complete figures displaying the ray distribution, initial and final parameterization are provided in the appendix (figures 6.2,6.3 and 6.4).

The density model   Starting from the vertical model derived above, the density model is parameterized by subdivision of the layers into blocks. The block sizes increase with distance from the observational points.

To enable a correlation between velocity and density model parameters also in larger depths of the model space, the block size in the first layer is chosen to have 4 km side length. The corresponding minimum block sizes in the deeper layers are displayed in table **??**. The minimum block size in the uppermost layer is based on the distance between the observational points (2 km) and the Nyquist frequency. However, this dense spacing of observations is only realized in the area of the seismic stations. Beyond that, the spacing is larger and the model is overparameterized. Since the density model parameters, that are not constrained by the velocity model are subject to smoothing during the joint inversion, this overparameterization is not a problem.

Beyond the region of gravity observations, the blocksizes increase.

Analogously, the blocks in the deeper layers become larger with distance from the surface gravity measurements. In contrast to the velocity model, the lateral extensions of the density model are very large in order to avoid edge effects. The complete parameterization of the velocity and density models are also shown in the appendix.
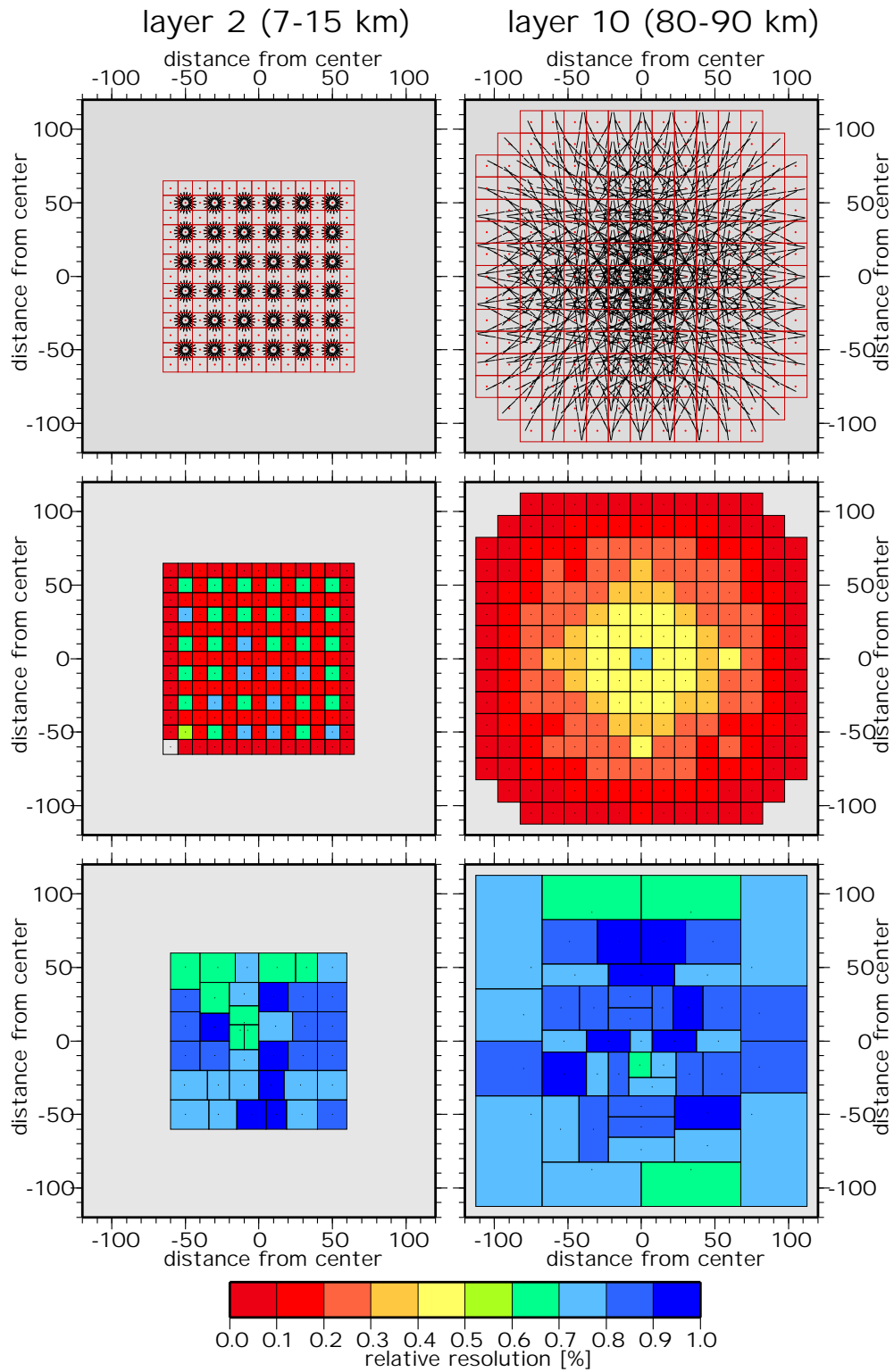
Figure 2.14: *Examples for layers 2 (left) and 10 (right) of the ray distribution in terms of the ray segments inside the layers (top), the initial parameterization and the respective diagonal elements of the resolution matrix (center) and the final parameterization also with resolution (bottom).*

2.3.4 Determination of amplitudes

For given data errors, the amplitudes of the inversion result mainly depend on the chosen *a priori* variances of the model parameters. Similar to ACH, several inversions are calculated for different *a priori* model variances and the respective data fit is plotted against the squared model length. The resulting trade-off curve is shown in figure 2.15.
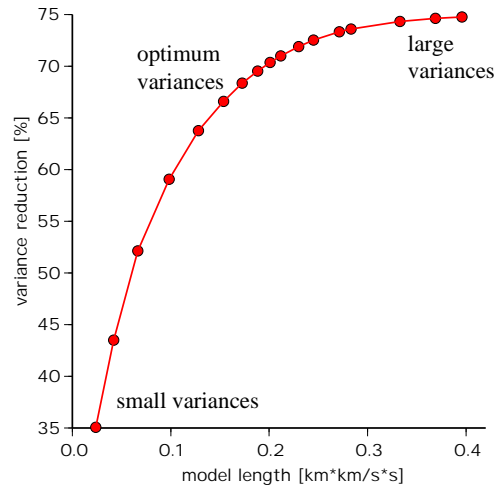
Figure 2.15: *Example of a typical trade-off curve. Each circle represents an inversion, calculated for different* a priori *model variances. The optimum variances are located at the bend of the curve where the models are still simple (small squared model length), but the data can be explained by the model to a satisfying degree. Small variances will lead to a model with little data fit and small amplitudes, large variances will lead to a large data fit at the expense of an oscillating model.*

The optimum model has to explain the data to a large degree and should be simple and non oscillating. Since both conditions cannot be met perfectly at the same time, there has to be a trade off between model length and variance reduction. Starting from small *a priori* model variances, the variance reduction increases quickly, while the squared model lengths only increase moderately. For large *a priori* model variances the model length increases rapidly while the variances do not improve significantly. The optimum model is situated in the bend of the trade-off curve where both variance reduction and model length are reasonable. Often there is not a single optimum model due to the absence of a clear sharp bend. In these cases the determination of an optimum *a priori* model variance and the corresponding amplitudes is done by fitting the residual traveltimes of single rays that sample perturbations in the spacially well resolved regions of the model space to calibrate the model. It is important to do this in the fine parameterized part of the model space in order to minimize the effect of contradicting data. Figure 2.16 illustrates the influence of the parameterization on the possible data fit. We consider two rays, one samples a velocity heterogeneity, the other one does not. In case a) the volume is parameterized with small blocks which enables a good data fit for both rays. In case b) both rays pass the same, large block which can only be resolved as an average over the whole region. However, since the rays sample different structures in the same block, it is not possible to obtain a reasonable data fit.
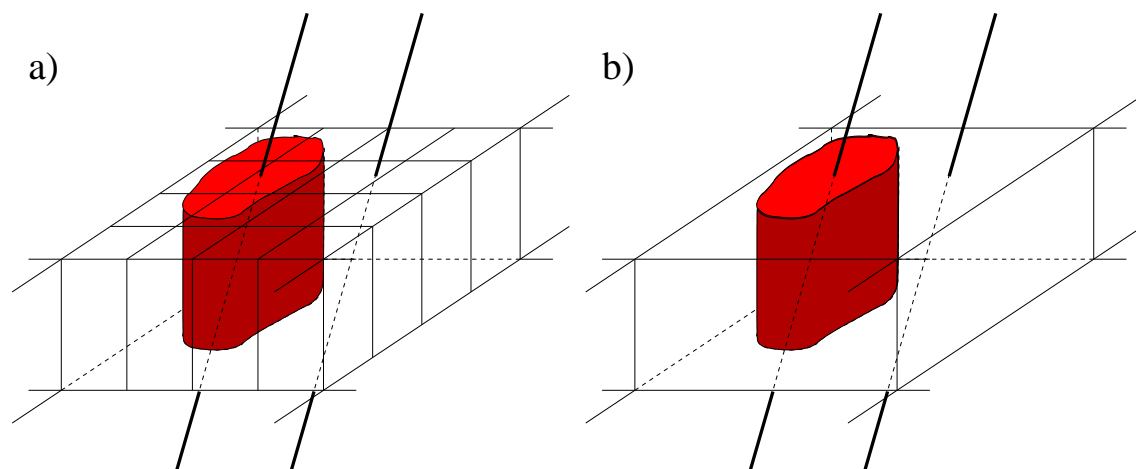
Figure 2.16: *Illustration, how the parameterization influences the possible data fit. Two rays pass a certain region of the model space which contains an anomalous structure (red area). In figure a) the residual traveltime of both rays can be explained by the model, which is not possible in figure b) since the data contradict each other.*

This effect, which is typical for inversion methods using large blocks (see also Spakman and Bijwaard (1998)) may reduce the possible data fit, depending on ray distribution, parameterization and the 3D velocity structure in the model volume. However, although the data fit is worse, the inversion result is not, as the JI-3D inversion -in contrast to ACH- is well controlled. In the iterative case, the optimum *a priori* model variance is different from the one of the non-iterative case. This is caused by the fact, that there are different data sets and starting velocity models for each iteration. In addition to that, there is also an automatic adaption of the *a priori* model variances, according to the achieved variance reduction and a change in the inversion algorithm as well due to additional terms (see figure 2.21). Therefore the trade-off procedure has to be done iteratively as well.

The optimum *a priori* density model variances in the case of the Bouguer inversion is done non iteratively, since the inversion problem itself is linear.

In case a joint inversion of teleseismic relative traveltime residuals and Bouguer gravity data is performed, also an iterative trade-off has to be done, since the inversion problem is strongly nonlinear.

Since the joint inversion process is strongly nonlinear, the inversion is done iteratively (see equation 2.25). Consequently, also the trade-off procedure has to be done iteratively. Unlike the single (and simple) delaytime and gravity inversions the joint inversion result depends on a set of *a priori* model variances. The resulting multidimensional trade-off procedure turns out to be very difficult. As already described in Zeyen and Achauer (1997) the nonlinearity introduced by the variation of the correlation parameter B often leads to instable results. Therfore, the *a priori* variances of the different parameters and the smoothness have to be chosen within narrow bands. A rule of thumb is presented how the optimum *a priori* model variances can be derived to keep the inversion stable. This is done by a serial multi-dimensional trade-off procedure. Since the variation of all *a priori* model variances at the same time would lead to a huge number of trial inversions this is computational absolutely prohibitive. Hence, the following proceeding is suggested: The joint inversion is started with *a priori* variances that are already close to the

optimum ones. First, a conservative starting velocity model is determined iteratively, so that the raypaths and the structures are close to the real ones already. Then an iterative trade-off for the starting model is calculated , providing the desired *a priori* velocity model variances. Second, the *a priori* density model variances are determined via a non iterative trade-off procedure. Third, the variances for the smoothing of the density model parameters that are not constrained by the velocity model are determined for non correlated velocity and density models. After that, the *a priori* variances of the B model parameters are varied until an optimum set of varaince reductions and model lengths is found for the different sets of data and model parameters. Since the inversion is clearly governed by the delaytime data, we prefer *a priori* variance sets with a good delaytime data fit. Finally, the resulting variances are varied in narrow bands around the obtained starting values in order to find the optimum models.

2.3.5   Data inversion

In this section, the details of the proposed inversion procedure are discussed for the application to the synthetic data set.

Overview

This section provides an overview of the proceedings that are recommended to obtain stable and reasonable joint inversion results. The choice of the *a priori* model variances and their combination are vital for achieving stable and reasonable inversion results. Special attention is paid to illustrate the procedures towards the optimum inversion results. In addition to that, intermediate results are shown to demonstrate the effects of the iterative and the joint inversion.

It is computational prohibitive to perform a full multi-dimensional trade-off procedure to search for the optimum combination of *a priori* model variances. Hence, the proceedings suggested are based on the separate determination of the optimum variances, prior to the search for the optimum variance of the linearity. The velocity and density variances are obtained using single inversions, the variances of the coupling parameter are obtained using joint inversions. The resulting combination of *a priori* model variances is a good base for further optimizations of the inversion.

In this section the determination of the optimum *a priori* model variances for the joint inversion is shown for the synthetic data set, presented above. Moreover, the evolution of the resulting models after the subsequent inversion steps are presented, to demonstrate the effects and benefits of the iterative and joint inversion.

The proceedings are as follows: First, the single delaytime and gravity inversions are carried out and the different methods for the calibration of the respective inversion results are presented.

Second, a starting model for the delaytime inversion is determined in an iterative inversion. Third, the optimum *a priori* variances are determined for the density and velocity models. Finally, the optimum variances of the B-Parameter is determined during a joint inversion trade-off procedure.

Delay time data inversion

The non iterative case   To illustrate the effect of the iterative inversion, first the results for the non iterative inversion and how they are obtained are presented. The non iterative approach has relevance to real applications, if simple structures with small amplitudes are examined, so that the raypaths are not affected by heterogeneities. The setup and the used algorithm of the

non-iterative case are comparable to ACH. However, the optimized parameterization provides a uniform mathematical resolution in all parts of the model space, leading to stable results. In addition to that, a high spacial resolution is achieved in well sampled areas.

The proceedings are as described in chapter 2. For a given parameterization, the optimum *a priori* variances of the free model parameters have to be determined. The optimum model, in terms of correct amplitudes and stability, is determined via a trade-off curve (see figure 2.17). Several inversions are calculated for different *a priori* variances. The input consists of the respective *a priori* standard deviations or the "allowed errors", referred to as ddv-values in this work. Though the dimension of the standard deviation is $\frac{km}{s}$, the ddv-values are merely absolute.

For constant ddv-values $1/ddv = \theta$, the "damping parameter" as used in ACH. Since the inversion results are stable over a large range of ddv-values which means, that the results merely differ in amplitudes, there is no clear bend in the trade-off curve. This leads to a variety of possible models with ddv values ranging from 0.05 to 0.12, indicated by larger red dots in figure 2.17. This difficulty is mainly a consequence of the variable parameterization and the effect of large blocks (see figure 2.16). In order not to be affected by contradicting data to much, the optimum ddv-value can also be determined using subsets of rays sampling only the central, well sampled and finely resolved region of the model space. This results in a trade-off curve with a more pronounced bend, as shown in blue in figure 2.17. The resulting set of possible models and ddv-values ranges from ddv=0.10 to 0.12. Another possibility of determining the amplitudes is the calibration via two single rays. Both rays have to pass the model in fine parameterized regions of the model space. One ray should sample an anomaly while the other one is considered as reference ray and therefore should not pass the heterogeneous structure. Analogous to the proceedings for a normal trade-off, it is examined, how well the differential traveltime of the rays can be reproduced by the model for different ddv-values. According to the green curve in figure 2.17, the best match is obtained for ddv=0.12. In all cases, the respective variance reductions are plotted against the same squared model length, comprising all model parameters.

The resulting optimum model (ddv=0.12) is plotted layerwise in figure 2.19. The velocity nodes are indicated by black dots while the black boxes mark the region of the model space that contributed to the calculation of the velocity perturbations of the respective node. Since relative traveltime residuals are used, there is no information about absolute velocities but we rather have to consider velocity contrasts for the interpretation. Positive velocity perturbations are indicated by bluish, negative ones by reddish colors.

Generally, both input structures (figure 2.11) can be recovered. However, also the general problems of teleseismic delaytime inversions, vertical smearing, especially into the first layer, become obvious.

While the positive NS-tending anomaly is shifted from layer 3 and 4 to layer 1 and 2 the main part of the negative EW-tending anomaly is shifted from layer 7 and 8 to layers 5 and 6 with a small rest in layer 7. This is due to vertical smearing, which is typical for teleseismic delaytime inversion. The inversion is especially affected by the missing crossfire in layer 1 and little crossfire in layer 2, so that the whole inversion result is shifted to the surface. Vertical smearing also is found in layers 3 and 4 which are affected by both, smearing from above and below. The horizontal extent of the anomalies is very well recovered and, considering the smearing, the amplitudes are well recovered.

However. there is a "counterstructure" in the last layer which is wrong with respect to the synthetic input anomaly. It is important to note, that this is no oscillation since the structure is also found in inversions with very small ddv-values (very large damping). The main reason for the false anomaly in the last layer and the strong upward shift of the whole structure are the (with respect to the true structure) wrong raypaths, calculated for the homogeneous starting model. This is an
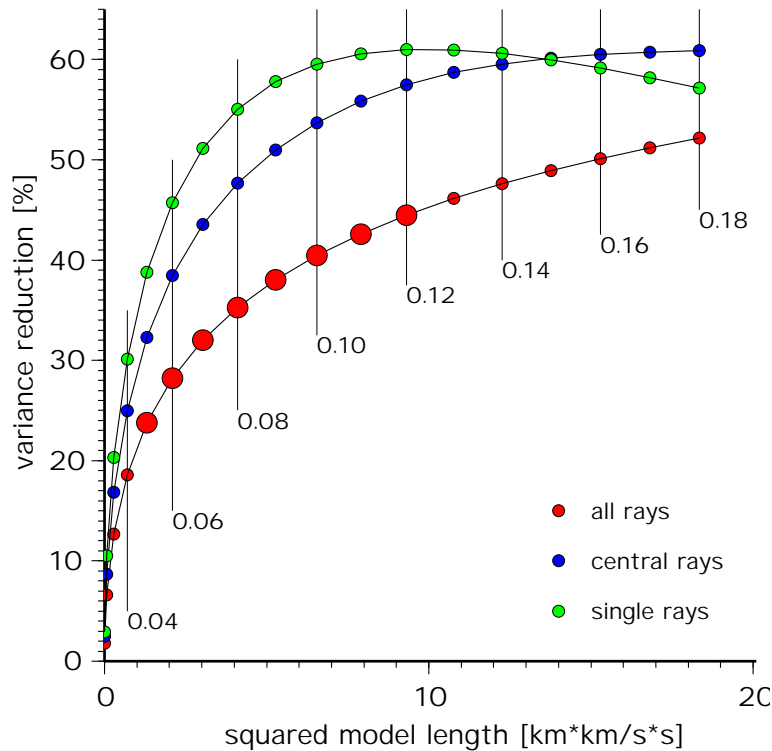
Figure 2.17: *Trade-off curves for the non-iterative delay time inversion. The three curves correspond to the same inversions (identical ddv-values), but different data (ray) subsets for the calculation of the variance improvement. This reflects the influence of the parameterization on the data fit and demonstrates several possibilities of determining the optimum ddv-value. While it is difficult to determine the optimum ddv-value, taking into account all rays, it is obvious for the single ray case.*
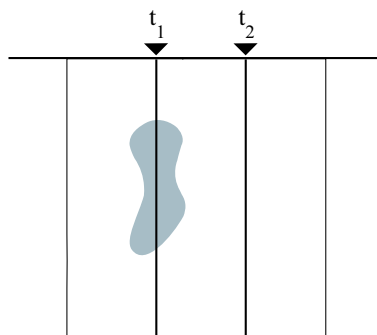

Figure 2.18: *Calibration of the amplitudes via two single rays. The left ray samples the anomaly which results in a residual traveltime t_1. The ray on the right hand side is used as reference for the unperturbed model, with residual traveltime t_2. Both rays have to pass the model through regions with high spacial resolution. Using the differential residual travel time $t_1 - t_2$, the amplitude of the anomaly can be well controlled and the optimum ddv-value can be found reliably.*
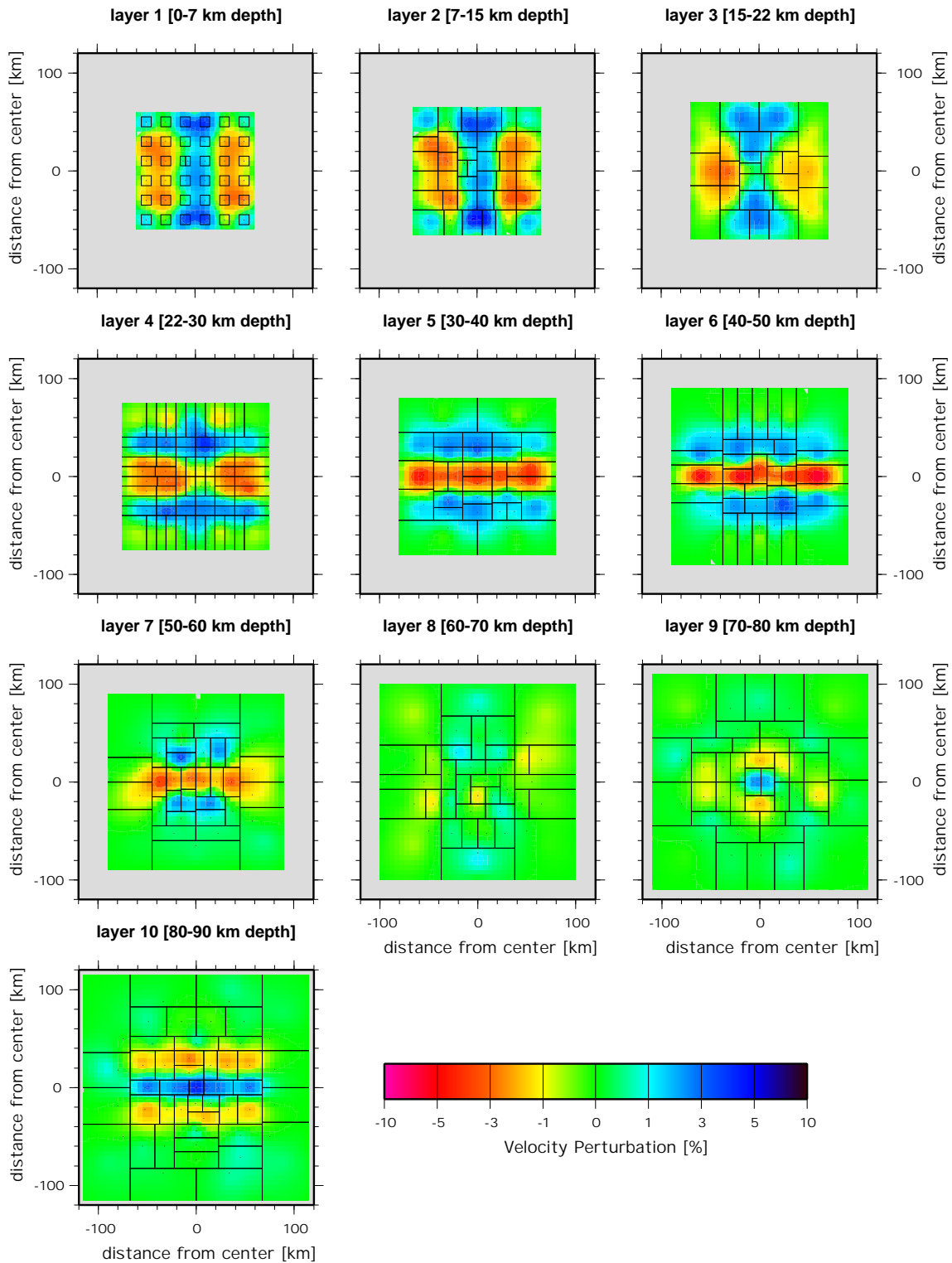
Figure 2.19: *Horizontal cross sections through the velocity model resulting from the non iterative delaytime inversion (ddv=0.12). Negative velocity perturbations are plotted in reddish and positive ones in bluish colors. The black dots and lines indicate the underlying parameterization in terms of nodes and blocks. The grey regions are outside the model.*

inherent problem of non iterative inversions of strongly contrasting anomalies.

Offset and average   As described in chapter (2.2.4) there is the possibility of applying an offset and average procedure to the inversion, which is shown here for completeness reasons. Figure 2.20 shows the respective offset and average result for the same setting as described before (ddv=0.12). The parameterization of the velocity model is arbitrary with respect to the true (unknown) velocity structure. Compared to the ordinary delaytime inversion result (figure 2.19), the offset and average result (figure 2.20) is able to image the horizontal boundaries of the recovered structures more clearly. The effect of the offset and average inversion is shown by two examples: First, the central block (0 km / 0 km) in layer 6 of the non offset and average result, is not centered on the anomaly which leads to a small bend of the recovered structure to the North. This bend is not found in the offset and average result, where the corresponding anomaly has smooth boundaries. Second, in layer 8 of the non offset and average result, no anomaly can be determined, however in the same layer of the offset and average plot, a EW-tending anomaly is discovered.

The iterative case   In the case of an iterative inversion there is an additional contribution to the inversion algorithm of the non-iterative case. The term $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ controls how far the inversion model of iteration n may deviate from the initial starting model of iteration 0, $\vec{p}_0$. Since for iteration 0, $\vec{p}_0 = \vec{p}_{n0}$, this term only contributes to iteration 1 and larger iteration numbers. (Zeyen and Achauer, 1997) state, that this term can be omitted, if non specific *a priori* information is used. This would lead to a less computational intensive, faster converging and more or less identical result. However, since their inversion scheme is not fully iterative, the influence of $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ on the inversion is examined. Figure 2.21 shows the comparison of the convergence behavior of inversions with (white symbols) and without (red symbols) the additional term for four selected ddv-values. In both cases the convergence is determined by the chosen ddv-values. Obviously, the inversion results are identical for iteration 0, since the additional term vanishes. For the first iteration there is a clear drop in the variance reductions for almost all inversions which is caused by two effects:

- The variance reduction is reduced in the case of missing $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ by the changes of the model and the corresponding raypaths. If the variances are small, then the changes of the raypaths are not too large and there is no drop of the variance reduction, as observed for ddv=0.05 and 0.06.

- In the case including $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ both effects are observed, the one mentioned above and the influence of the (here) homogeneous starting model. This leads to a large drop of the variance reductions for all variances.

The inversions without $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ converge to different models, as indicated by the different variance reductions (see also figure 2.22). Since the differences of the initial starting model and the current starting model (i.e. the inversion result of the last iteration) are incorporated in the inversion, the new model "reacts" to the behavior of the previous inversion and oscillates around the final result, until it eventually converges.
Summarized for all iterations, the inversions that neglect $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ tend to reach larger variance reductions, unreasonable large model lengths and eventually diverge, especially for large variances. In case the term $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ is considered, all inversions converge. Inversions with small initial ddv-value converge sooner than inversions with large ddv-values. Since for different ddv-values different models are obtained, a trade-off procedure has to be used to determine the optimum one. Trade-off curves for the successive iterations are shown in figure 2.23. Model
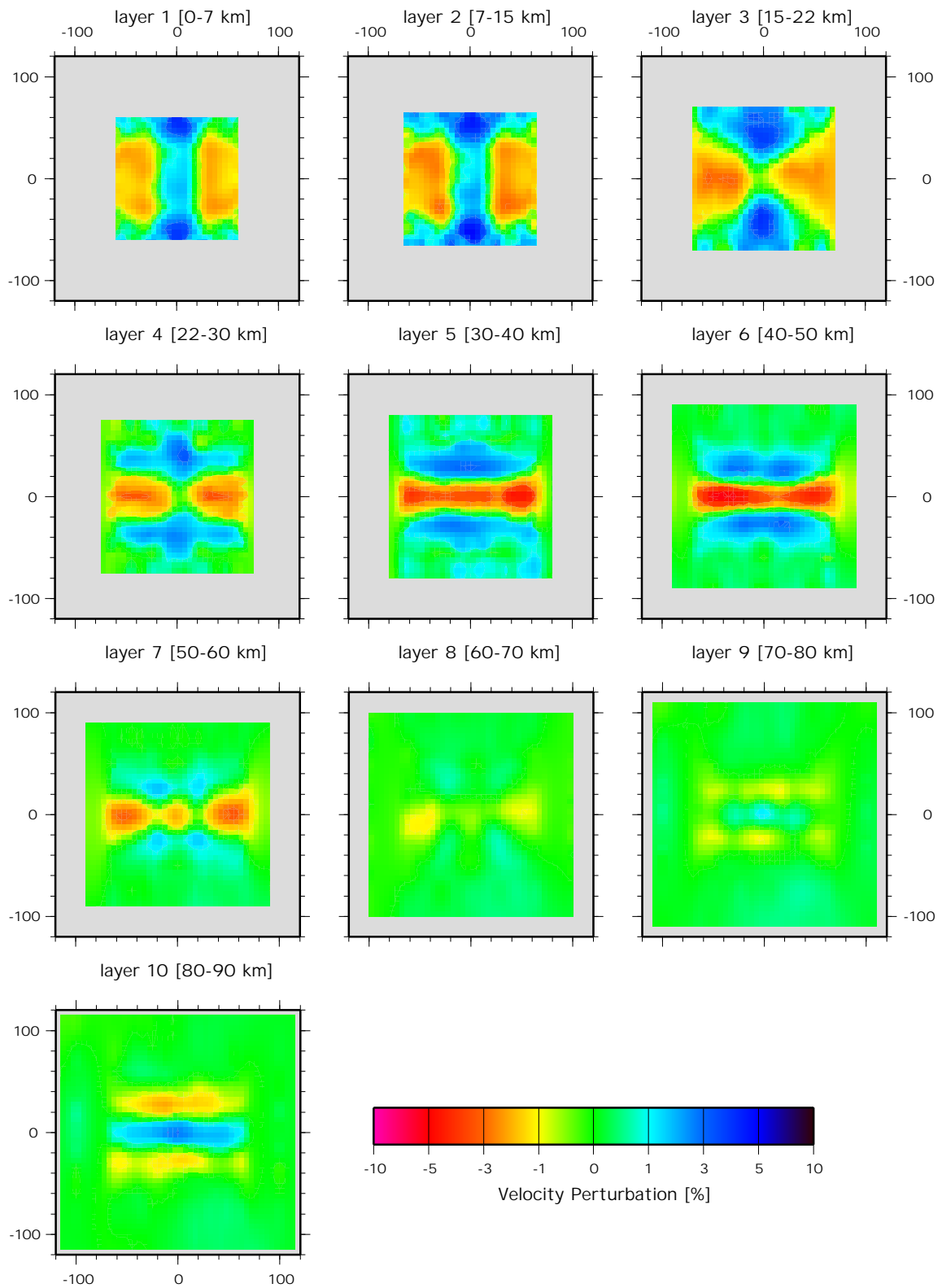
Figure 2.20: *Horizontal cross sections through the velocity model resulting from the non iterative delaytime inversion (ddv=0.12). Positive velocity perturbations are plotted in reddish and negative ones in bluish colors. An offset and average procedure is used leading to a model, more independent from the original parameterization.*
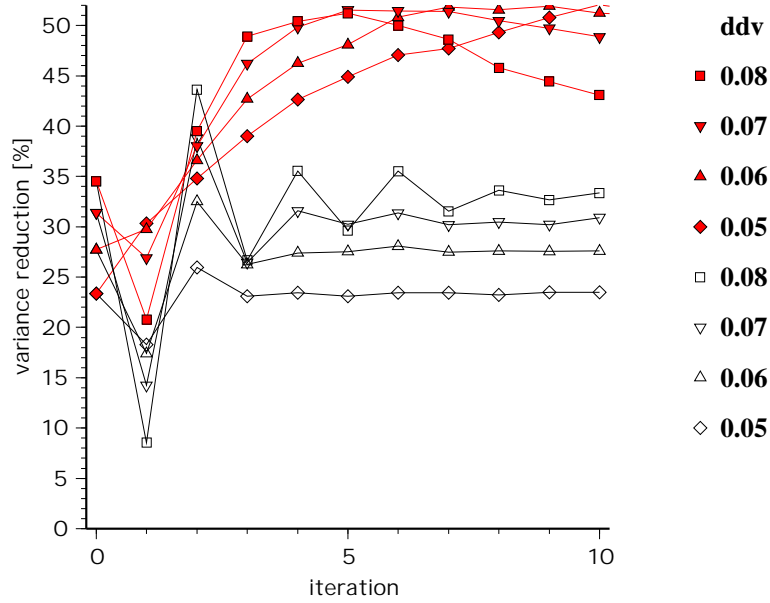
Figure 2.21: *Influence of the term $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ on the convergence behavior for different ddv-values. Each curve represents the variance reduction for successive iteration steps of a certain ddv-value. The inversions plotted in red do not include $C_p^{-1}(\vec{p}_0 - \vec{p}_{n0})$ while the unfilled symbols do.*

lengths of iteration 0 tend to become far too large and due to the "indifferent" shape of the trade-off curve, it is difficult to determine the optimum model. This works far better in later iterations. Two branches of the trade-off curves evolve, a stable one for small ddv-values and an instable one for large ones. The optimum ddv-value depends on the number of iterations used, since the stable branch grows from iteration to iteration containing larger ddv-values, until the global optimum is reached (see e.g. figure 2.23).

A suitable velocity starting model for the joint inversion has to provide a velocity structure that is close to the real one. Together with the resulting realistic ray paths this should help the joint inversion to converge quickly and to improve stability and uniqueness. Since the model is expected to change due to the influence of the gravity data, the velocity starting model should be chosen conservatively, so that joint inversion is not biased too much by the velocity results. The resulting "optimum" ddv-values is chosen to be 0.08. According to figure 2.22 the variance reduction for the ddv-value 0.08 is almost constant in iterations 8, 9 and 10. Figure 2.23 shows, that ddv=0.08 plots on the stable branch of the trade-off curves in iteration 6 and later.

Trade off for the heterogeneous velocity model   In order to determine the optimum *a priori* velocity model variance for the joint inversion, again an iterative trade-off procedure is applied. The heterogeneous velocity starting model (figure 2.24) provides "close to true" raypaths, leading to a fast converging result. This is obvious in figure 2.25. Inversions converge at higher levels and with less difference than in figure 2.25. The optimum ddv-value (0.09) is marked in red. According to the trade-off curves in figure 2.28 the optimum ddv-value (0.09) is located on the stable branch of the trade-off curves in iteration 2 and later. With respect to the heterogeneous starting model, the model length is increased from about $2\frac{km^2}{s^2}$ to about $4\frac{km^2}{s^2}$ while the variance
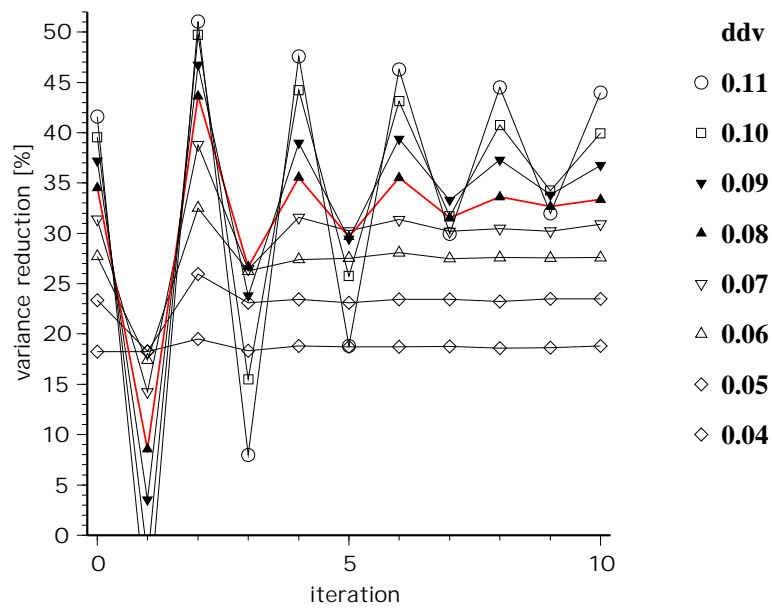
Figure 2.22: *Convergence charts for iterative inversions with ddv-values ranging from 0.04 to 0.11. Iterations converge sooner for smaller ddv-values than they do for larger ones. Since the inversions converge to different variance improvements (i.e. different models), the optimum ddv-value has to be determined by an iterative trade-off procedure (see figure 2.23). The convergence chart of the inversions that result in the starting model for the joint inversion is plotted in red.*
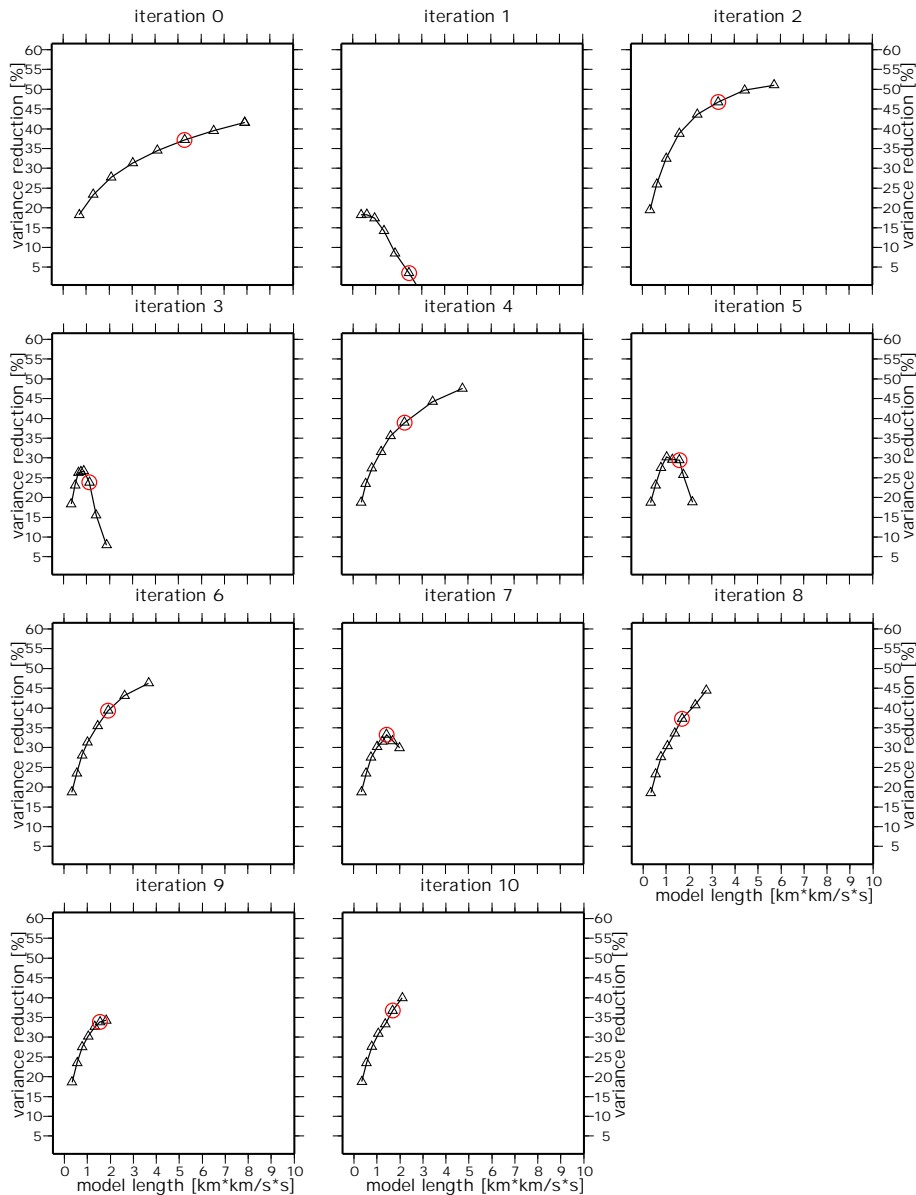
Figure 2.23: *Trade-off curves for iterations 0 to 10. Eight inversions are calculated with ddv-values of 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11. Results for small ddv-values plot towards small model lenghts. Considering all inversions, the largest model lengths are obtained in iteration 0, where the trade-off curve does not have a clear bend. In the successive iterations two branches evolve, a stable one, towards smaller model lengths and an oscillating one towards larger model lengths. As the inversions converge with larger iteration numbers, also the ddv-values on the stable branch increase. The results for the starting velocity model of the joint inversion (ddv=0.08) are indicated by red circles and plot on the stable branch in iteration 7 and later.*
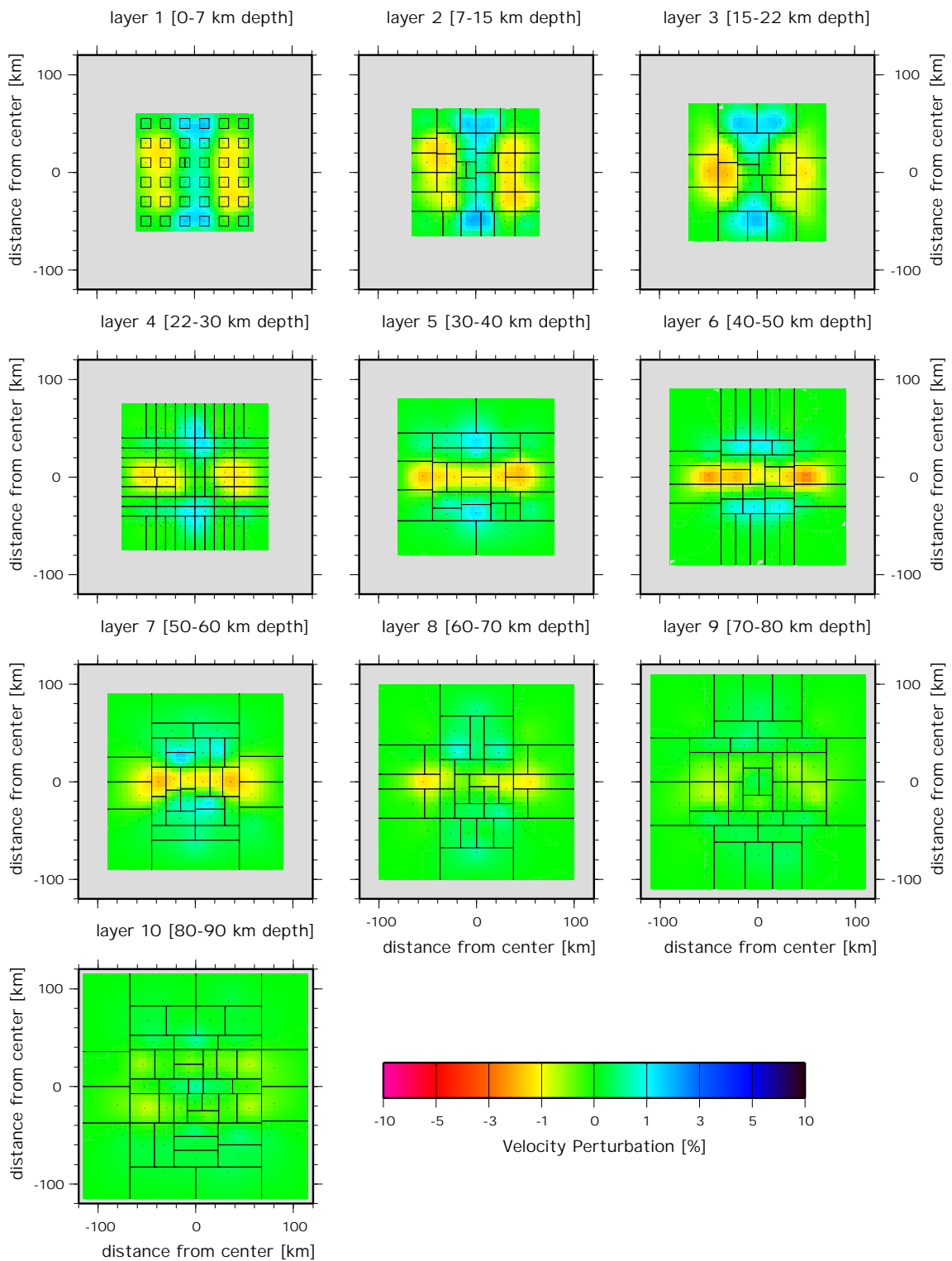
Figure 2.24: *Velocity model after 10 iterations (ddv=0.08). Since this model is intended to be the starting model for the joint inversion, a conservative ddv-value of 0.08 provides the velocity structure, realistic raypaths but not the full amplitudes.*
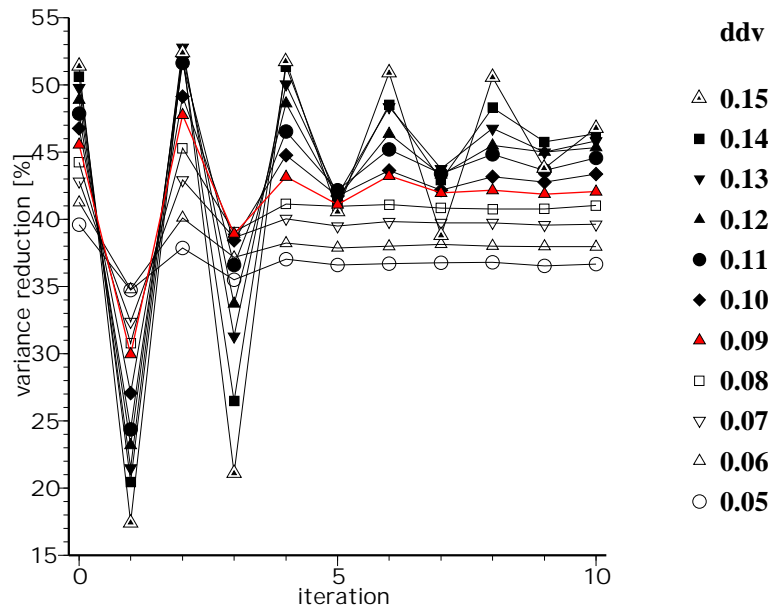
Figure 2.25: *Convergence curves for the heterogeneous velocity starting model. Ddv-values range from 0.0 to 0.15. All inversions converge, small ddv-values lead to fast convergence, large ddv inversions converge later. The optimum ddv (0.09) is indicated in red, and used as* a priori *model variance for the joint inversion.*

improvement is increased by more than 5% (see figures 2.22 and 2.25). The resulting velocity model (figure 2.27) is used for the direct evaluation of the effect and benefit of the joint inversion. Apart from the amplitudes, the velocity model resembles the starting model (2.24).
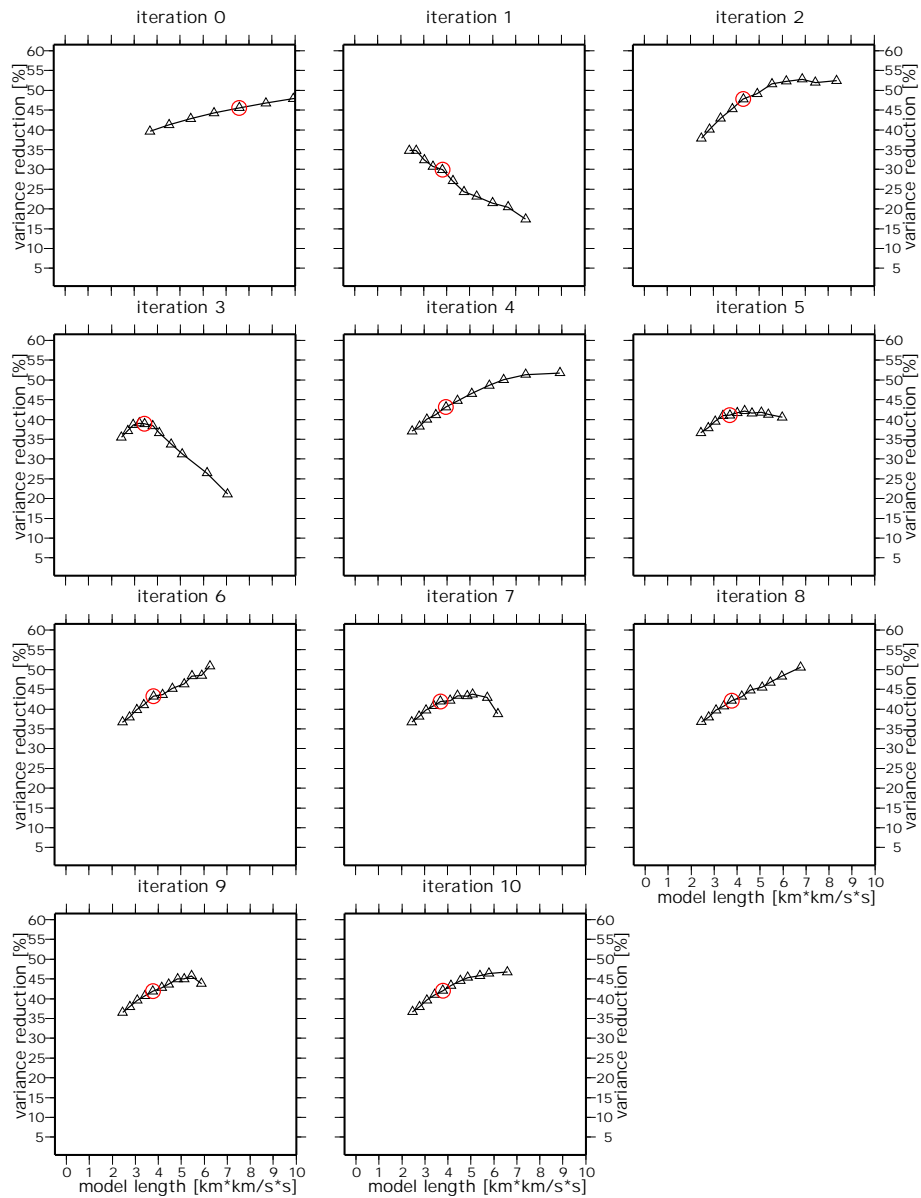
Figure 2.26: *Trade-off curves for the iterative inversions using a heterogeneous start-ing model. Ddv-values range from 0.05 to 0.15 with the optimum value (ddv=0.09) indicated by a red circle. The inversion with ddv=0.09 is located on the stable branch of trade-off curve from iteration 2 on.*
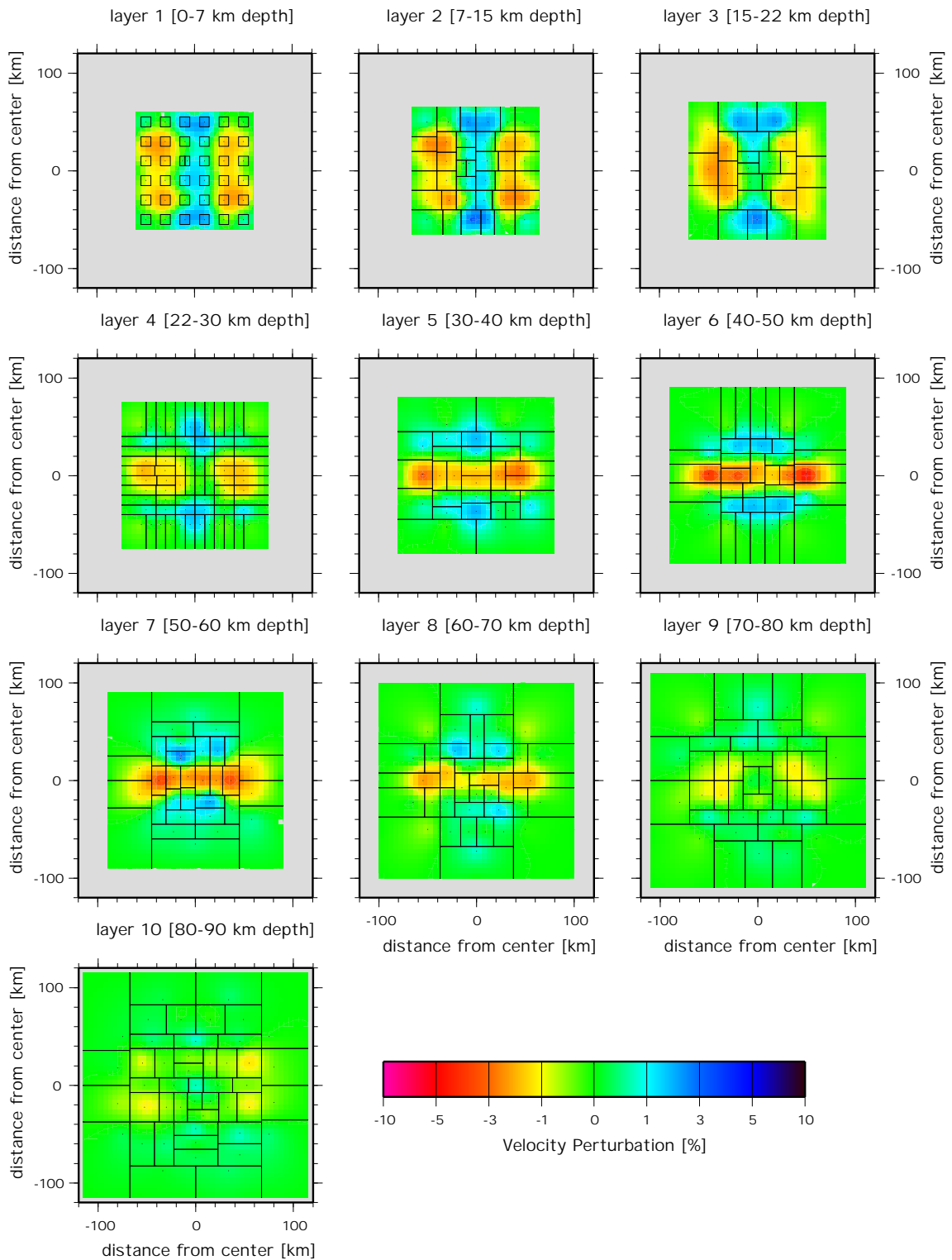
Figure 2.27: *Velocity model after 10 iterations using a heterogeneous starting model (ddv=0.09). Since both starting model and the ddv-value are used for the joint inversion this result can be used for evaluating the effect of the joint inversion.*

Bouguer data inversion

Since the inversion of Bouguer data is absolutely non-unique, it is very important to start the joint inversion with reasonable *a priori* variances for the density parameters, so that the data can be explained and the resulting model is simple. In order to determine the optimum model variances a trade-off procedure is applied, analogously to the delay time case.

However, the choice of the respective *a priori* density model variance is more obvious. The trade-off curve in figure (2.28) reflects the nonuniqueness of the Bouguer data inversion by the plateau in the trade-off curve, close to 100% variance reduction, independent of the model lengths and *a priori* variances, except for very small ones. Since one is looking for a model with reasonable amplitudes, we choose $d\rho = 0.1\frac{g}{cm^3}$ since the data are explained by more than 99% while the model length is small. This is done non iteratively, since the inversion of the gravity data is a linear problem.
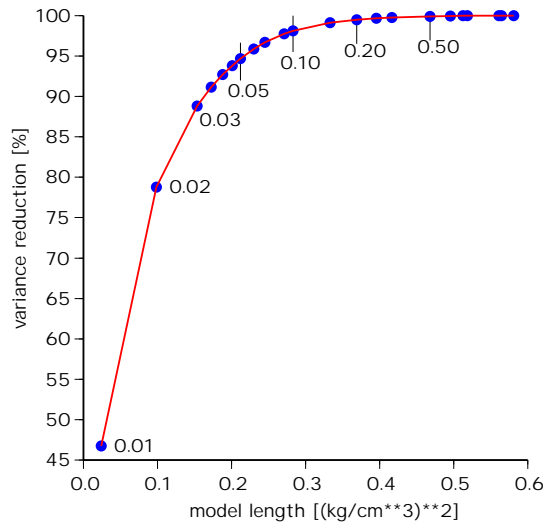

Figure 2.28: *Trade-off curve for the determination of the optimum* a priori *density variance. Since one is looking for models with large variance reduction and small model length, $d\rho = 0.1\frac{g}{cm^3}$ is chosen to be the starting $d\rho$-value for the joint inversion.*

Figure 2.29 shows the 'optimum' density model, according to the trade-off curve. However, a density model, derived from mere Bouguer gravity inversion without other constraints, is not very significant. Structure is found in all layers, which is an effect of the optimized parameterization, but meaningless for the real structure due to the nonuniqueness. For not adapted parameterizations the density variations tend to be found in the uppermost layers. Moreover, surprising little anomaly is found in layers 1 and 2 which is a realistic feature.

It is important to note, that this model is not used in the joint inversion, since the starting density model is homogeneous.
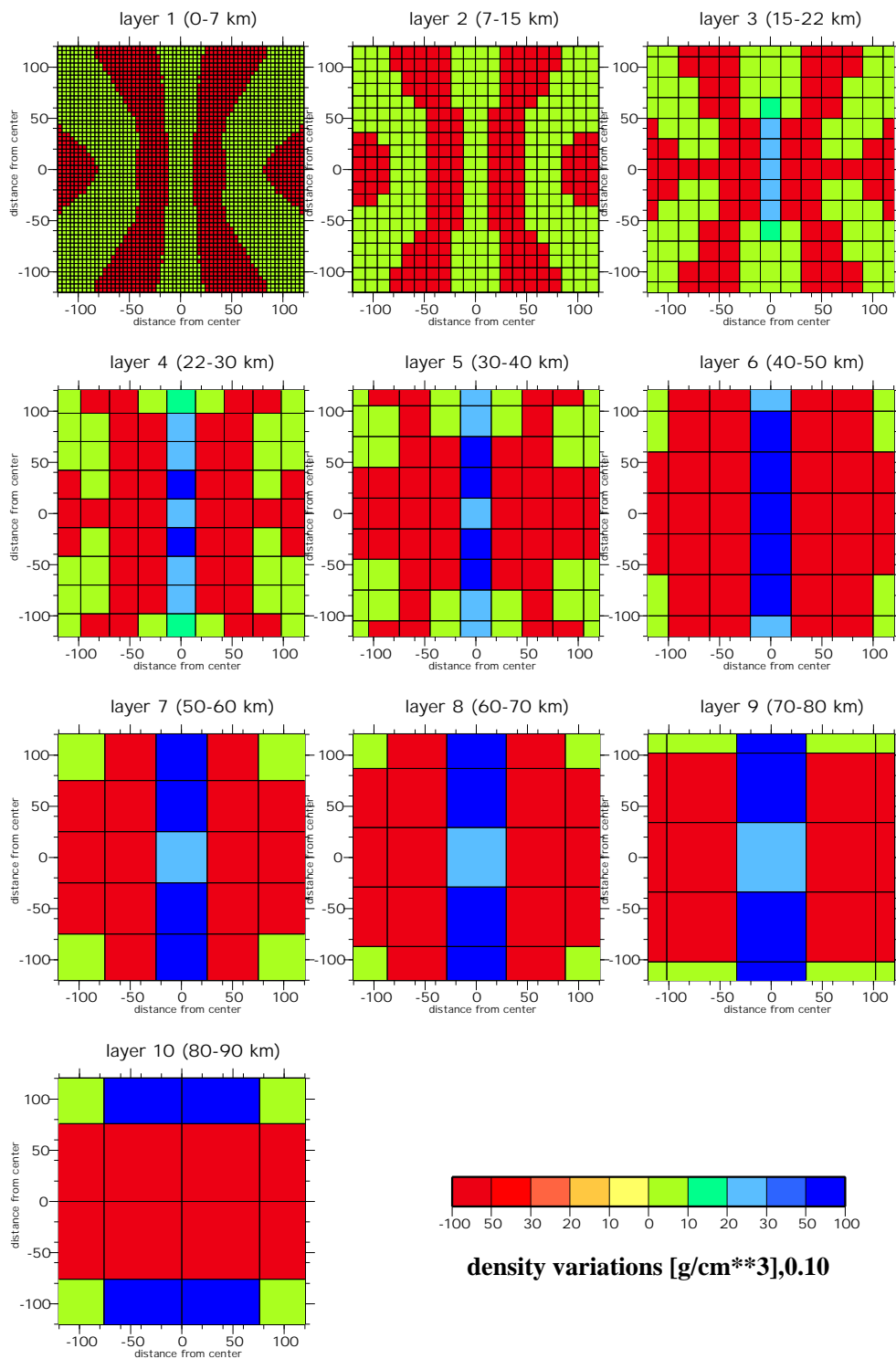
Figure 2.29: *Model of density variations, obtained by the mere inversion of Bouguer gravity data ($d\rho = 0.1 g/cm^3$). It is one of the "simplest" models that can explain the data. Nonetheless, this result is not unique.*

2.3.6 Joint inversion

By jointly inverting teleseismic delaytimes and Bouguer gravity data, models of velocity and density variations and their correlation are determined. The resulting models should be correlated and explain the respective data set. Moreover, the model lengths have to be consistent with the ones derived from the single inversions. Due to the nonuniqueness of the gravity inversion it is demanded, that the joint inversion results are mainly determined by the delaytime data.

The joint inversion depends on 6 sets of *a priori* (co-)variances usually given as vectors or matrices. Appropriate scaling of the involved quantities is vital for the joint inversion. The *a priori* variances for data and model parameters already were determined in the single inversions, providing the weighting within the individual data and model parameter sets. In order to be able to apply weighting to the importance of the individual sets of data, models and the smoothness constraint, scaling factors are introduced:

The 'scaling factors' are:

- ddt, for the data errors of the relative residual delaytimes

- ddv, for the *a priori* variances of the model parameters

- dBoug, for the Bouguer gravity data errors

- $d\rho$, for the *a priori* variances of the density model parameters

- dB, for the *a priori* variances of the correlation parameters

- dS, for the smoothness of the density model

These factors are dimensionless and are also referred to as inversion parameters, which have to be carefully determined in order to obtain optimum inversion results, since the inversion problem is strongly nonlinear. Note, that the weighting of the *a priori* variances within a set of data (e.g. errors) or model parameters (e.g specific *a priori* information) are not affected by the scaling.

Hence, the straightforward proceedings would be to fix the *a priori* variances of the delaytime data and the velocity model parameters as derived by the iterative single delaytime inversion and to vary the remaining 4 inversion parameters accordingly, so that the criteria for optimum models are met. However, this is computational absolutely prohibitive and a different proceeding is suggested: As shown in the previous sections, the *a priori* variances of both data sets and models can be derived from single (iterative) inversions. The variances for the gravity data and the density model have to be considered as preliminary since the optimum scaling between the data sets is not known and may have to be adjusted.

In a next step, the variance for the smoothing of the density model is determined. Since the lateral extension of the density model exceeds the one of the velocity model by far, many density blocks are not constrained by velocity nodes. Due to the nonuniqueness of the gravity inversion, these density model parameters have to be damped, which is done by a smoothness condition, as shown in equation 2.23. Without smoothing, the gravity signal, that cannot be explained by the correlated density model, is projected to the unconstrained region of the model, leading to unrealistic, often oscillating results. In case the smoothing is too strong, any density structure, which is not constrained by the velocity model cannot be recovered by the inversion.

Hence, the proceeding is as follows: The smoothing parameter dS is determined for the case of a very weak coupling (i.e. the correlation between the individual velocity and density model parameters may vary freely) by trial and error. Starting from small dS-values (i.e. strong smoothing)

proceeding to larger ones, the optimum is found if the joint inversion is able to reproduce the single inversion results in the noniterative case.

Then dS is fixed and the dB inversion parameter is changed together with $d\rho$, trying to find an optimum combination. After that, the scaling of the Bouguer gravity data dBoug is varied around the previously fixed, to examine if the applied scaling is the correct one.

Application to the full data set

Determination of variances and scaling   The optimum dS is determined as described above for ddv=0.09, $d\rho$=0.10 and dB=10.0 (totally decoupled velocity and density models). The corresponding optimum dS-value is dS=0.0001, which is fixed during the further inversions. The inversion results resemble those of the non joint inversions (figures 2.27 and 2.29) closely.

There are many possible strategies for the determination of the remaining inversion parameters dB, $d\rho$ and dBoug with affordable effort. Two possibilities, leading to different sets of inversion parameters but very similar and consistent results, are presented here:

First, as suggested above, dB and $d\rho$ are varied in a 2D trade-off scheme to find an optimum combination. Variance reductions and model lengths are derived and plotted (figures 2.30 and 2.31). In figure 2.30, the normalized variance improvements for both data sets are plotted with blue colors indicating good variance reduction.
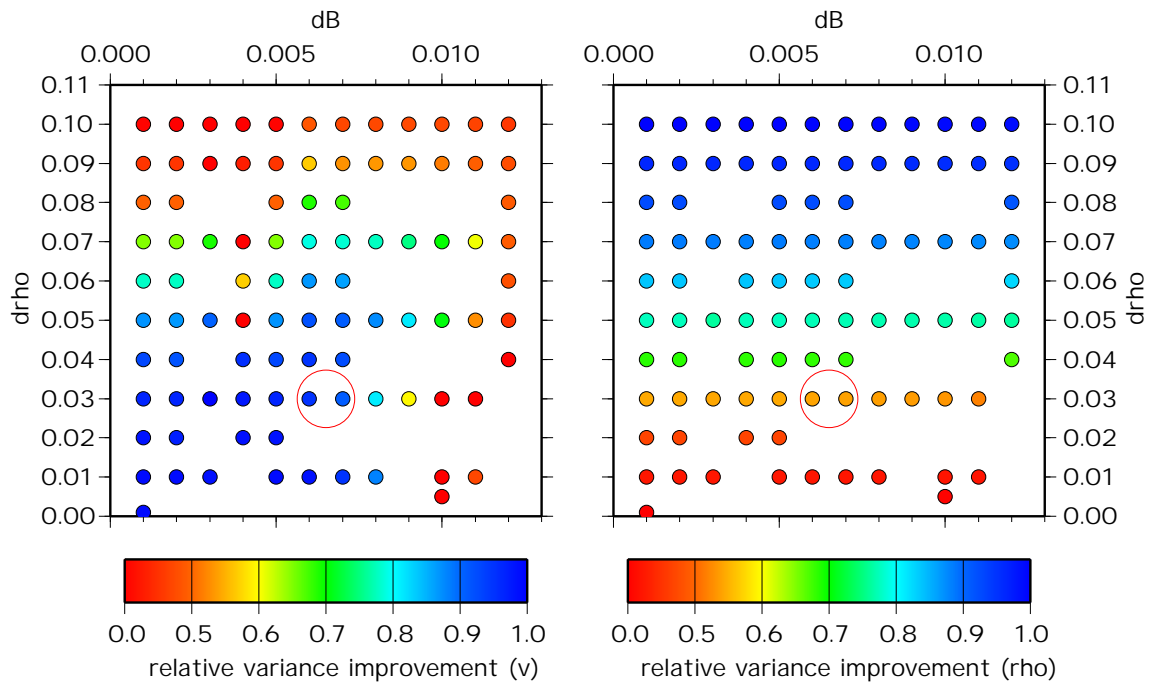

Figure 2.30: *Plots of the variance improvements of inversions using different dB-dρ combinations. dB and dρ (drho) are the dimensionless scaling parameters. Blue colors indicate high, red colors indicate low variance reductions. The variance improvements for both, the delaytime inversions (left) and the Bouguer gravity inversions (right) are normalized and the values regarded as optimum are indicated by a red circle.*

To evaluate both models at the same time, also the product of both normalized variance reduction plots is given in figure 2.32. Consequently, the points in blue indicate combinations of dB and $d\rho$
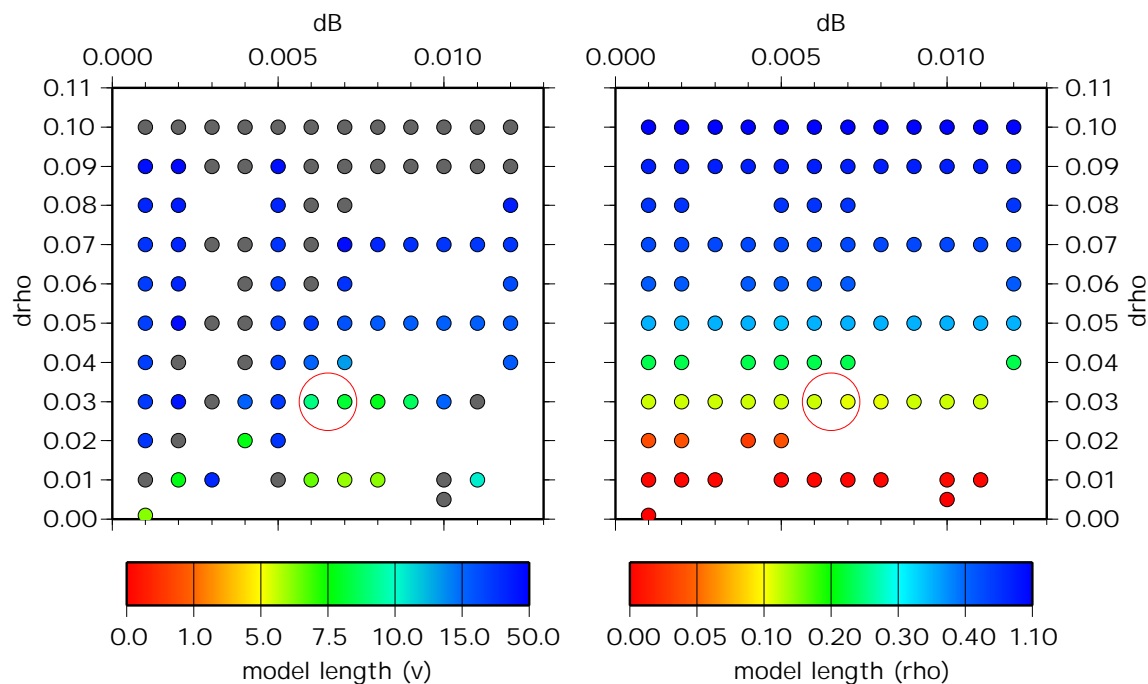
Figure 2.31: *Plots of the model lengths of the velocity models (left) and the density models (right) are displayed. dB and dρ (drho) are the dimensionless scaling parameters. The model length is are given as squared model lengths. The units are $\frac{km^2}{s^2}$ and $\frac{kg^2}{dm^6}$. The 'optimum' values which are consistent with the non-joint inversion results are plotted in green. The red circle indicates the optimum combination of the dB and dρ inversion parameters.*

where both data sets are reasonably explained. Optimum variance improvements of the delaytime data are obtained for dB-values smaller than 0.008 and $d\rho$-values smaller than 0.05 (figure 2.30). The Bouguer data residual variances are least for large $d\rho$ values (e.g. 0.10). This is an effect of the nonuniqueness of the gravity inversion, since any signal can be explained if the model parameters may vary sufficiently (i.e. large $d\rho$ values).

The optimum combination in terms of common variance improvement (figure 2.32) are $d\rho$ values between 0.03 and 0.09 and dB values ranging from 0.001 to 0.01 approximately. It is obvious, that there is no 'smooth' maximum of the common variance reduction but that it is interrupted by parameter combinations, affected probably by nonlinearities, leading to low variance improvements. The number of possible combinations of $d\rho$ and dB values is further reduced by the condition, that the model length of the resulting models have to be within reasonable limits. As reference for suitable model lengths we use the ones derived during the single delaytime and gravity inversions. Figure 2.31 shows the lengths of the velocity and density models for several parameter combinations. Realistic (squared) model lengths appear as yellow to lightblue colors, corresponding to $5 - 10 \frac{km^2}{s^2}$ and $0.1 - 0.3 \frac{kg^2}{dm^6}$. 'Optimum' values are plotted in green.

The reduced number of tested possible combinations provides $d\rho = 0.03$ and $dB = 0.006$ or $dB = 0.007$ as optimum combinations, indicated by red circles. The variation of dBoug and other parameters around the current settings does not change the results significantly, so that this set of inversion parameters is assumed as the basis for an iterative joint inversion.

As a second possibility, the desired coupling strength (dB) is preselected (from the previous con-

siderations dB=0.007 and dS=0.0001 are chosen) and dBoug is varied. Ddv and $d\rho$ are fixed, using the values optimum for the non joint inversions, 0.09 and 0.10 respectively. The results of the variation of dBoug in terms of variance improvements and model lengths are shown in figure 2.33. Optimum results are obtained for dBoug=0.25, leading to joint inversion results that resemble the ones obtained above very closely, with slightly increased amplitudes, model lengths and improved variance reductions.

A complete summary of the parameters is given in table 2.3:

ddv	$d\rho$	dB	dS	dBoug	VI(v)	ML(v)	VI(ρ)	ML(ρ)
0.09	0.03	0.007	0.0001	0.10	37.8%	$7.63\frac{km^2}{s^2}$	19.9%	$0.04\frac{kg^2}{dm^6}$
0.09	0.10	0.007	0.0001	0.25	36.2%	$8.17\frac{km^2}{s^2}$	27.2%	$0.09\frac{kg^2}{dm^6}$

Table 2.3: *Comparison of the variance improvements (VI) and the model lengths (ML) for the velocity and density models derived for two different sets of inversion parameters (ddv, dρ, dB, dS and dBoug).*
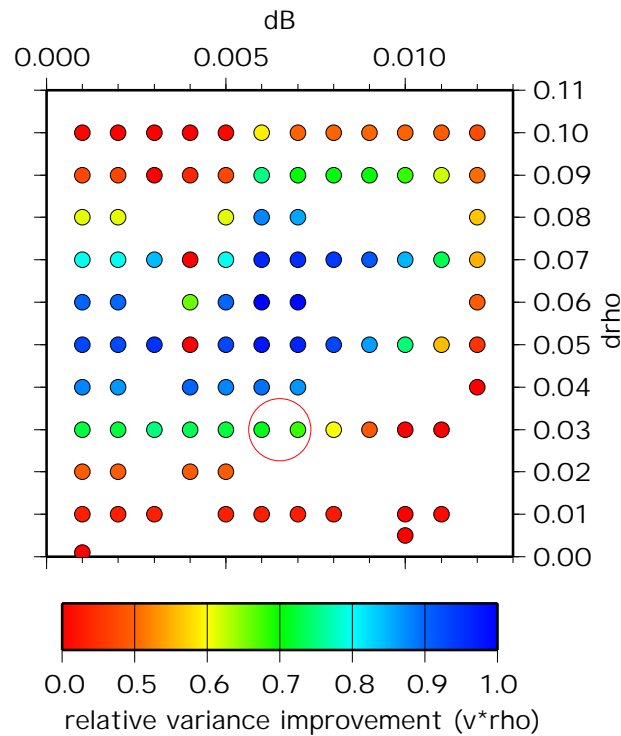
Figure 2.32: *Products of the normalized variance improvements for the delaytimes and the gravity data. Blue colors indicate high common variance reductions, red ones low common variance reductions. The optimum parameters combination is indicated by a red circle.*

Inversion results The inversion results for the non iterative joint inversion are presented in figure 2.34 and 2.35.

Since the inversion results for the two inversion parameter sets resemble each other closely, only the ones for the latter case are presented and evaluated. The results are compared to the synthetic input model 2.11 and the non joint inversion results, given in figures 2.27 and 2.29.

The NS-tending positive velocity structure is found continuous in layers 1 to 4 while the signal in layers 1 and 2 is an effect of vertical smearing, due to the unfavorable near surface ray distrubution. The structures found in layers 3 and 4 are correct. In the pure delaytime case (figure 2.27), the structure also appears in layers 1 to 3 but merely disrupted in layer 4. In both, the joint inversion and the pure delaytime inversion vertical smearing affects layers 5 and 6 which should contain no structure according to the synthetic input model.

While in the pure delaytime inversion result the deeper negative EW-tending anomaly is smeared into layers 5 and 6, in the joint inversion result contributions from both, the upper and lower structure are visible. In the original input structure, this anomaly is located in layers 7 and 8, which is recovered correctly by the pure delaytime inversion. However, in the joint inversion result it only appears in layer 7 while in layer 8 it is strongly disturbed by a strong positive signal that is induced by the coupling to the density model, which is shown later. Layers 9 and 10 should contain no anomaly. This is recovered to a reasonable extent by the pure delaytime inversion while the joint inversion velocity model reflects the density model with strong positive and negative anomalies.

The original density model as derived by the non joint inversion shows anomalies in all layers which are reasonable with respect to the amplitudes, but do not reflect the synthetic input structure
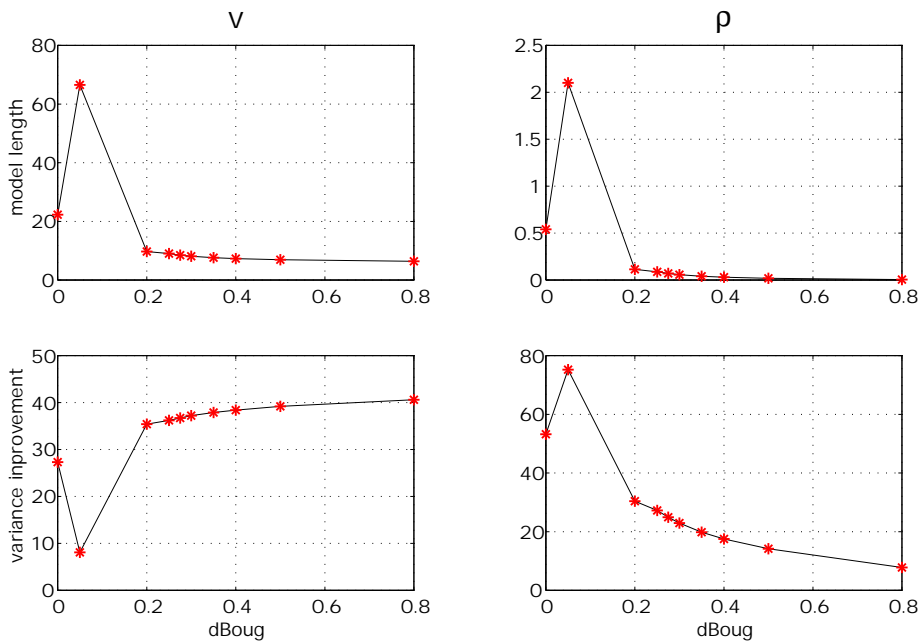
Figure 2.33: *Model lengths (top) and variance reductions (bottom) are displayed of the velocity (left) and density models (right) for different dBoug-values.*

(figure 2.11) too much. In the joint inversion result (2.35), layers 1 and 2 are homogeneous which probably is an effect of the smoothness constraint, since many density blocks do not spacially coincide with velocity nodes. Layer 3 shows a NS-tending positive density anomaly which is too low in amplitude, while layer 4 only recovers a single high density block, surrounded by low density parameters. Similar results are observed for layers 5 to 10 which do not reflect the true structure, but rather affects the velocity model. This is believed to be an effect of the density block parameterization. If the density blocks are larger than the velocity blocks or if they overlap several velocity blocks a simple correlation between the models is difficult. The block structure of the density model is well visible in the resulting velocity model.
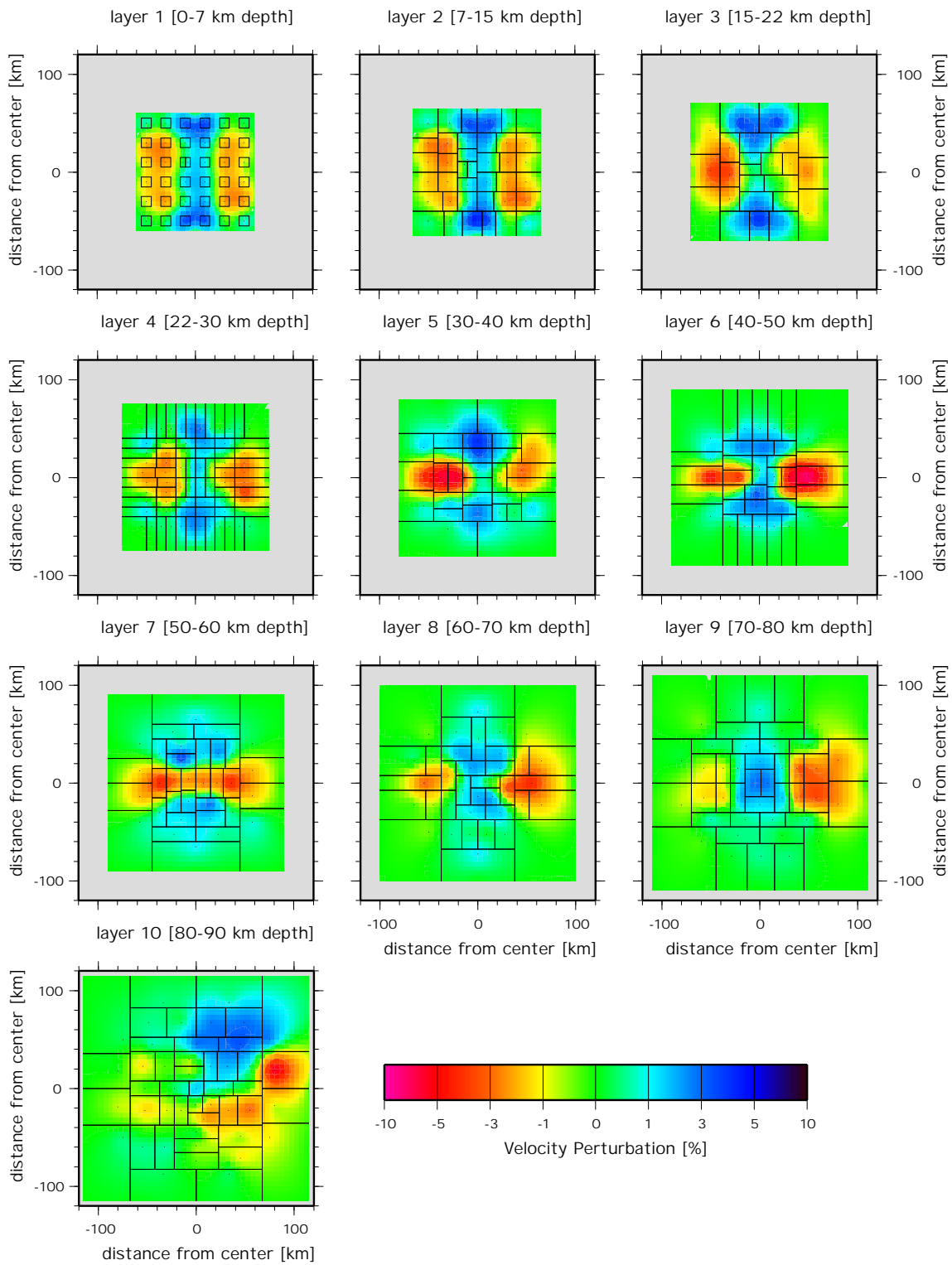
Figure 2.34: *Velocity model resulting from the non iterative joint inversion. The scaling factor used are ddv=0.09 dρ=0.10, dB=0.007, dS=0.0001 and dBoug=0.25.*
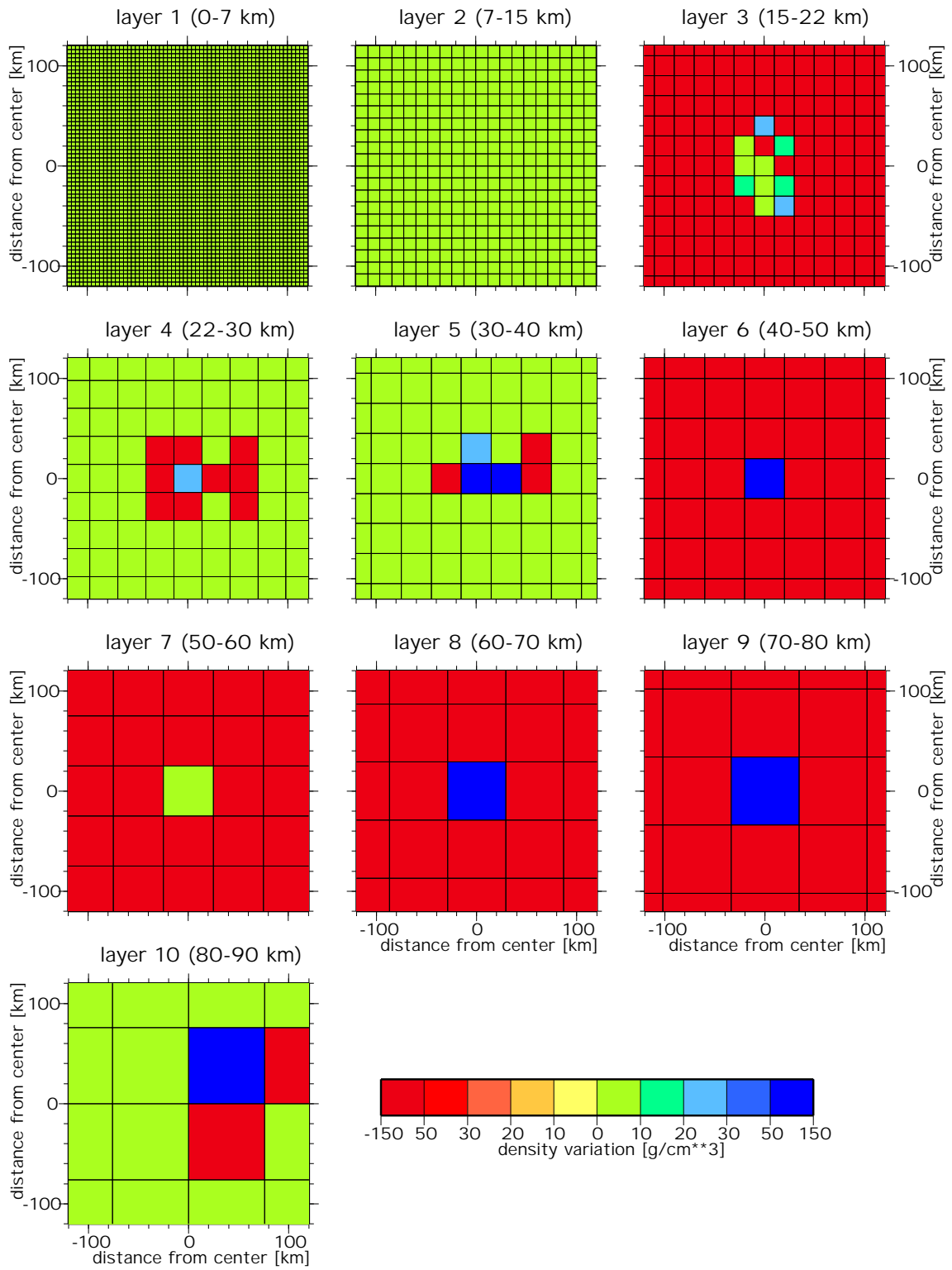
Figure 2.35: *Density model resulting from the non iterative joint inversion. The scaling factor used are ddv=0.09 dρ=0.10, dB=0.007, dS=0.0001 and dBoug=0.25.*

Summarizing the results, the upper part of the model space (i.e. layer 1-7 for the velocity model and 1-3 for the density model) can be reasonably resolved in the joint inversion, slightly better than in the case of the single inversions. However, in the (deeper) rest of the models, the relevant structures could not been resolved and the results are worse than for the single inversions in these regions. Moreover, the blockstructure of the density model is mapped into the velocity result.

This is mainly an effect of the influence of the density block sizes on the correlation between density and velocity model parameters. The velocity model parameters are defined on nodes whereas the density model parameters are defined on blocks. If each density block is constrained by a velocity node, there is a simple one to one correlation. In case the density blocks are smaller than the spacing of the velocity nodes, several density blocks will not be 'hit' by a velocity node. However, the respective density parameters are not unconstrained, but coupled to the rest of the density model by the demanded smoothness of the density model. The density-velocity correlation is built for the spacially coinciding parameters. If the density blocks are much larger than the spacing of the velocity nodes, the blocks may contain several nodes with different velocity perturbations that have to correlate to the same density parameter. This leads to the effects observed in the deeper layers of the velocity and density models, that a reasonable correlation cannot be realized and that the density blocks map into the much more densely spaced velocity model. As a consequence, density blocks larger than the minimum spacing of the velocity nodes should be avoided.

In addition to that, the strong vertical smearing into the uppermost layers of the velocity model could not be reduced by the joint inversion at all, which is not surprising, since the ray distribution is the same as for the pure delaytime case and a minimum depth cannot be determined from gravity data. Therefore the uppermost parts of the model space should be constrained by additional *a priori* information, which is easily available for many regions of the earth.

A real inversion for the correlation parameter B cannot be realized since B has to be fixed during the inversions via small dB-values. This is also observed by Tiberi, (2003, pers.comm.) who uses the same algorithm but a different realization. On the one hand this is an effect of the difficult correlation due to large density block sizes, on the other hand it is an effect that this joint inversion algorithm depends on many unknown parameters which are at least partially interdependent.

Being aware of avoiding too large density blocks, one has to keep the block sizes extremely small in the uppermost layers (here 4 km sidelength is used in the first layer). This, in turn, leads to an enormous number of density model parameters in the inversion, causing computational problems. As the inversion results suggest, the block sizes still are too large to obtain reasonable joint inversion results.

Several approaches can be considered to improve the situation:

- Depth depending coupling of density and velocity model (also depending on block sizes)

- Application of identical parameterizations to the velocity and density model. This is similar to the parameterization of Zeyen and Achauer (1997) who use constant block sizes per layer. Applied to the variable parameterization this means that the density model would be equal to the block structure of the velocity model, however laterally extended. The very negative consequence would be, that many small density parameters in the central part of the model would in fact not contribute to the joint inversion since the effect of very small volumes on the forward calculated gravity data is tiny, in contrast to the larger blocks which would dominate the density inversion.

- Application of the joint inversion only to the crustal depths by filtering the Bouguer data and applying the joint inversion only to selected layers of the model.

- Elimination of the uncontrolled crustal vertical smearing by applying specific *a priori* information about the respective model parameters.

Additional constraints

In order to improve the joint inversion results the space of possible solutions is further reduced by introducing additional constraints.

As suggested above, specific *a priori* information is applied to layers 1 and 2, from the surface to 15 km depth. In real applications, information about the velocity and density structure in near surface regions are widely available, e.g. from geological maps, reflection- and refraction seismic studies. Thus, information about the part of the velocity model, that is most difficult to resolve independently (i.e. the region directly beneath the stations), is incorporated into the model and fixed by small *a priori* model variances. The same applies to layers 1 and 2 of the density model. In this synthetic application, it is assumed, that the uppermost two layers are homogeneous.

The results of the iterative pure delaytime inversion are fair to good for the subcrustal part of the model space. In the crust, however, additional information is needed to improve the inversion results in this region. Therefore, the Bouguer gravity data are filtered, so that the remaining signal only may be caused by anomalies, located within the upper 30 km of the model space (i.e. the crust). According to Jung (1960), the maximum wave length of a density anomaly situated within 30 km depth is approximately 90 km. Hence, a highpass filter with corner and cut-off frequencies corresponding to wave lengths of 90 km and 100 km, respectively, is applied.
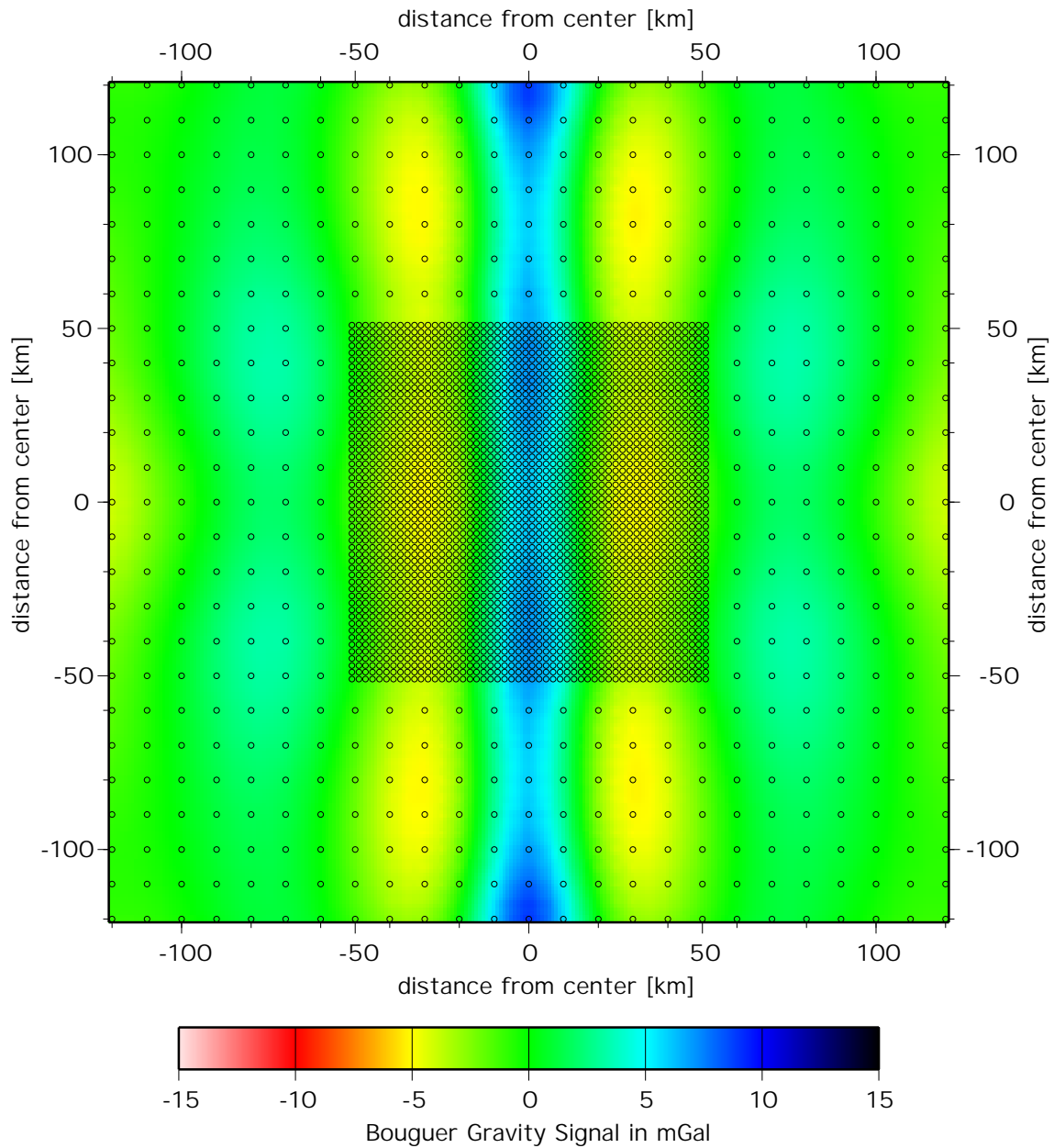
Figure 2.36: *Bouguer signal derived from the 'observed' one (figure 2.13) by high pass filtering. The remaining signal is assumed to be caused exclusively by density structure located no deeper than 31 km.*

Figure 2.36 shows the filtered Bouguer data set. The signal reflects the NS-tending anomaly of increased density of the synthetic input model (2.11), located in the depths range between 15 km and 30 km. The filtered amplitude is 15 mGal instead of 30 mGal in the original data set. This probably is an effect of the rather conservative choice of the corner wave length, since Jung (1960) considered the maximum realistic density contrasts in the earth for the derivation of his rule of thumb, which is too large in this case. In addition to that, it has to be mentioned, that highpass filtering of the Bouguer data results in signal reflecting structure located above a certain depth, but that does at no means guarantee, that all structure in that depth range can be recovered in the filtered data. Hence one has to be aware, that by filtering one may lose desired information.

Using *a priori* information about layers 1 and 2 and filtered Bouguer data, containing only information about layers 1 to 4, leads to the fact, that joint inversion actually is only applied in layers 3 and 4.

Therefore, the parameterizations of the density- and B-model are adapted as given in table 2.4. On the left hand side, the parameterization of the full model as used in section 2.3.6 are given with the numbers of the velocity and density model parameters and the *a priori* variances of the correlation parameter B (dB). Since layer 1 and two are fixed in the velocity and density model due to the *a priori* information, the reduced density model, as shown on the right hand side of table 2.4, only contains 1 parameter per layer, reducing the original number of 4772 parameters to 2. Since the number of velocity parameters in layers 1 and 2 are small, the parameterization is left unchanged. Since in layers 5 to 10, only a teleseismic inversion is performed, the number of density parameters is also set to 1 per layer. Accordingly, the *a priori* variances of the B parameter is set to 10.0 in layers 1 and 2 and 5 to 10, resulting in totally decoupled velocity and density model parameters in those layers.

layer n	full model			reduced model		
	v_{all}	ρ_{all}	B_{all}	v_{part}	ρ_{part}	B_{part}
1	37	4156	0,007	37	1	10,0
2	38	616	0,007	38	1	10,0
3	24	373	0,007	24	373	0,007
4	72	189	0,007	72	189	0,007
5	29	257	0,007	29	1	10,0
6	35	201	0,007	35	1	10,0
7	33	157	0,007	33	1	10,0
8	36	105	0,007	36	1	10,0
9	42	137	0,007	42	1	10,0
10	46	116	0,007	46	1	10,0

Table 2.4: Comparison of the initial full and the reduced models. Since joint inversion is only applied to layers 3 and 4, the other layers are decoupled by dB=10.0, and the number of density parameters can be reduced greatly. The columns give the numbers of parameters used in the velocity and density model and the respective *a priori* variances for the coupling between the models in the individual layers.

After all, this also results in a greatly decreased number of model parameters leading to less computational problems.

The heterogeneous starting model for the joint inversion   Analogously to the proceedings of the previous joint inversion a heterogeneous velocity starting model and the corresponding *a priori* model variances for the velocity and density models have to be derived first.

Therefore, layers 1 and 2 of the homogeneous model, described in paragraph 2.3.3 are fixed by very small *a priori* model variances, so that the velocity model parameters are not allowed to vary. The optimum *a priori* model variances for the rest of the model is determined as ddv=0.09 in an iterative trade-off procedure. The resulting velocity model is the heterogeneous starting model for the joint inversion. The optimum *a priori* model variances for the now derived starting model of the joint inversion are determined in the same way, leading to an optimum ddv-value of 0.10. The corresponding velocity result is shown in figure 2.38 as the iterative inversion result of the pure delay time data inversion. This is used to evaluate the resulting velocity model of the joint inversion. The variance improvement achieved here is 42.2% and the corresponding (squared) model length is $4.23 km^2/s^2$. The result is evaluated together with the corresponding joint inversion result.
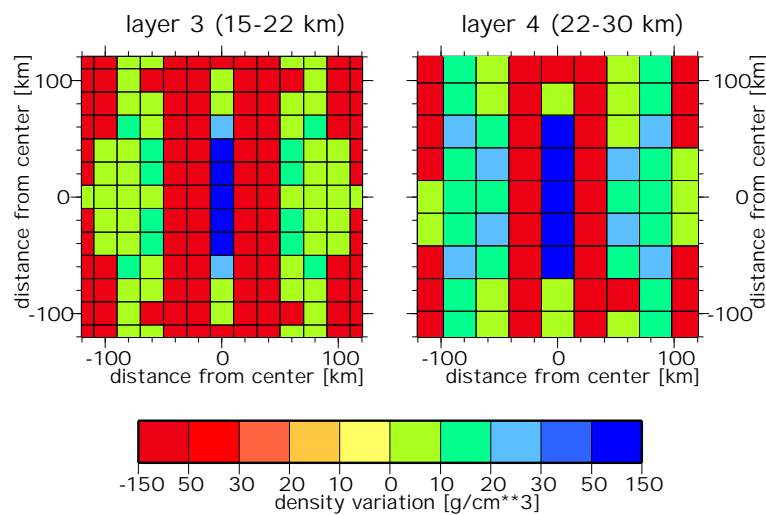

Figure 2.37: *Density model obtained from the non iterative inversion of the filtered Bouguer data (2.36).*

The optimum *a priori* model variances for the density model parameters of layers 3 and 4 are also determined in a trade-off procedure. Inverting the filtered gravity data set, using the optimum *a priori* density variance of $d\rho = 0.7$, results in the model shown in figure 2.37. These two layers are the reference for the density model derived by the joint inversion. The variance improvement for this inversion is 96.5% and the (squared) model length is $0.15 \frac{kg^2}{dm^6}$. Also this result is evaluated together with the corresponding joint inversion result.
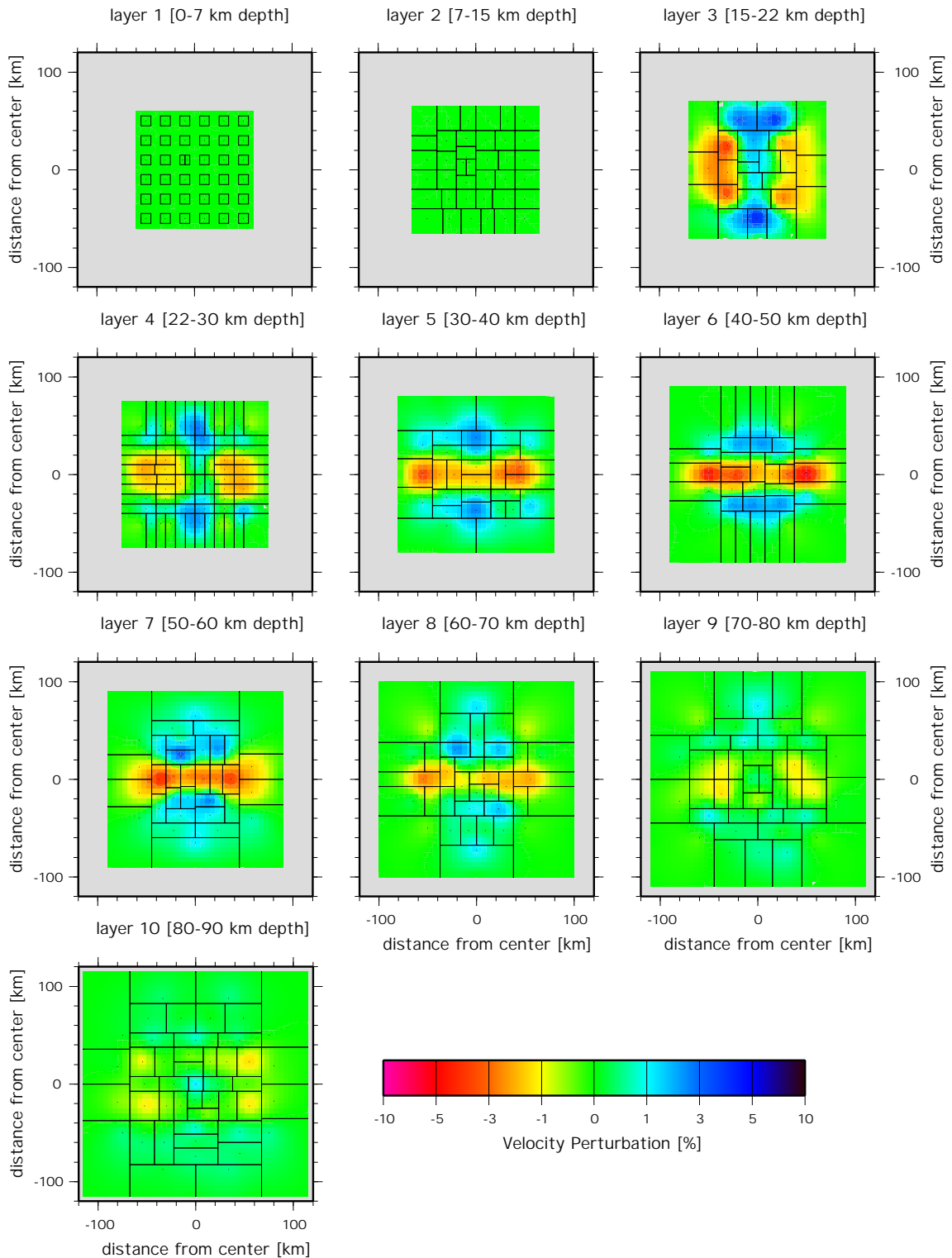
Figure 2.38: *Velocity model after 10 iterations of pure delaytime inversion (ddv=0.10), with* a priori *information. The same heterogeneous starting model as for the joint inversion, later in this chapter, is used. Therefore the results are directly comparable to the velocity model obtained from the joint inversion.*

The non iterative joint inversion   In analogy to the previous section, the smoothing factor is determined as dS=0.0001. This inversion parameter is fixed during the further proceedings and the optimum combination of dB and dρ, the scaling factors for the coupling strength and the density model parameters are derived. This is done by searching a large number of possible combinations in a trial and error manner, using non iterative joint inversions. The resulting common variance reduction, and model lengths for the velocity and density models are presented in figures 2.39, 2.40 and 2.41, respectively. Each point corresponds to the inversion results for one specific combination of dB and dρ. The optimum combination of variance improvements and model lengths is determined in a 2D-trade-off fashion. The range of dB and dρ values searched comprises 10^{-5} to 10.0 and 0.01 to 1.00, respectively. As explained before, the criteria for such an optimum combination are a high common variance reduction (blue colors in figure 2.39), and model lengths that are consistent with the ones obtained for the non-joint inversions. These are $0.15 \frac{kg^2}{dm^6}$ for the result of the gravity- and $4.23 \frac{km^2}{s^2}$ for the delay time inversion. Starting with figure 2.39, two main regions providing a large common variance improvement are found:

- dB ranges from 0.7 to 10.0 and dρ from 0.15 to 1.0. This corresponds to the case, where density and velocity model are not or only weakly coupled, leading to inversion results as presented in figures 2.38 and 2.37. For smaller dρ values, the variance reduction of the delaytime inversion still is good. However, due to the small allowed amplitudes in the density model the gravity data cannot be explained to a large degree.

- dB ranges from 10^{-5} to 0.05 and dρ from 0.5 to 1.0. In this region, the coupling has 'saturated', i.e. there is no more change for smaller dB values. In this dB range, dρ values between 0.5 and 1.0 provide models that can explain the data to a very large degree.

Proceeding to figure 2.40, in the respective regions, the model lengths of the velocity models are more or less constant and in the correct range, with only few exceptions. This is a strong hint, that the joint inversion is determined to a very large degree by the delay time inversion.
The model lengths of the density models shown in figure 2.41 show a quite different behavior. Similarly to the common variances, there are also two regions that provide results which satisfy the criteria, defined above:

- dB ranges from 1.0 to 10.0 and dρ from 0.5 to 1.0. This corresponds to the decoupled case, as already described above.

- dB ranges from 10^{-5} to 0.4 and dρ from 0.5 to 0.7, showing one exception, which can only be explained by nonlinearities.

Merging the results, two regions of inversion parameter combinations remain:
First, dB > 1.0 and dρ > 0.5 represent the decoupled joint inversions that are not relevant here. Second, dρ values between 0.5 and 0.7 and dB values between 10^{-5} to 0.05 comprise a set of inversions that do not differ significantly in terms of the resulting velocity and density models and the corresponding model lengths and variance reductions. The correlation parameters between density and velocity model parameters are fixed and the dρ value is very similar to the one determined from the Bouguer gravity trade-off. The relatively large region of stable and consistent inversion results at least partially is due to the good correlation between velocity and density model, which is enabled by the small density blocks in layers 3 and 4, compared to the nodal distances in the velocity model (see table 6.1).
Variations of the scaling factor for the influence of the gravity data were tested in the non iterative joint inversion case and no significant improvements could be found.
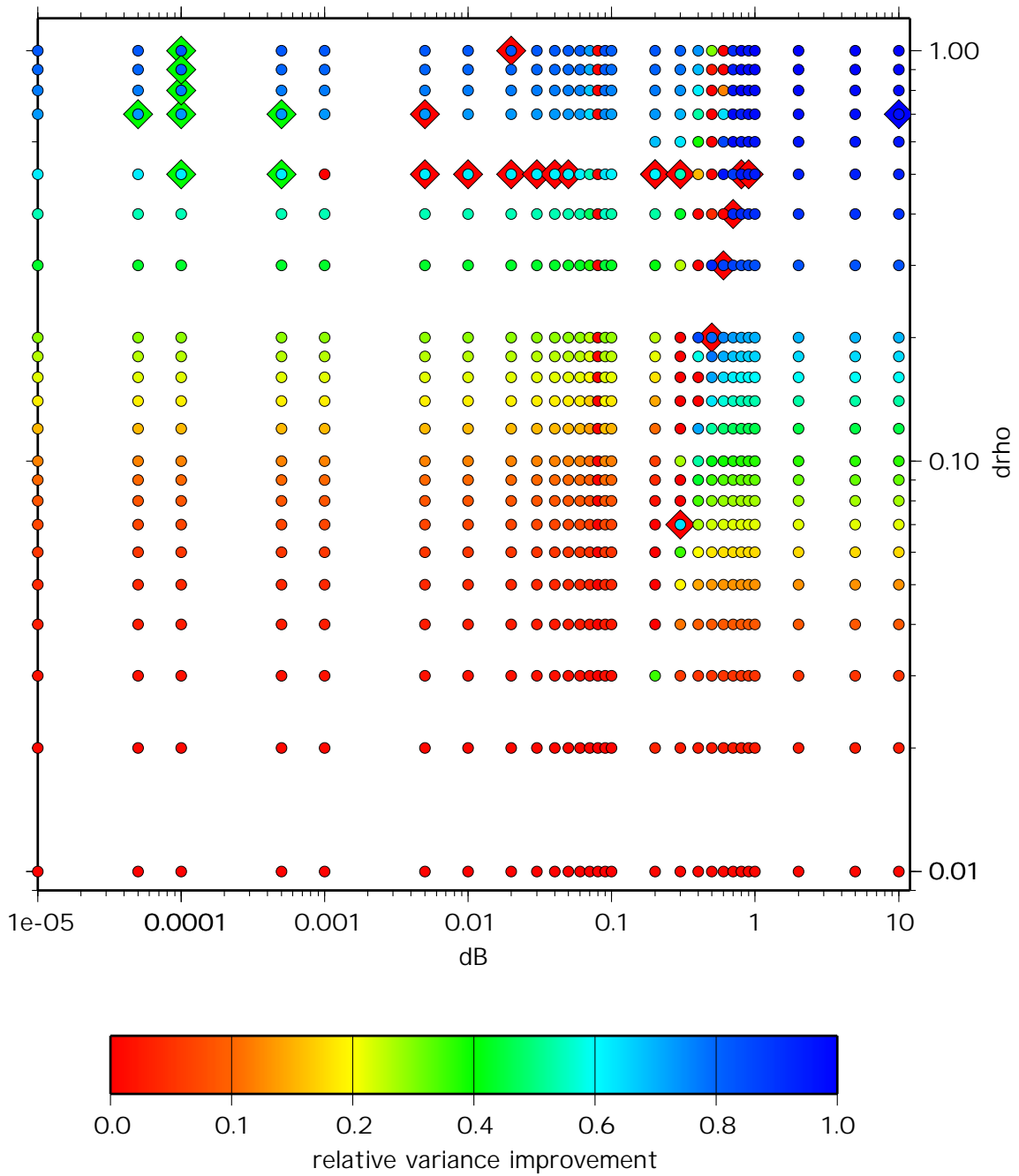
Figure 2.39: *Common variance reductions (product of the normalized variance reductions of the delay time and the gravity inversions), for a 2D trade-off to determine the optimum combination of the dB and the dρ inversion parameters. While the variance reductions are given by colored dots, the diamonds indicate the parameter combinations where iterative joint inversions are calculated. Red diamonds indicate divergent, green (close to) convergent behavior. The blue symbol denotes the decoupled case.*
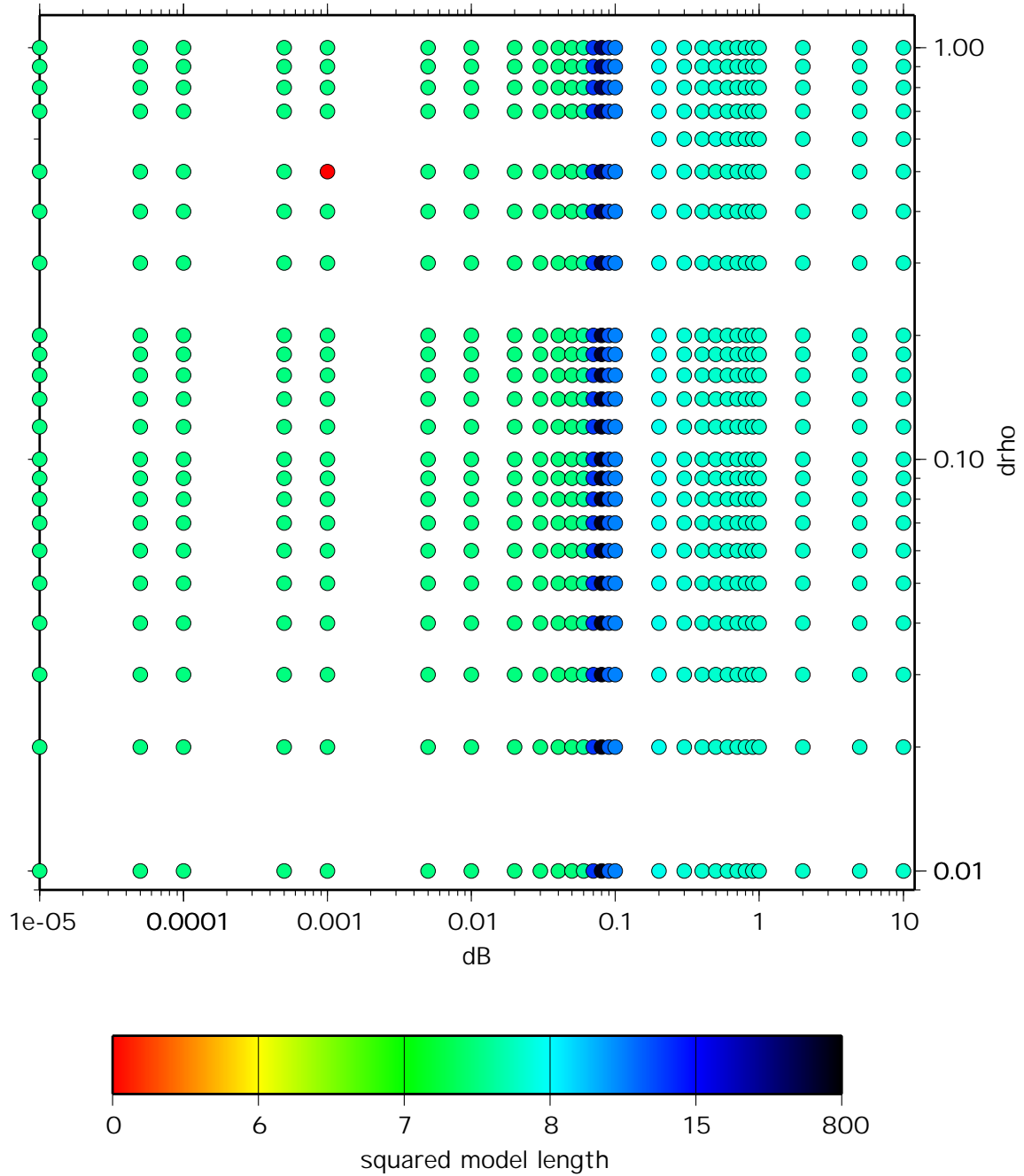
Figure 2.40: *Squared model lengths of the velocity model in $\frac{km^2}{s^2}$, derived from inversions for different combinations of the dB and the dρ inversion parameters. The optimum model lengths in terms of consistency with the single delaytime inversions is indicated by green colors. This is used for a 2D trade-off to determine the optimum combination of dB and dρ.*
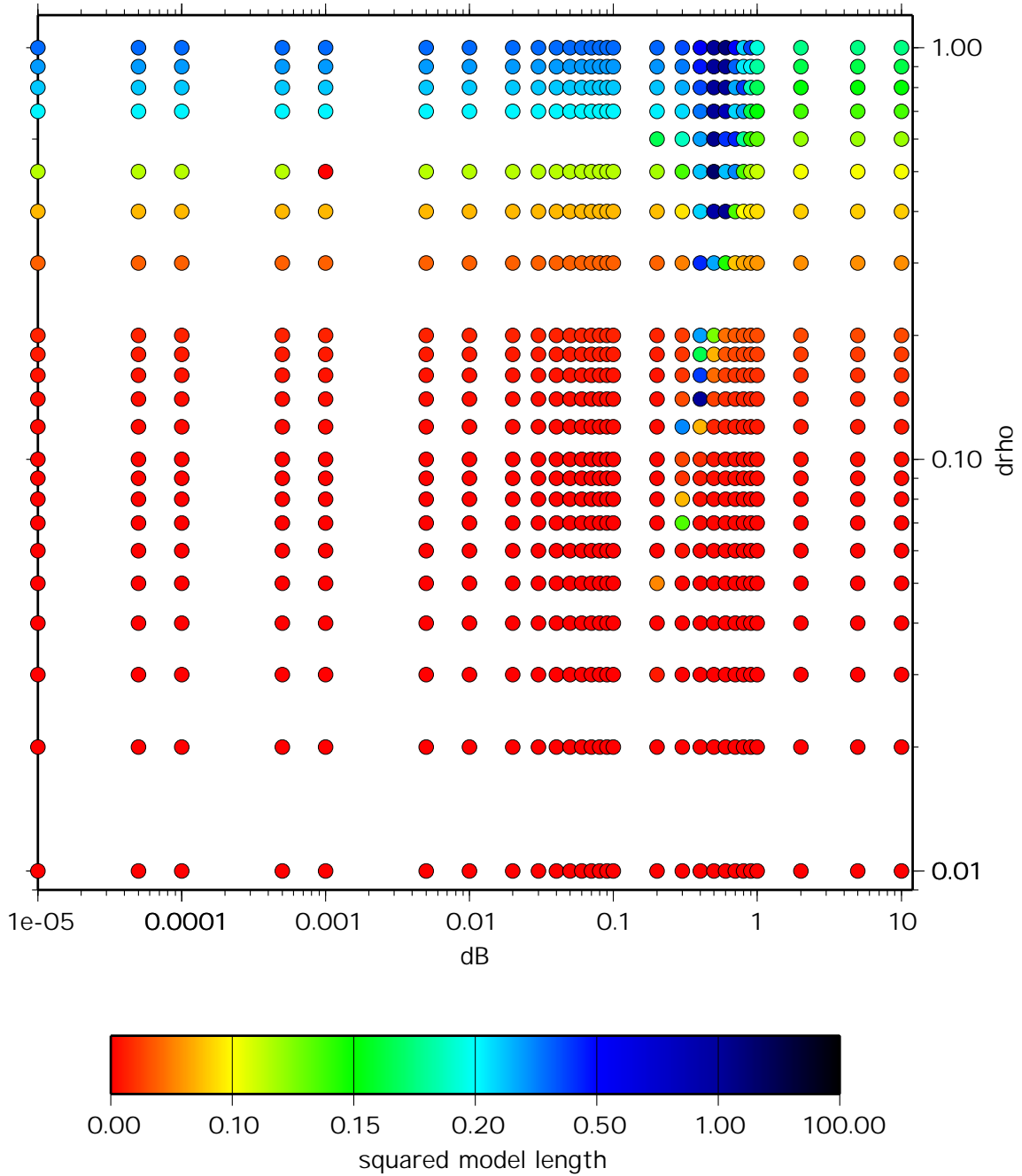
Figure 2.41: *Squared model lengths of the density model in $\frac{kg^2}{dm^6}$ derived from inversions for different combinations of the dB and the $d\rho$ inversion parameters. The optimum model lengths in terms of consistency with the single gravity inversions is indicated by green colors. This is used for a 2D trade-off to determine the optimum combination of dB and $d\rho$.*

Hence, the resulting models for the non iterative joint inversion are given in figures 2.42 and 2.43 for $d\rho = 0.7$ and $dB = 0.0001$.

First, the joint inversion results are evaluated with respect to the non joint inversion results that also incorporated *a priori* information in layers 1 and 2 (figure 2.27) .
Apart from the higher amplitudes and model lengths ($4.23km^2/s^2$ vs. $7.2km^2/s^2$) of the partly joint inversion velocity result, the two models are identical. This means, that the noniterative joint inversion cannot resolve the synthetic input structure better than the iterative pure delaytime inversion, based on the same heterogeneous starting model, which is a strong hint, that the inversion is strongly determined by the delay time data set.
The resulting velocity model is compared to the iterative pure delaytime inversion results (figure 2.27) and the synthetic input structure (figure 2.11). Obviously, the incorporation of the *a priori* information prevents the smearing of the positive NS-tending anomaly of the input model into layers 1 and 2, in contrast to the delay time inversion without *a priori* information. Although this very structure is found correctly in layer 3 of both models, the inversion using *a priori* information finds larger amplitudes and a more continuous structure. The results for layers 4 to 10 are to a very large degree identical, which clearly is an effect of the pure delaytime inversion applied to layers 5 to 10 in the joint inversion case.
According to the synthetic input model one would expect the same positive anomaly, found in layer 3 also to appear in layer 4, but both inversion results show the positive anomaly interrupted by a negative EW-tending signal which is caused by vertical smearing from lower layers. Both models find the EW-tending anomaly, originally located in layers 7 and 8 of the synthetic input model in layers 5 to 8. Due to the vertical smearing, the amplitudes are merely about half the expected. Layers 9 and 10 are relatively homogeneous in the inversion results which corresponds well to the synthetic input structure. Apart from layers 1 to 3 which are directly affected by the *a priori* information the inversion results are not improved by the non iterative joint inversion using the filtered gravity data set.
The comparison to the non iterative full joint inversion results shows, that the joint inversion with *a priori* information provides better results for layers 1 to 3, while the full joint inversion is able to recover the positive NS-tending anomaly in layer 4 as continuous structure. Apart from the fact, that the EW-tending anomaly can be recovered continuous in layer 7, the lower part of the model including the *a priori* information and the partial joint inversion reflects the synthetic input model better than the one of the full joint inversion.
Compared to the pure gravity inversion of the filtered data set (2.37) the results from the joint inversion (2.43) are in good agreement in layer 3. The structure in layer 4 obviously influenced by the velocity model. However, on the whole, the result of the partial joint inversion is reasonable.
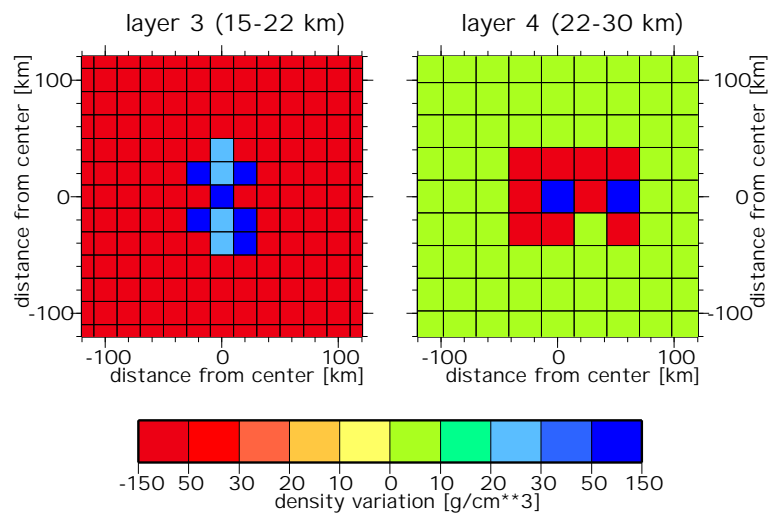
Figure 2.42: *Velocity model resulting from the non iterative joint inversion, using* a priori *information and a filtered gravity data set.*

Figure 2.43: *Velocity model resulting from the non iterative joint inversion, using* a priori *information and a filtered gravity data set.*

The iterative joint inversion In analogy to the proceedings prior to the iterative pure delay time inversion, the *a priori* variances, or in this case rather the scaling factors, have to be rechecked if they are still appropriate for the iterative inversion.

Though the most impartial way would be to recalculate the inversions given in figures 2.39, 2.40 and 2.41 iteratively in order to obtain the analogical maps of common variance improvement, and the respective model lengths for dB-dρ combinations, this is not done here, since this would be extremely demanding in terms of time and computational resources. Hence, a different approach is chosen: Iterative joint inversions were calculated for several dB-dρ combinations, located in the two regions defined as optimum during the non iterative trade-off process. In addition to that, inversions were also calculated, sampling possible other combinations. The parameters that were rechecked that way are indicated by diamonds in figure 2.39. According to the results for the iterative joint inversions the symbols are colored red, green and blue. Red denotes the dB-*d*ρ combinations that did not work for an iterative inversion and green indicates successful iterative inversion. Successfull is defined here as non divergent within three iterations. The blue diamond indicates that this inversion did converge, but did not provide any new insights, since the velocity and density models are decoupled. Green symbols are located in the area where the previous optimum was determined and the results do not differ significantly. This, on the one hand is an effect of the saturation of the coupling, as described already for the non iterative 2D trade-off procedure. On the other hand this also indicates that the inversion is to a very large degree determined by the delaytime data and that changing the scaling factor for the density model parameters has only little influence on the inversion result.

For consistency reasons, the same dB and dρ inversion parameters are used as in the non iterative joint inversion (i.e. dρ=0.7 and dB=0.0001).

The corresponding iterative inversion results for the velocity and density model are given in figure 2.44 and 2.45.

The result is given for the first iteration. Layers 1 and 2 are homogeneous due to the *a priori* information as in the non iterative case (hereafter also referred to as iteration 0). The anomaly recovered in layers 3 and 4 has the same shape in both iterations but is much stronger (twice as strong) for the 1st iteration. Considering the peak to peak amplitudes and calculating the average over layers 3 and 4, iteration 0 recovers 86.7% of the amplitude of the synthetic model, whereas the result from iteration 1 is about 70% too large. In both cases, the input structure can be recovered in layer 3 while it is disturbed by vertical smearing in layer 4. In layers 5 to 10 the amplitude for the velocity model is slightly decreased with respect to iteration 0. This results in very low amplitudes in the two lowermost layers which are homogeneous in the synthetic input model. Moreover, the decrease of amplitude is also strong in layer 5 which also corresponds to the originally homogeneous layer. The EW-tending structure of layers 7 and 8 in the synthetic model appear in layers 6 to 8 in the velocity model of the first iteration. This is in complete accordance with the results observed in further iterations: Too strong amplitudes in layers 3 and 4 while the amplitudes in layers 5 9 and 10 are greatly reduced. The inversion results of the two iterations are summarized in table 2.5.

Since the inversion results completely diverge for iteration 3 and higher the set of inversion parameters used for the iterative joint inversion cannot be the optimum one. The correct parameters have to be derived iteratively, like for the iterative pure delaytime inversion as e.g shown in figure 2.23 but in at least 3 dimensions (dBoug, *d*ρ and dB) in order to get objectively stable results.

Both 2-layer density models of iterations 0 and 1 basically show the same anomalies: In layer 3, a NS-tending high density structure corresponding to positive velocity perturbations in the velocity
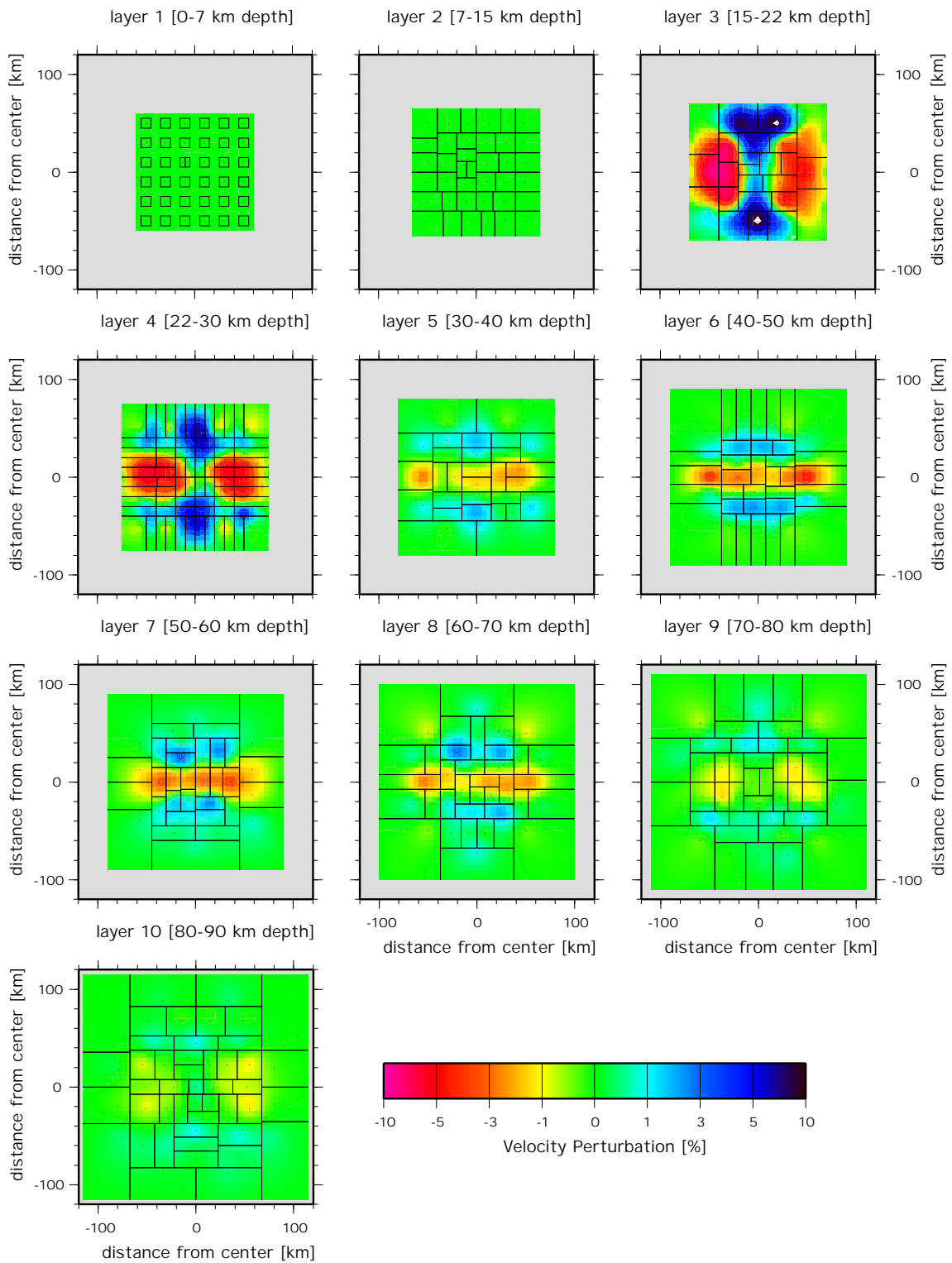
Figure 2.44: *Velocity model resulting from the iterative joint inversion, using* a priori *information and a filtered gravity data set.*

iteration	VI(v) [%]	ML(v) $[\frac{km^2}{s^2}]$	VI(ρ) [%]	ML(ρ) $[\frac{kg^2}{dm^6}]$
0	42.7	7.50	65.5	0.21
1	28.7	9.19	15.6	0.03

Table 2.5: *Variance improvements (VI) and model lengths (ML) for iterations 0 and 1 of the iterative partial joint inversion.*

model and a single high density parameter surrounded by low density blocks. An additional high density block is found to the right. The contrasts in both layers become smaller in the first iteration, which is the opposite of what is observed for the amplitudes of the velocity model in layers 3 and 4.
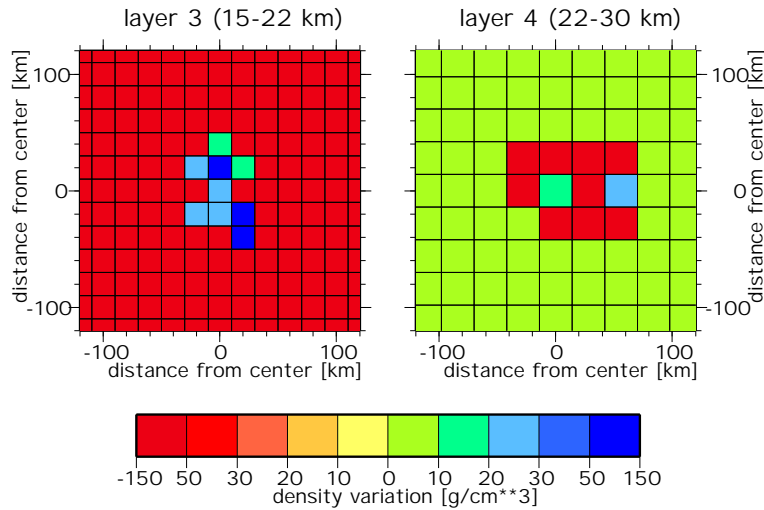

Figure 2.45: *Density model resulting from the iterative joint inversion, using a priori information and a filtered gravity data set.*

2.4   Discussion

The iterative pure delay time inversion provides stable and high resolution inversion results. This is mainly due to the 3D raytracing, the iterative inversion and, most important, the optimized variable parameterization. Best results are obtained beyond 50 km depth where no significant vertical smearing is observed. Above 50 km depth, the observed vertical smearing mostly is directed towards the surface, being largest directly underneath the seismological stations. This is an effect of the poor ray crossing, which, even for the optimized parameterization, cannot be completely avoided. However, the use of *a priori* information prevents the inversion from smearing anomalies into *a priori* known regions of the model space.

The inversion of Bouguer data using the optimized parameterization provides stable results, which are, however, due to the nonuniqueness, of no use without additional information. However, high pass filtering of the Bouguer gravity data set allows the assumption of a maximum depth of the structures contained in the filtered data set.

During the full joint inversion of the delay time and gravity data sets a large number of *a priori* variances have to be determined in a multi dimensional trade-off procedure. Together with the very large number of the density model parameters due to the optimized parameterization, this means very large expenses in terms of computational power and time. So this is prohibitive for large real applications and for the iterative full joint inversion. If the scaling between the individual data sets, models and additional constraints do not fully match the inversion problem, no stable result will be obtained. An additional problem is the difficult correlation of the density and velocity model parameters beyond 60 km depth due to the large blocks in the optimized density model. After all, the solution is hardly improved, also where a good correlation is possible, compared to the pure delaytime inversion. In addition to that, the joint inversion is not able to reduce the vertical smearing into the uppermost layers (1 and 2) and the velocity model is deteriorated beneath 60 km depths due to the coupling to the large blocks of the density model.

The partial joint inversion uses additional constraints like specific *a priori* information about the velocity and density model parameters in layers 1 and 2 and the filtered gravity data set, which only contains signal from structures no deeper than 30 km. Since the region, where the inversion is done jointly, only comprises layers 3 and 4, the number of density parameters is greatly reduced and the blocksizes in the respective layers allow a good correlation between density and velocity model parameters. This makes the partial joint inversion problem much more manageable than the full joint inversion. Vertical smearing into the two top layers is prevented by the use of *a priori* information, improving mainly layer 3. For the non iterative joint inversion, the resulting velocity model resembles closely the corresponding iterative pure delaytime inversion, which means, that the joint inversion is mainly determined by the delaytime data set with only little influence of the gravity data. However, first results of the iterative partial joint inversion suggest, that the inversion results are improved by the influence of the gravity data, already after the first iteration and the separation of the two bar like structures seems to be possible. Even if the partial joint inversion problem can be managed better than the full joint inversion, the multidimensional trade-off to find the optimum scalings is still enormously consumptuous in terms of time and computer power, especially in the iterative case. Due to the fact, that the iterative inversion is not yet completely stable for the applied combination of scaling factors and that the inversion results are determined by the delay time data to a very large degree, it is concluded, that the optimum combination of inversion parameters has not been found yet.

2.5 Conclusions

Iterative teleseismic delay time inversion using the variable, optimized parameterization and 3D ray tracing provides stable 3D velocity models with high resolution where the ray distribution is appropriate. However, even this inversion concept is not free of typical vertical smearing effects, induced by the subvertical incident teleseismic rays. This affects strongly the region directly beneath the stations and to a moderate extent the upper 50 km, using a station spacing of 20 km. The vertical smearing effects cannot be reduced by jointly inverting the delay time and the gravity data significantly. Only with additional constraints, like specific *a priori* information about the subsurface structure beneath the stations and a filtered Bouguer data set, the iterative joint inversion is able to reduce the vertical smearing in the top 50 km range. Since the inversion results, especially of the joint inversion, depend to a very large extent on the choice of the *a priori* variances of data, models and the additional constraints, it takes great efforts to determine the optimum combinations, especially in the iterative case. The determination of the correlation parameter B between velocity and density model parameters was not possible, mainly due to the nonlinearity. Considering the little improvement of the joint inversion results with respect to the models derived from the iterative pure delaytime inversion and the synthetic input structure and the enormous additional effort it must be said, that the method of choice clearly is the iterative pure delaytime inversion with specific *a priori* information about the crustal or lithospheric structure.

3 Application to teleseismic delaytime data in the Eifel region

In this chapter, the JI-3D method is applied to a real data set in the Eifel region, Germany. The inversion is done non iteratively for delay time data, only. As the velocity variations are in the range of -2% and the velocity gradients associated with the resulting plume are very small, an iterative inversion is neglected, assuming that the raypaths are not much influenced by the plume structure.

3.1 Introduction

Major hot-spots, centers of massive volcanism not linked to plate boundaries, such as Hawaii or Iceland, are commonly explained by mantle plumes (Morgan, 1971). These are believed to be 100-200 km wide columns of hot rock rising from great depth in the mantle that partially melt beneath the lithospheric plates. At a classical hot-spot, continuous magma production together with plate motion relative to the plume creates a hot-spot track, progressing in age away from the hot-spot. In contrast, for many intraplate volcanic centers with small eruption volumes and lack of a well-defined track, a plume-related origin is not clear. However, an extrapolation of the frequency-size distribution of observed hot-spots suggests that numerous smaller plumes may be present in the upper mantle Malamud and Turcotte (1999) and may feed relatively small eruption centers. In the following we report results of a major seismic tomography experiment, addressing the mantle structure below the Eifel volcanic fields, Germany.

3.2 Setting

In the East and West Eifel volcanic fields in Germany, west of the river Rhine (Fig. 3.1), about 300 small eruptions occurred between 700,000 and 10,800 BP (Schmincke et al., 1983). The total erupted volume of <15 km^3 represents an average magma flux that is several thousand times smaller than that of the Hawaiian hot-spot (White, 1993). About 250 m of uplift in this time interval (Meyer and Stets, 1998), ongoing exhalation of mantle helium (Griesshaber et al., 1992) and isotopic and trace element
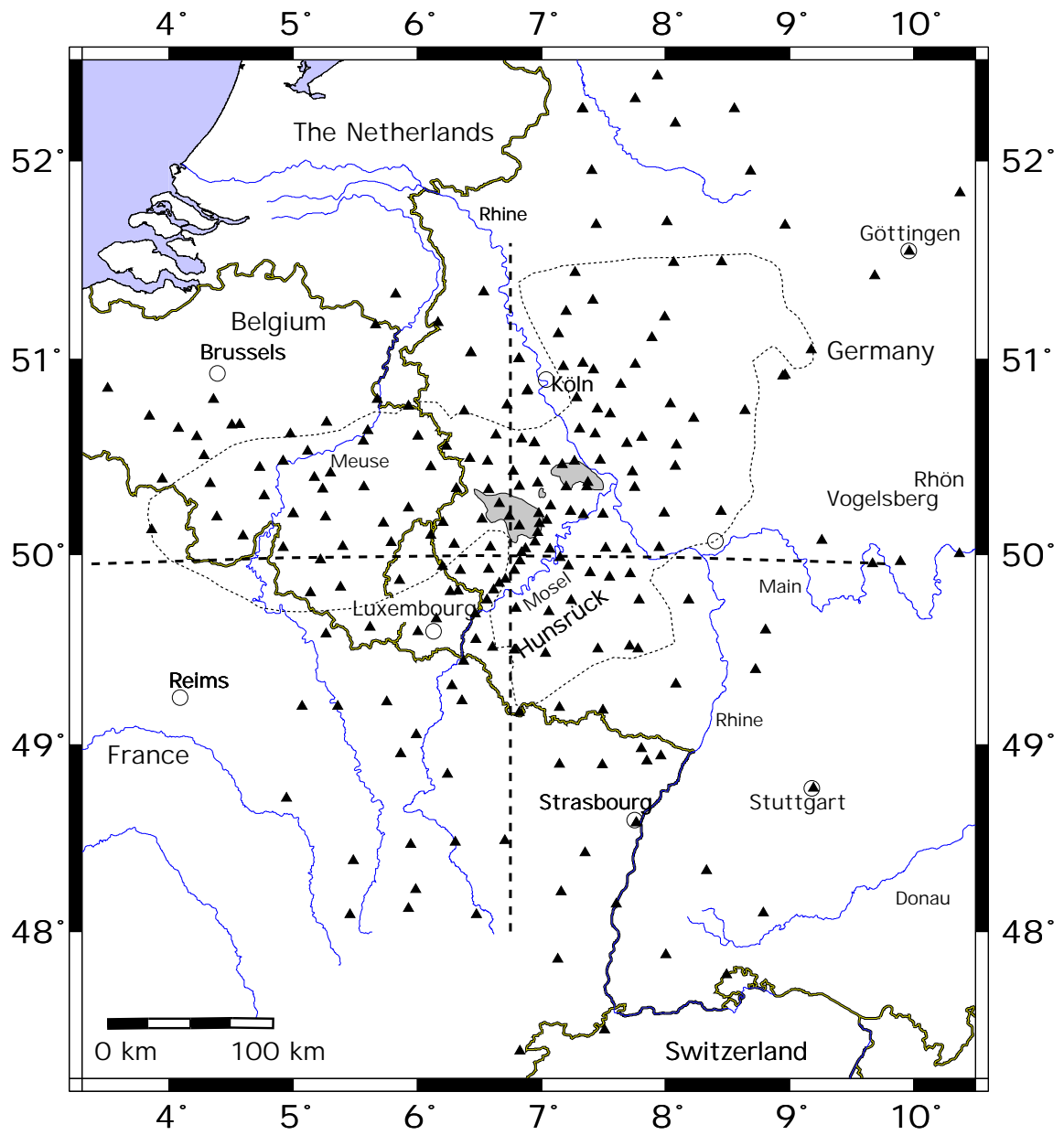
Figure 3.1: Location map with seismic station distribution (triangles), volcanic fields in the Eifel region (in grey) and the outline of the Rhenish Massif (dotted curve). The two dashed lines indicate the positions of the cross-sections in Fig. 3.3 a, b.

signatures (Hoernle et al., 1995 , Wedepohl and Baumann, 1999) all support a plume origin for the volcanism in the Eifel. It had been proposed that the various scattered volcanic fields (e.g. Vogelsberg, Rhön) stretching from the Eifel towards Eastern Europe represent a hot-spot track (Duncan et al., 1972). However, later radiometric dating (Lippolt, 1983) indicated that a clear age progression does not exist. As an alternative to a possible deep-mantle origin it has been suggested that the Central European volcanism is related to the lithospheric stress-field associated with the Alpine mountain building (Regenauer-Lieb, 1998). Earlier seismic studies indicated low velocities in the shallow mantle below the Eifel (Raikes and Bonjer, 1983 , Passier and Snieder, 1996), but had no resolution below 200 km depth. A global tomography study shows a wide plume-like structure in the lower mantle below Central Europe (Goes et al., 1999), but no clear connection through the transition zone to the shallow mantle.

3.3   Method and results

To image the mantle structure of the proposed Eifel plume with high resolution to a depth of at least 400 km, ten European institutions operated a seismic network with an aperture of 500 km by 500 km (Ritter et al., 2000). 158 mobile recording stations (Fig. 3.1) were deployed from November 1997 to June 1998. Data from 84 permanent stations add more observations of tele-seismic waves. Details on the design of the network and instrument types are given in Ritter et al.,(2000). To account for the different types of sensors, we deconvolved the instrument response functions. For a three-dimensional seismic tomography we determined 7319 P- and PKP-wave hand-picked arrival times with high precision at the restituted velocity seismograms. Based on the estimated accuracy of the timing, the data were sorted into three quality classes (± 0.01 s, ± 0.03 s and ± 0.05 s) and were weighted accordingly in the inversion.

We achieve a good coverage of azimuths and ray parameters with 66 teleseismic events (Fig. 3.2). For our P-wave model we use 203 stations west of 10.5°E. P-wave travel time residuals are calculated by subtracting the theoretical travel time for the IASP91 (Kennett et al., 1991) model from the observed travel time. To exclude travel time effects due to heterogeneities in the source region or event misslocation, we subtract the mean residual of each earthquake. This results in relative residuals that are mainly generated by seismic velocity perturbations underneath the station network. These relative residuals vary from -0.8 s to +0.8 s at our stations. There is a clear regional trend for late arrivals near the volcanic fields in the Eifel region, indicating reduced seismic velocity at depth.

To map the seismic velocity variations in the mantle a 3-D tomographic inversion, the JI-3D method, is used as described in chapter 2. The parameterization is determined exactly after the ray distribution depending on ray density and angular distribution of rays. This results in 20-200 km wide blocks in 10 layers down to 480 km depth. The inversion algorithm is a Bayesian scheme accounting for data quality (Zeyen and Achauer, 1997). The starting model is equivalent to the 1-D IASP91 (Kennett and Engdahl, 1991) velocity distribution. The uppermost layer at 0-5 km depth with low seismic velocity (5.1 km/s) serves mainly to account for stations statics. The resulting 3-D velocity model explains 88.6 % of the measured relative travel time residuals. In Figs. 3.3 a, b we present perturbations of the compressional wave speed relative to the IASP91 model (Kennett and Engdahl, 1991) for two vertical cross-sections (S-N and W-E). The point of intersection of the cross-sections corresponds approximately to the center of the main anomaly. Velocity perturbations larger than about 0.5 % are significant as found in resolution analyses (see below). A clear low-velocity anomaly (LVA) from about 70 km to at least 400 km depth is found, centerd slightly southwest of the West Eifel volcanic field. The width of this anomaly is about 100 km on average,

Figure 3.2: Distribution of events. The distance is given in degree relative to the center of the station network (triangle). The data set contains travel time readings from 51 events with P as first arrival (upper map). The PKP arrivals are from 15 events with an epicentral distance of more than 120° (lower map). The body wave magnitude of the events ranges from about 4.9 to 6.8. Precise hypocentral parameters were taken from Engdahl et al. (Engdahl et al., 1998).
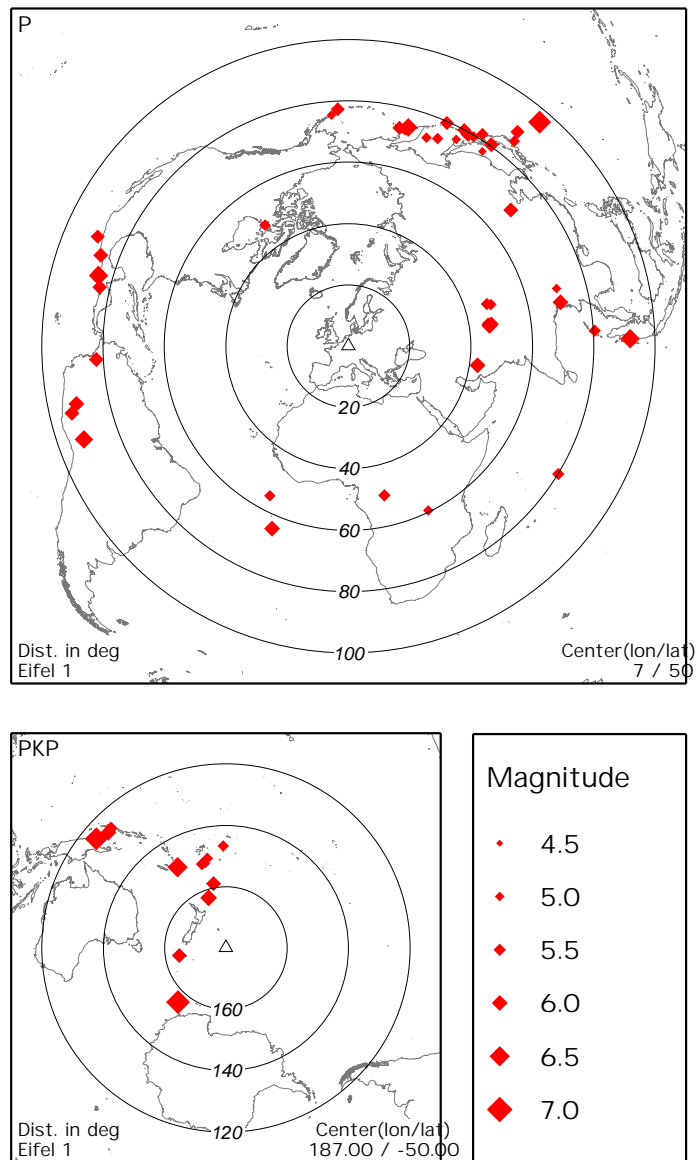
and the velocity contrast to the surrounding mantle reaches up to 2 %. Structure below 410 km depth is not well resolved, and the LVA may extend further down into the transition zone.

3.4   Resolution

Figs. 3.5 a, b show the model parameterization and ray density along the cross-sections. There are at least 30 rays per block with an adequate azimuthal coverage (Fig. 3.2). We performed reconstruction tests (van der Hilst, 1993) to estimate the spatial resolution in the model space. Synthetic travel time residuals were calculated for various assumed velocity structures and for the ray geometry of the actual experiment. Then we invert the synthetic data after adding Gaussian noise. Figs. 3.4 a, b show an example in which we want to test whether the deep plume structure seen in Figs. 3.3 a, b is true or the result of vertical smearing, along the ray paths, of a shallow low-velocity structure. The synthetic plume is 100 km wide, extends from 70 to 270 km depth and has a seismic velocity contrast of -3 %. The recovered images in Fig. 3.4 a, b are close to the input structure with only minor smearing of the velocity perturbations along the steeply inclined ray paths. The bottom of the synthetic plume at 270 km depth is blurred, however, a depth resolution of about 50 km can be achieved, suggesting that the deep plume structure seen in Figs. 3.3 a, b has been reliably mapped. Checkerboard and spike tests confirm that velocity anomalies of ± 2 % and about 50 km lateral extent can be recovered in the upper mantle below the center of the station network with our dataset. The largest artefact in our synthetic test is found at shallow depths above the plume (Figs. 3.4 a, b). Because the teleseismic ray paths hardly cross at crustal depths, the upper 30 km are poorly resolved. Therefore, the low velocities at shallow depths in Figs. 3.3 a, b might represent crustal heterogeneities, but they may be also artefacts of the inversion method.
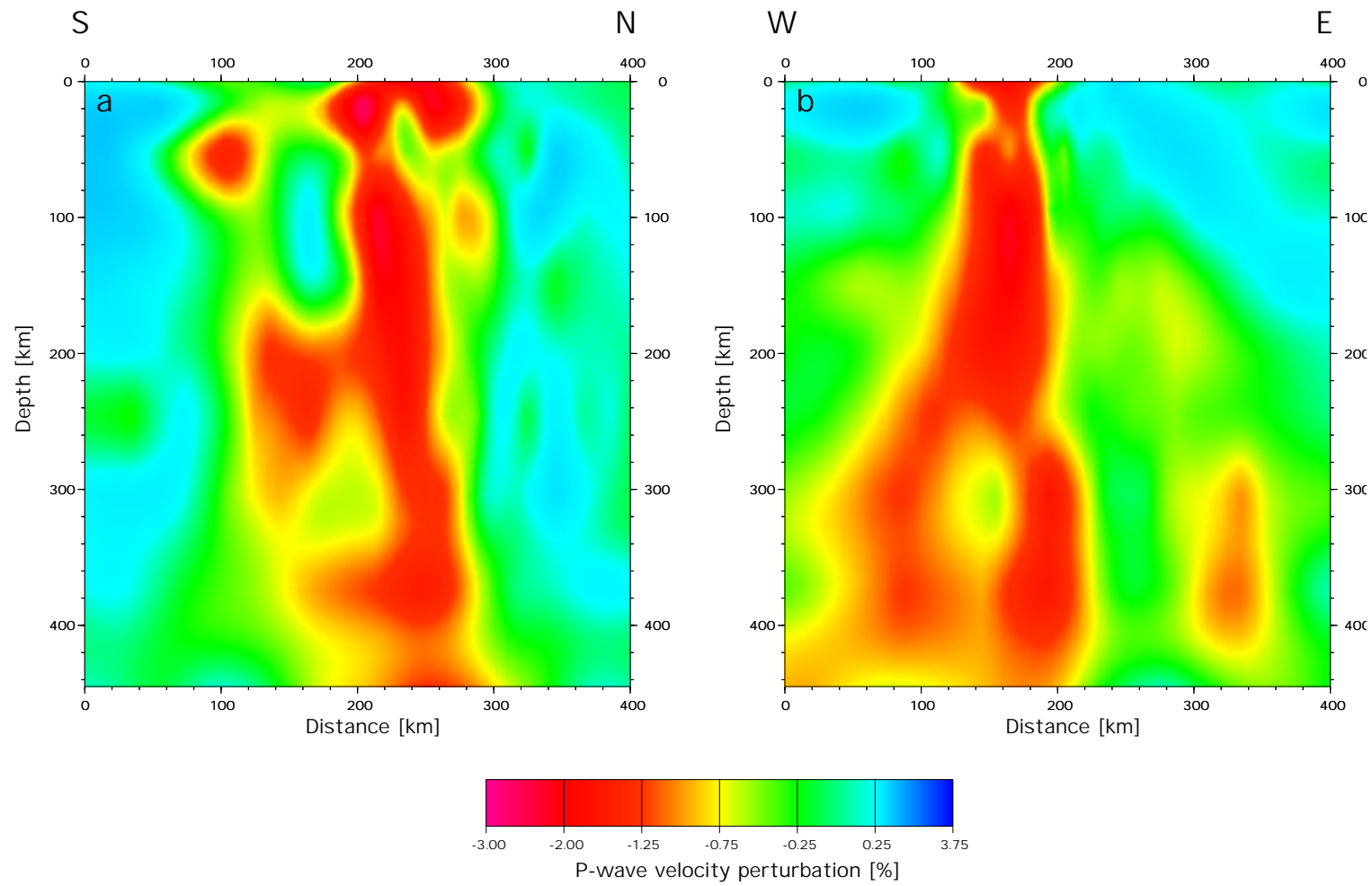
Figure 3.3: Displayed are perturbations of the P-wave velocity relative to the IASP91 model (Kennett and Engdahl, 1991), on the sections indicated in Figure 3.4. a, S-N cross-section and b, W-E cross-section. The reddish areas in the mantle are characterized by relatively low seismic velocity which most likely indicate increased temperature.
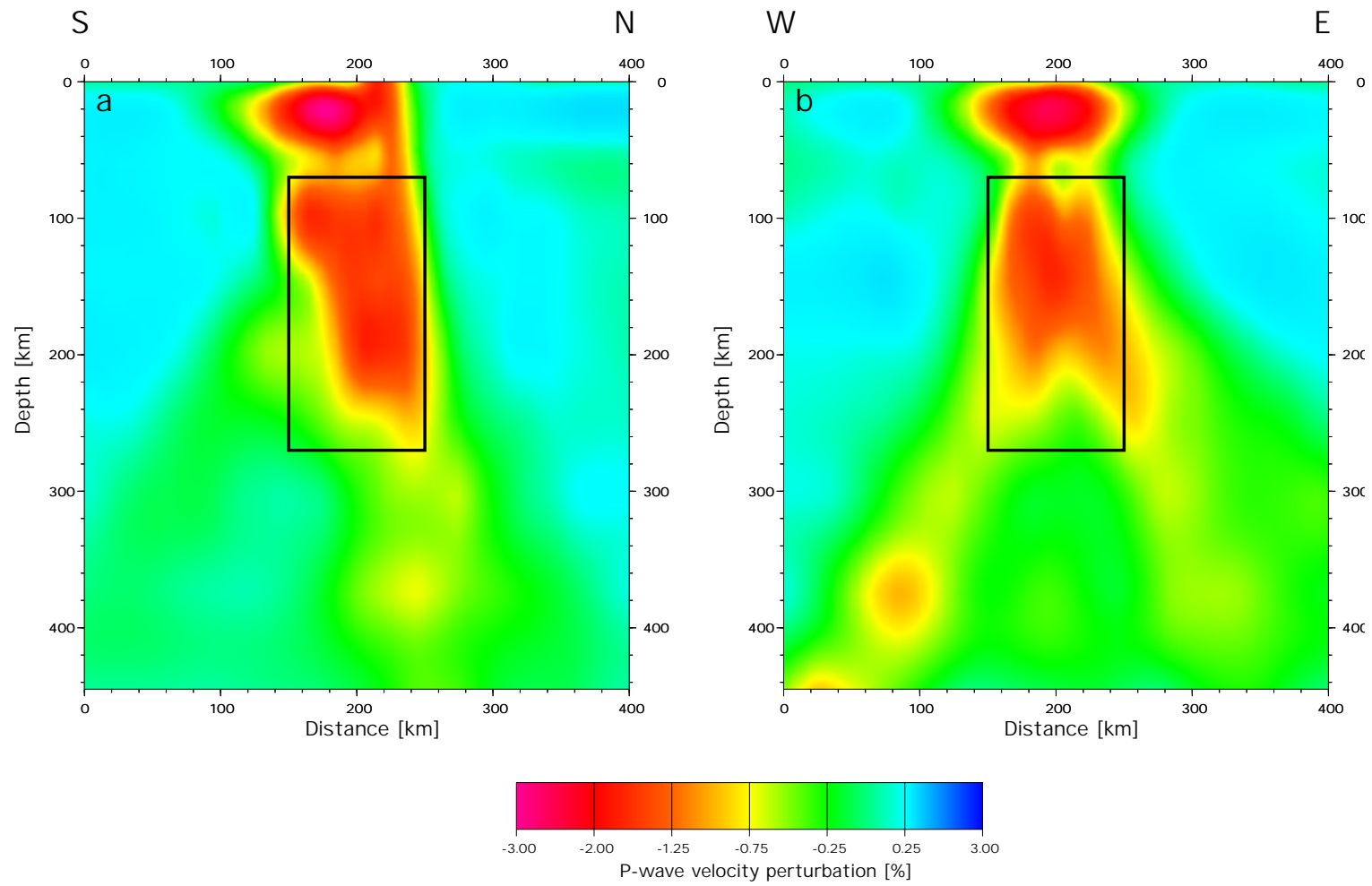
Figure 3.4: These figures show reconstruction tests to estimate the effect of vertical smearing in the model space. As input structure a 100 km wide synthetic plume from 70 to 270 km depth (outlined in black) with -3 % velocity contrast is used.
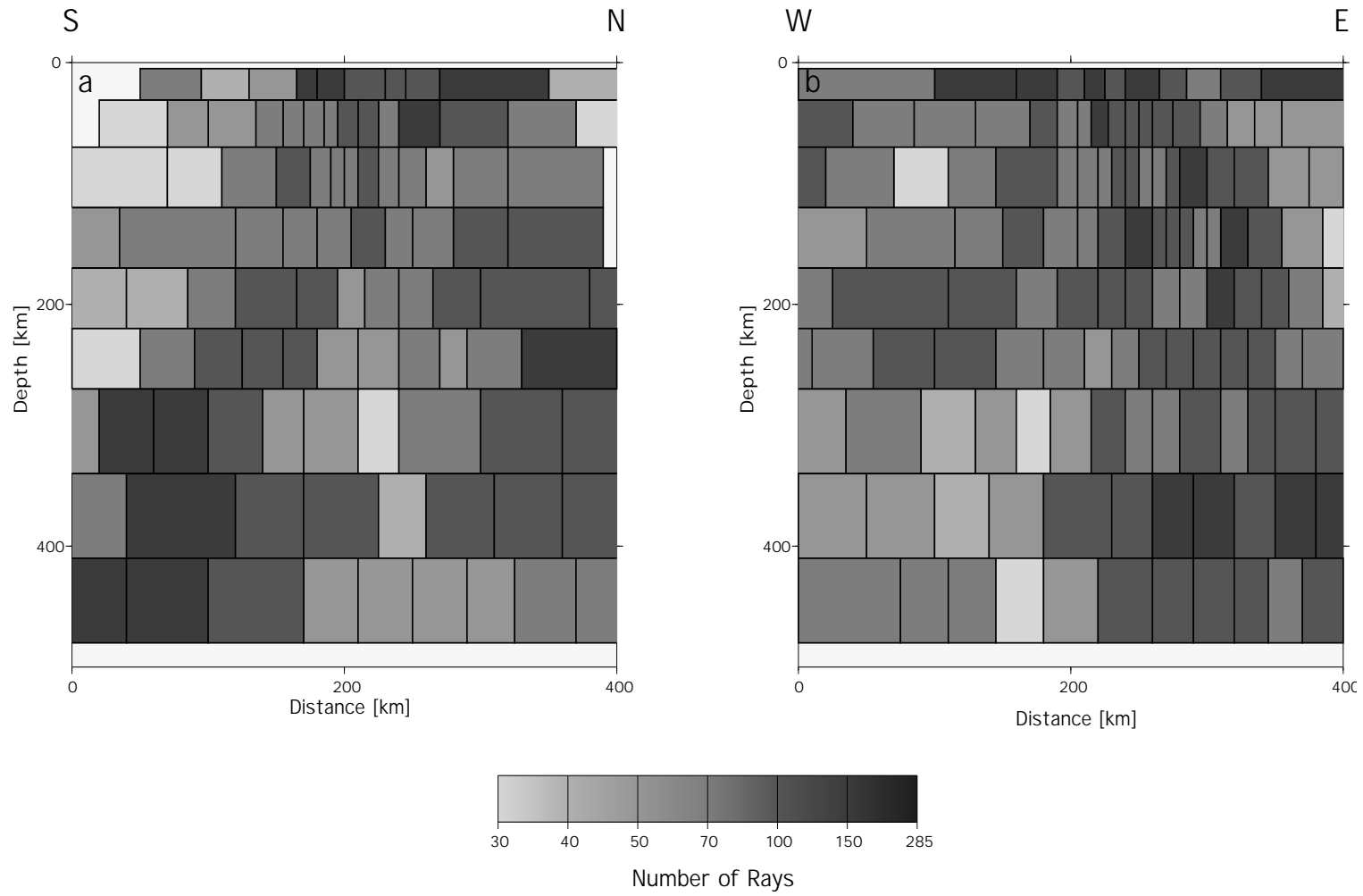
Figure 3.5: These figures display the block structure and ray coverage. Note that at least 30 rays with a good azimuthal coverage pass through each block.

3.5   Discussion and conclusions

The low velocity anomaly (LVA) in the upper mantle under the SW Eifel region is presumably caused by elevated temperature, because large P-velocity perturbations due to compositional variations are unlikely (Sobolev et al., 1997). Accounting for first-order effects like anharmonicity (Anderson et al., 1992 , Duffy and Anderson, 1989) and anelasticity (Sobolev et al., 1997 , Karato, 1993), the amplitude of the LVA underneath the Eifel can be explained by an excess temperature of 150-200 K ($\pm$100 K). The width of the anomaly and the inferred excess temperature agree approximately with estimates for plume conduits in the upper mantle from geodynamic and petrological models (White and McKenzie, 1995 , Ribe and Christensen, 1999). The temperature anomaly of the Eifel plume is similar to estimates based on seismic data for the plumes underneath the Massif Central (150-200 K) (Sobolev et al., 1997) and Iceland (about 150 K) (Shen et al., 1998), but is lower than for Hawaii (250-300 K) (Li et al., 2000). In seismic tomography the amplitude of anomalies is often reduced by smearing of structure. Because of the much higher station density than in comparable studies, our result is probably less affected, which could imply that the maximum temperature anomaly might be slightly lower for the Eifel than it is for other plumes.

The main implication of our result is that even small intra-continental volcanic fields can draw their magma supply from much more voluminous upflows, rising from greater depth in the mantle. The melt production in the plumes is controlled by variations in buoyancy flux, volatile content, thickness of the overlying lithosphere, and excess temperature of the plume (White, 1993 , White, 1995 , Ribe and Christensen, 1999 , Albers and Christensen, 1996). Subtle differences in the last two parameters can have a strong influence on the magma generation rate. A slightly reduced plume temperature or thicker lithosphere can make a plume appear insignificant or entirely invisible in terms of magmatism.

A common source for the two currently active European upper-mantle plumes (Massif Central and Eifel) and possibly for other Tertiary volcanic fields in Europe could be the broad anomaly (500 km by 500 km) in the lower mantle that has been identified with global tomography (Goes et al., 1999). Convection modeling has shown that the endothermic phase boundary at 660 km depth could hold up a lower-mantle upwelling (Brunet and Yuen, 2000) with ponding of plume material in a broad reservoir in the transition zone, from where several narrow plumes could be launched through the upper mantle. Although a connection of the Eifel plume through the transition zone is not yet proven, testing such a model is a challenge for future investigations.

3.6   Acknowledgments

4 Influence of seismic mantle discontinuities on teleseismic tomography

This chapter was inspired by the finding of chapter 3, that the Eifel plume apparently penetrates the seismic discontinuity at 410 km depth. Since the plume temperature is increased with respect to its surroundings, it will interact with the 410-phase boundary (Bina and Helffrich (1994)). The influence of the resulting deflections and the corresponding lateral velocity contrasts on the delay time data and their inversion are estimated.

4.1 Introduction

The increasing number of high-quality digital recording systems allows the installation of temporal seismic networks with large apertures of several hundred kilometers such as the TOR array (Gregersen et al., 1999) or the Eifel Plume experiment (Ritter et al., 2000). These station networks can be used e.g. for the detailed imaging of mantle plumes to great depth - even down into the lower mantle. Two seismological methods are particularly suited: Seismic tomography (Iyer and Hirahara, 1993) can image voluminous three-dimensional (3D) velocity anomalies due to hot plume material (e.g. Nataf, 2000 ; Foulger et al., 2000). Receiver-function analyses of compressional- (P) to shear- (S) wave converted phases can detect sharp velocity contrasts at seismic discontinuities (Shen et al., 1998 ; Li et al., 2000). The input data for the tomographic inversions are mostly (tele-)seismic travel-time residuals for P- or S-waves relative to a standard 1D Earth model, e.g. ak135 (Kennett et al., 1995).

The cold downward flows (slabs) and hot upward flows (plumes) in the mantle cause depths variations (deflections) of the 410 and 660 km discontinuities which confine the mantle transition zone. These discontinuities are thought to represent phase transition boundaries of the olivine component in the mantle. When the transition is in thermodynamic equilibrium, a lateral temperature difference will lead to variations in transition pressure, hence depth (e.g. Bina and Helffrich, 1994). The direction and the amount of the vertical deflection depend on the sign and magnitude of the Clausius-Clapeyron slope which is characteristic for the specific isochemical phase transition. The resulting deflections have been modeled with numerical simulations (Christensen, 1998 ; Brunet and Yuen, 2000), and they are observed seismologically underneath mantle plumes with receiver-function analyses (Shen et al., 1998 ; Li et al., 2000).

Across the 410 km and 660 km discontinuities large velocity contrasts of about 4 % and 6 % occur. However, the lateral travel-time anomalies resulting from the vertical deflections at these

discontinuities are mostly neglected in tomographic modeling, because their typical maximum deflection amplitude of about 25 km (Shen et al., 1998 ; Li et al., 2000) is below the spatial resolution of the tomographic methods. In the following we study the influence, which arises from the complex 3D seismic velocity distribution by plume-discontinuity interaction, with synthetic P-wave data and 3D inversion modeling. The input parameters such as plume structure and temperature variation or station and event distribution are taken from results of geodynamic modeling (Christensen, 1998 ; Brunet and Yuen, 2000) and recent seismic experiments (Shen et al., 1998 ; Foulger et al., 2000 ; Li et al., 2000 ; Ritter et al., 2000). Several models are examined to estimate the influence of the different structural parameters on the tomographic results. It is also tested how results of receiver function studies may help to improve tomographic mantle images by incorporating a priori information. Future experimental studies may benefit from this work, because the resolution and modeling limits of the tomographic approach are outlined.

4.2 Inversion strategy and model description

4.2.1 Tomographic inversion method

Based on a realistic station and source distribution (see below) ray paths and travel times are calculated through 3D synthetic plume structures. The forward calculation assumes that a plane wavefront of a teleseismic event enters the model space from below. Using the approximate 3D raytracing routine by Steck and Prothero (1991), ray paths are calculated from the wavefront to the recording stations. The timing accuracy along the more than 800 km long ray paths is about ± 0.02 s as tested against analytical solutions (Portmann, 2000). This part of the computation is the most time-consuming work (about 60 hours CPU time), and it is done with a parallel IBM SP2 computer. Afterwards travel-time residuals are determined relative to the ak135 1D Earth model (Kennett et al., 1995). By subtracting the mean residual of each event, relative travel-time residuals are determined for the further modeling. To achieve realistic data, Gaussian noise with a mean value of 0.04 s is added. The amplitude of the noise is used for weighting the data during the inversion.

The forward calculated travel times are the input for the tomographic modeling. For the imaging we use the JI-3D method, described in chapter 2. The model space is parameterised at velocity nodes. The spatial distribution of the nodes depends on the ray density as well as the distribution of ray directions and incidence angles. In this way the model space is about equally well resolved as controlled by an inspection of the resolution matrix. The influence of the parameterization on the inversion result is reduced by applying an offset-and-average technique for horizontal smoothing (Evans and Zucca, 1988). All models are averaged from 9 independent inversions whose parameterization nodes have been shifted by 1/3 of their shortest distance in E-W and N-S directions. The inversion is calculated with a Bayes algorithm as described by Zeyen and Achauer (1997). The covariance term in the Bayes inversion is used to incorporate a priori information. E.g. it is possible to fix the seismic velocity at specific parts of the model space such as at seismic discontinuities. As background model for the forward calculations and starting model for the inversion we use the 1D ak135 velocity model (Kennett et al., 1995). To display the results the images are smoothed with splines in tension (Smith and Wessel, 1990) to avoid plotting artefacts.
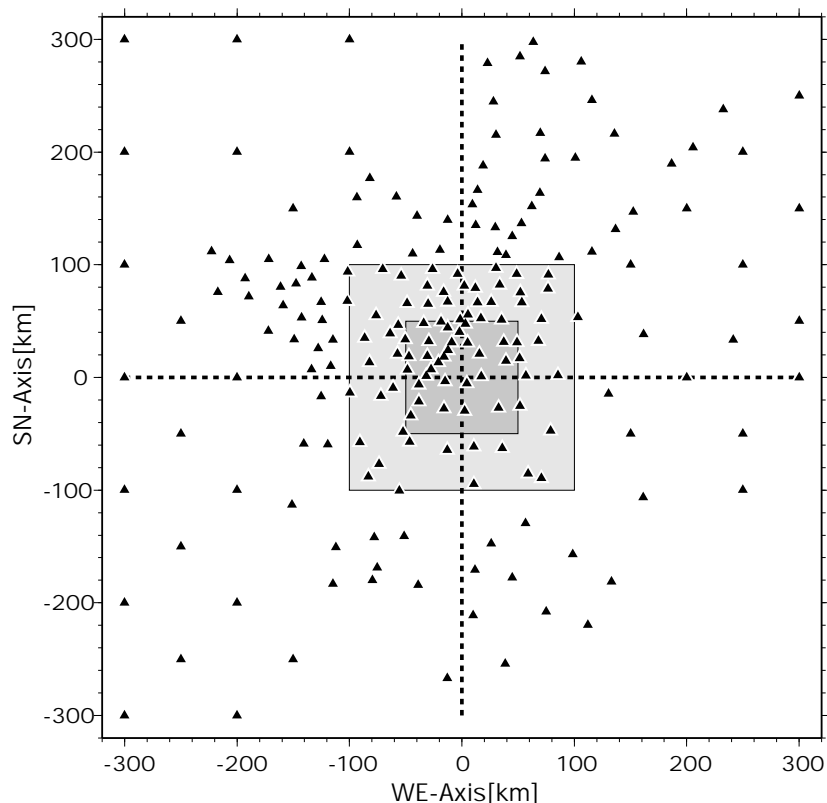
Figure 4.1: Station distribution used for the synthetic modeling. The 206 stations are arranged similar to the recent Eifel Plume experiment (Ritter et al., 2000). The asymmetry of the station network is typical for real applications. The small and big squares indicate the position of the input plume structures in the upper and lower mantle, respectively.

4.2.2 Station and epicenter distribution

The distribution of the 206 recording stations (Fig. 4.1) is typical for a large-scale mobile 2D network. The chosen experimental design is close to that of the Eifel Plume Project (EPP) (Ritter et al., 2000) and includes a central region with dense station spacing ($\sim$15 km) and larger distances ($\sim$50 km) between stations at the edges. Some 20 stations were added around the edges of the original EPP network to increase the aperture and hence the resolution properties in the transition zone and uppermost lower mantle. A similar station distribution may be applied in oceanic environments around volcanic archipelagos with land stations on islands and surrounding ocean bottom recorders. The aperture of the network in Fig. 4.1 is 600 km by 600 km and allows to recover velocity perturbations to some 700 km depth.

The 52 teleseismic sources (Fig. 4.2) are at least 40° apart from the stations and provide first arrivals from P and PKP phases. The number of the travel-time residuals is correlated with an assumed body-wave magnitude between 5 and 7 of the corresponding event. Thus strong events provide more data than weak events as is the case in real applications. The epicenter distribution is also similar to the EPP with four additional sources in former azimuthal gaps (NW

87

Figure 4.2: Source distribution around the station network (triangle). a) 43 event locations with mantle P waves as first arrivals; b) 9 event locations with PKP core phases as first arrivals. The distribution of the events and their assumed body-wave magnitude (Mb) are representative for temporary recording networks in Europe.
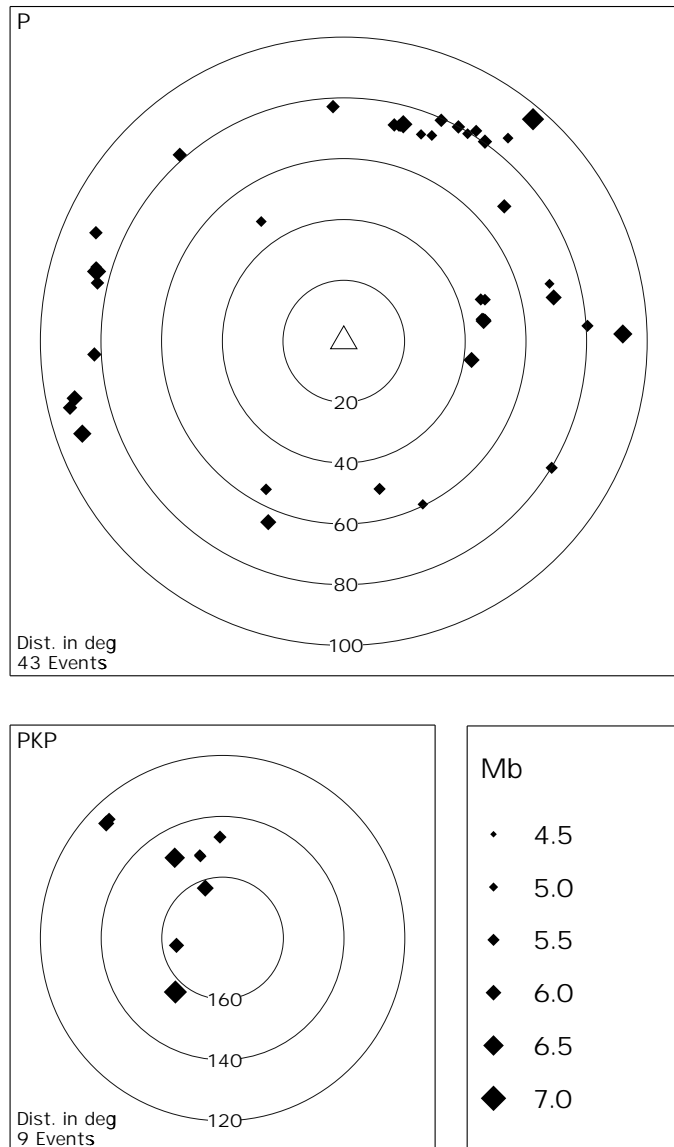
and S direction). In general, the overall ray distribution is typical for teleseismic tomography experiments in Europe, but may also be representative for numerous other targets on Earth. To simulate instrument failures, not all events are recorded at each station. All together 6900 synthetic seismic rays are calculated.

4.2.3 Forward models

As we want to examine systematically the combined effect of velocity perturbations due to a thermal plume and deflected mantle discontinuities, four different input models are used for the forward calculations (Fig. 4.3). First an upper mantle plume, starting in the transition zone, is used (Fig. 4.3a). Model 1 (M1) contains the upper mantle plume and flat, undisturbed mantle discontinuities. Model 2 (M2) contains the same plume plus a displaced 410 km mantle discontinuity as expected from the excess temperature of the plume. Second a lower mantle plume
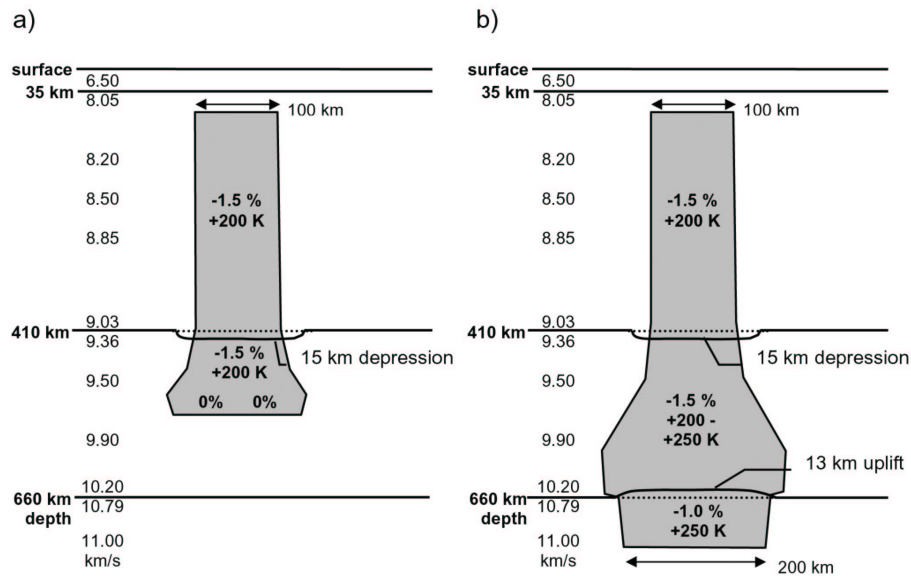
Figure 4.3: Side view of the three-dimensional input models used for the forward calculation. The background velocity is in km/s. Velocity and temperature contrasts are given relative to the surrounding mantle. a) Upper mantle plume for model M1 with a flat discontinuity at 410 km depth (dotted line) and for model M2 with a downwarped 410 km discontinuity (solid line). b) Lower mantle plume M3 with flat discontinuities and model M4 with deflected discontinuities due to the excess temperature of the plume.

is investigated (Fig. 4.3b). Model 3 (M3) contains the deeper mantle plume and flat, undisturbed mantle discontinuities at 410 and 660 km depth. Model 4 (M4) contains the same deep plume plus displaced discontinuities as expected from the excess temperature of the plume.

The background P-wave velocity in models M1 to M4 corresponds to the ak135 velocity model (Kennett et al., 1995). The upper mantle plume in M1 and M2 starts at 500 km depth and rises to 100 km depth near the bottom of an assumed lithospheric plate (Fig. 4.3a). At the deepest part its source region is 250 km wide, and the conduit in the upper mantle has a diameter of 100 km. Its assumed excess temperature is 200 K as expected from geodynamic modeling (Albers and Christensen, 1996 ; Brunet and Yuen, 2000). This corresponds to a reduction of the P-velocity of about 1.5 %. Between 410 km and 500 km depth the velocity contrast is reduced from 1.5 % to 0 %. Models M3 and M4 contain a 650 km high plume conduit between 100 km and 750 km depth (Fig. 4.3b). In the lower mantle the plume is 200 km wide and has 250 K excess temperature. After crossing the 660 km discontinuity the thermal anomaly increases laterally to 250 km width. Towards the top of the transition zone it narrows to about 100 km width at 410 km as in models M1 and M2. In the upper mantle the plume structure is 100 km wide and ends at 100 km depth. The P-velocity contrast is -1 % in the lower mantle and -1.5 % above. Deflections of the 410 km and 660 km discontinuities are neglected in M3 but included in M4. The maximum expected travel-time residual for the lower mantle plume is about 1.05 s delay for steeply incoming rays (e.g. core phases) which propagate completely through the plume.

inversion	A	B	C	D	E	F
input data from model	M1	M2	M2	M3	M4	M4
starting model	1D(M1)	1D(M1)	3D(M2)	1D(M3)	1D(M3)	3D(M4)
no. of travel-time data	6900	6900	6900	6900	6900	6900
no. of layers	14	14	16	14	14	16
no. of unknowns	1242	1242	1415	1242	1242	1415
variance reduction	45 %	42 %	43 %	54 %	50 %	52 %

Table 4.1: Parameters for inversion A-F.

The 410 km and 660 km discontinuities are deflected by +15 km (M2 and M4) and -13 km (M4), respectively (Fig. 4.3). These deflections are determined in the following way: Around the 410 km discontinuity we assume a temperature increase of 200 K. Using a Clausius-Clapeyron slope of 2.9 MPa/K (Bina and Helffrich, 1994) for the phase transition, a change in pressure of 0.58 GPa results which is equivalent to a 14.6 km depression. The P-wave velocity at 410 km depth jumps from 9.03 km/s to 9.36 km/s (Kennett et al., 1995) what corresponds to an increase of 3.65 %. The positive temperature anomaly of 200 K due to the plume causes a decrease in P-velocity of about 1.2 % above and below the 410 discontinuity. Around the 660 km discontinuity we assume a lateral temperature increase of 250 K. Using a Clausius-Clapeyron slope of -2.0 MPa/K (Bina and Helffrich, 1994), a change in pressure of -0.50 GPa results which is equivalent to a 13.3 km updoming of the phase boundary. The P-wave velocity at 660 km depth jumps from 10.20 km/s to 10.79 km/s (Kennett et al, 1995) and this corresponds to an increase of 5.78 %. The positive temperature anomaly of 250 K causes a decrease in P-velocity of about 1.3 % above and below the 660 discontinuity.

4.2.4   Inversion models

For six tomographic reconstructions (A-F) four data sets the with travel-time residuals obtained from models M1-M4 are inverted, using four different starting models: For inversion A the data set of the upper mantle plume M1 and the 1D background velocity of M1 is used (Table 1). In a second inversion B the data set from M2 is inverted, using the same 1D input model. Then the data set from M2 is inverted, using a background starting model as for inversions A and B, but with the 3D deflection of the 410 km discontinuity as fixed a priori information (inversion C). The same procedure is performed for the lower mantle plume. The data sets from M3 and M4 are inverted with the 1D velocity background (inversions D and E, respectively). Finally, the data set from M4 is inverted, using the same background starting model, but with the 3D deflections of the 410 and 660 km discontinuities as fixed a priori information (inversion F). For the insertion of the a priori information in inversions C and F a 15 km thick layer (410 - 425 km depth) and a 13 km (647 - 660 km depth) thick layer are included in the parameterization. These layers remain unchanged during the inversion. Table 1 summarises the inversion parameters.

4.3   Travel times

4.3.1   Theoretical travel-time anomalies

The maximum expected travel-time residuals due to the plume structures (Fig. 4.3) are delays of 0.6 s in the upper mantle (-1.5 % velocity perturbation) and 0.1 s in the lower mantle (down to 750 km depth, -1 % v.p.). In the transition zone (0 to -1.5 % v.p.) the delays are 0.2 and 0.4 s for the upper and lower mantle plume, respectively. The P-wave velocity contrasts across the 410 and 660 km discontinuities are 3.65 % and 5.78 %, respectively. Thus the 15 km downward and 13 km upward deflections in models M2 and M4 are related to +0.07 s and -0.06 s travel-time residuals. Along steeply incident rays (e.g. core phases such as PKP) that traverse both deflected regions in M4 the net travel-time effect is nearly zero. Compared to the travel-time anomalies related to the lower mantle plume, the residuals originating from the deflected 410 and 660 km discontinuities in M4 are much smaller (1-10 %). Therefore, the best possibility for finding the deflections of the discontinuities in a tomography study is along inclined rays which pass only through one deflected region. Steeply incident rays with a long travel path inside the plume are not much affected in terms of travel-time by the deflected discontinuities.

4.3.2   Modeled travel-time residuals

The spatial distribution of the forward calculated data in models M1 and M2 with the upper mantle plume is displayed in Fig. 4.4 for different backazimuth (BAZ) ranges to the events. The plume is centered at 0 km / 0 km. Each data point represents the average relative travel-time residual (mean residual) at a station for a 90° BAZ range. In Fig. 4.4a mean residuals of events to the north (315° < BAZ < 360° and 0° < BAZ < 45°) are displayed. In the southern part of the station network there is a clear delay of up to 1 s for P-phases which traverse the plume structure. This delay is in accordance with theoretical considerations (see above). Small deviations from the expected residuals are due to the added noise. The spatial distribution of the positive residuals as well as data points without delay outside the 'shadow' of the plume indicate the reliability of the forward calculation. In Fig. 4.4b the mean residuals calculated for events in the west are shown. East of the plume there are delayed arrivals as expected from the ray geometry. The mean residuals along steeply incident ray paths (PKP-phases, events in Fig. 4.2b) are displayed in Fig. 4.4c. These data clearly map the horizontal extension of the anomaly, but they have hardly any vertical resolution. A comparison between the residuals for models M1 and model M2 indicates that the influence of the deflected discontinuities is small on the travel-time data (see right column in Fig. 4.4). The differences (DIFF) are mainly below 0.1 s as predicted above (section 3.1). The biggest difference is found for the PKP data (Fig. 4.4c) which have an additional delay of about 0.1 s due to the downwarping of the 410 discontinuity.

In Fig. 4.5 the spatial distribution of the forward calculated data is displayed for the lower mantle plume (M3 and M4). In Fig. 4.5a mean residuals of events to the north are displayed. In the southern part of the station network there is a clear delay of up to 1.2 s for P-phases which traverse the deep plume structure. As in Fig. 4.4 the 'shadow' of the plume is clearly seen, but it is larger, because the plume is deeper. In Fig. 4.5b the mean residuals calculated for events in the west show delayed arrivals to the east of the plume as expected from the ray geometry. Again the mean residuals along steeply incident ray paths (PKP-phases) in Fig. 4.5c clearly map the horizontal extension of the anomaly. The comparison between the residuals for
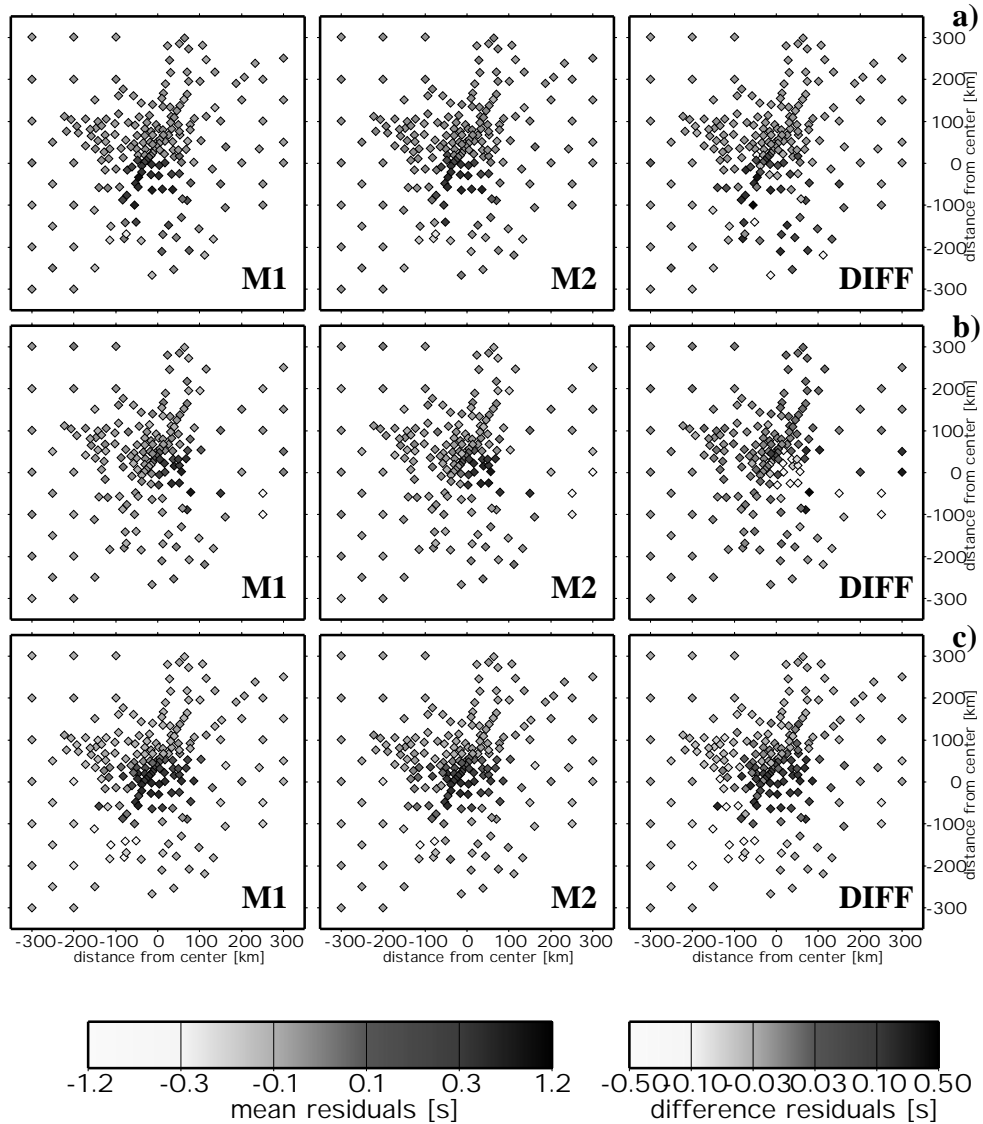
Figure 4.4: Average synthetic relative travel-time residuals (mean residuals) for the upper mantle plume as function of location (see network in Fig. 4.1) and backazimuth (BAZ) to the events. The residuals are calculated through input models M1 (first column), M2 (second column) and averaged for a 90° BAZ range. The differences arising from the deflected discontinuities are shown in the third column. a) Mean residuals from source locations north of the network (backazimuth 315°-45°) with delayed arrivals to the south. b) Mean residuals from events in the west (BAZ 225°-315°). c) Mean residuals from steeply incident PKP phases from events at distances of more than 120° (Fig. 4.2b). Note the different shading scale for the small differences displayed in the third column.
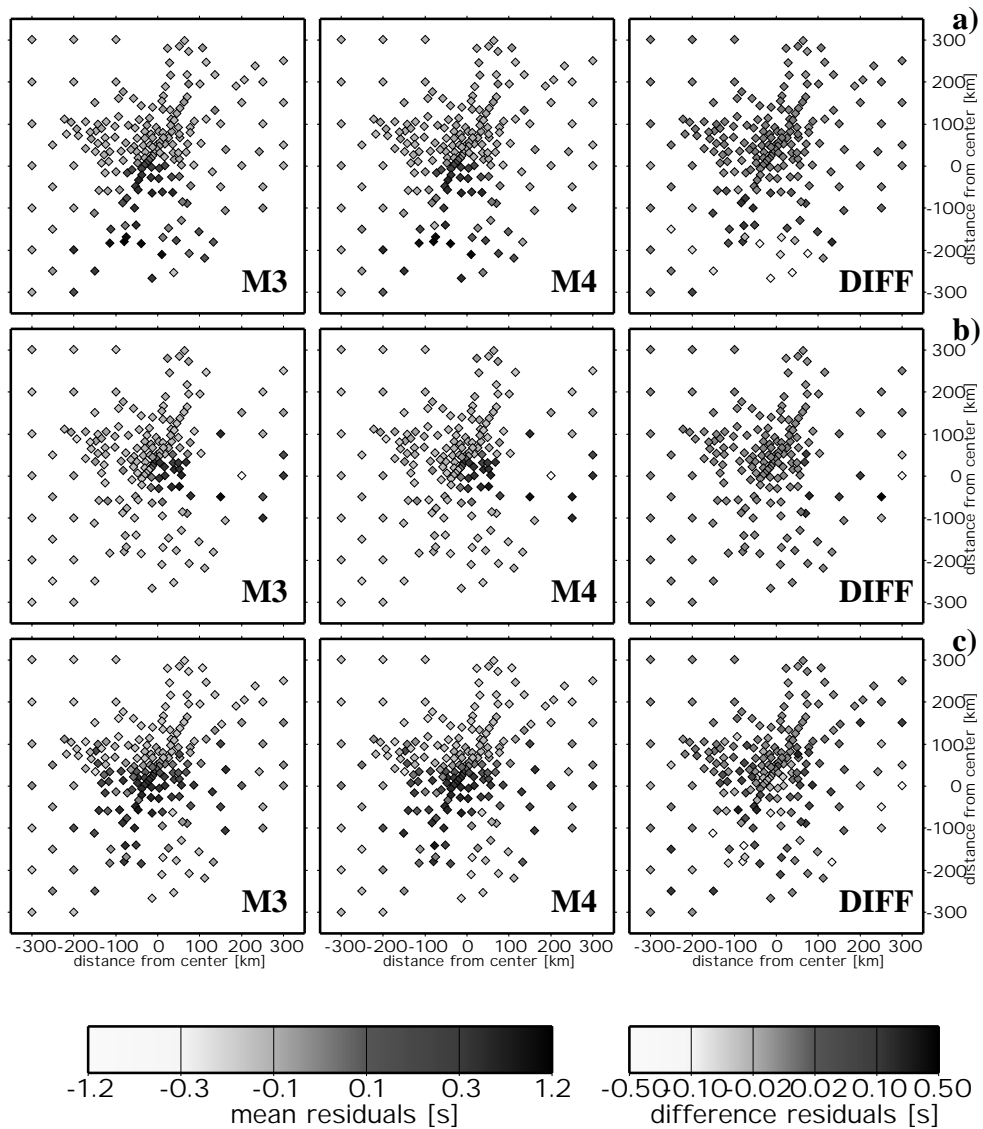
Figure 4.5: Average synthetic relative travel-time residuals (mean residuals) as for Fig. 4.4 but for the lower mantle plume from input models M3 (first column), M4 (second column) and their differences (third column). a) Mean residuals from source locations north of the network (backazimuth 315°-45°). b) Mean residuals from events in the west (BAZ 225°-315°). c) Mean residuals from steeply incident PKP phases (Fig. 4.2b). Note the different shading scale for the small differences displayed in the third column.

models M3 and M4 in the right column in Fig. 4.5indicates that the influence of the oppositely deflected discontinuities is small on the travel-time data. This is observed best for the PKP phases whose difference is much smaller for the lower mantle plume (Figs. 4.5c, right) than for the upper mantle plume (Fig. 4.4c, right). The differences (DIFF) in Fig. 4.5 are mainly below 0.05 s as predicted above (section 3.1), and they are mostly due to the added Gaussian noise.

4.4 Inversion results

The results of the travel-time inversions A - F are presented as horizontal (Figs. 4.6 and 4.8) and vertical (Figs. 4.7 and 4.9) cross sections through the velocity models. Inversion parameters are listed in Table 1. Since the input structure is well-known, we just briefly comment on the resolution of the inversion procedure.

4.4.1 Resolution limits

The general shapes of the assumed upper and lower mantle plumes are well recovered in all inversions. However, the upper boundary at 100 km depth is not resolved, because there is vertical smearing of the seismic perturbations towards shallow depths along the steeply incoming teleseismic ray paths. Such artificial crustal anomalies are also found in other applications (e.g. Ritter et al., 2001) and demonstrate the need for independent studies, focusing on the shallow seismic structure. The bottom of the synthetic lower mantle plume at 750 km depth is quite well recovered (Fig. 4.9), whereas the upper mantle plume is shifted out of the transition zone (Fig. 4.7). Because a realistic instead of a perfectly even distribution of epicenters is used, the lateral edges of the synthetic plumes are not exactly resolved (Figs. 4.6 to 4.9). This is best seen in the vertical N-S cross sections (Figs. 4.7b and 4.9b) where the southern bottom ends of the plumes are partly truncated, because there are too few rays from southern directions (Fig. 4.2). The upward smearing of the velocity perturbations along the rays paths into the crust leads to an underestimation of the amplitude of the plume-like anomaly itself, because artificial perturbations outside the actual plume are created. The images in Figs. 4.6 to 4.9 are typical for the resolution limits of teleseismic tomography experiments and demonstrate the effects of uneven ray coverage, model parameterization and dominating steeply incident ray paths, which must be kept in mind when interpreting the inversion results.

4.4.2 Comparison of inversion models

The horizontal cross sections in Fig. 4.6 compare the results for inversion models A, B, and C at four different depth levels. The comparison of A with B demonstrates that the travel-time anomalies due to the deflected discontinuity at 410 km depth are too small for significant changes in the distribution of the velocity perturbations, if the same starting model is used. In the depth range from 220 to 270 km (Fig. 4.6a) and 340 to 410 km (Fig. 4.6b) the small differences are below the resolution limit. Underneath the 410 km discontinuity, at 410 to 480 km depth, the small low-velocity anomaly (-1.5 % at 410 km to 0 % at 500 km, Fig. 4.3a) of the upper mantle plume is poorly recovered. Especially in the southern half, where there are very few seismic rays, the anomaly even has the opposite sign. This is better visible in Fig. 4.7. Due to the smearing
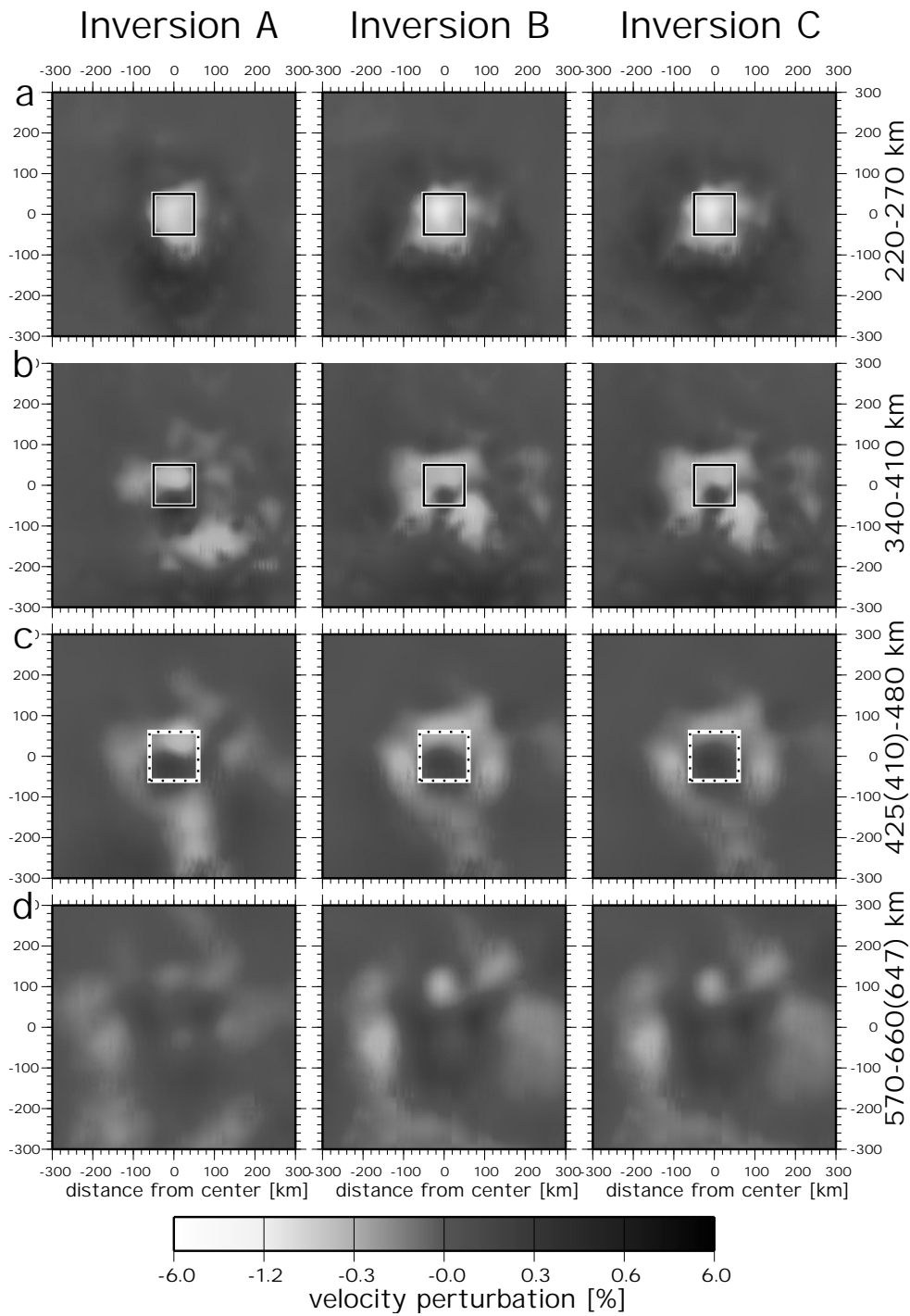
Figure 4.6: Horizontal cross sections through the tomographic models A-C with the upper mantle plume, displaying P-velocity perturbations in percent. The position of the input structure is indicated by a square. a) depth range 220-270 km; b) 340-410 km; c) 410-480 km for A and B, 425-480 km for C; d) 570-660 km for A and B, 570-647 km for C.
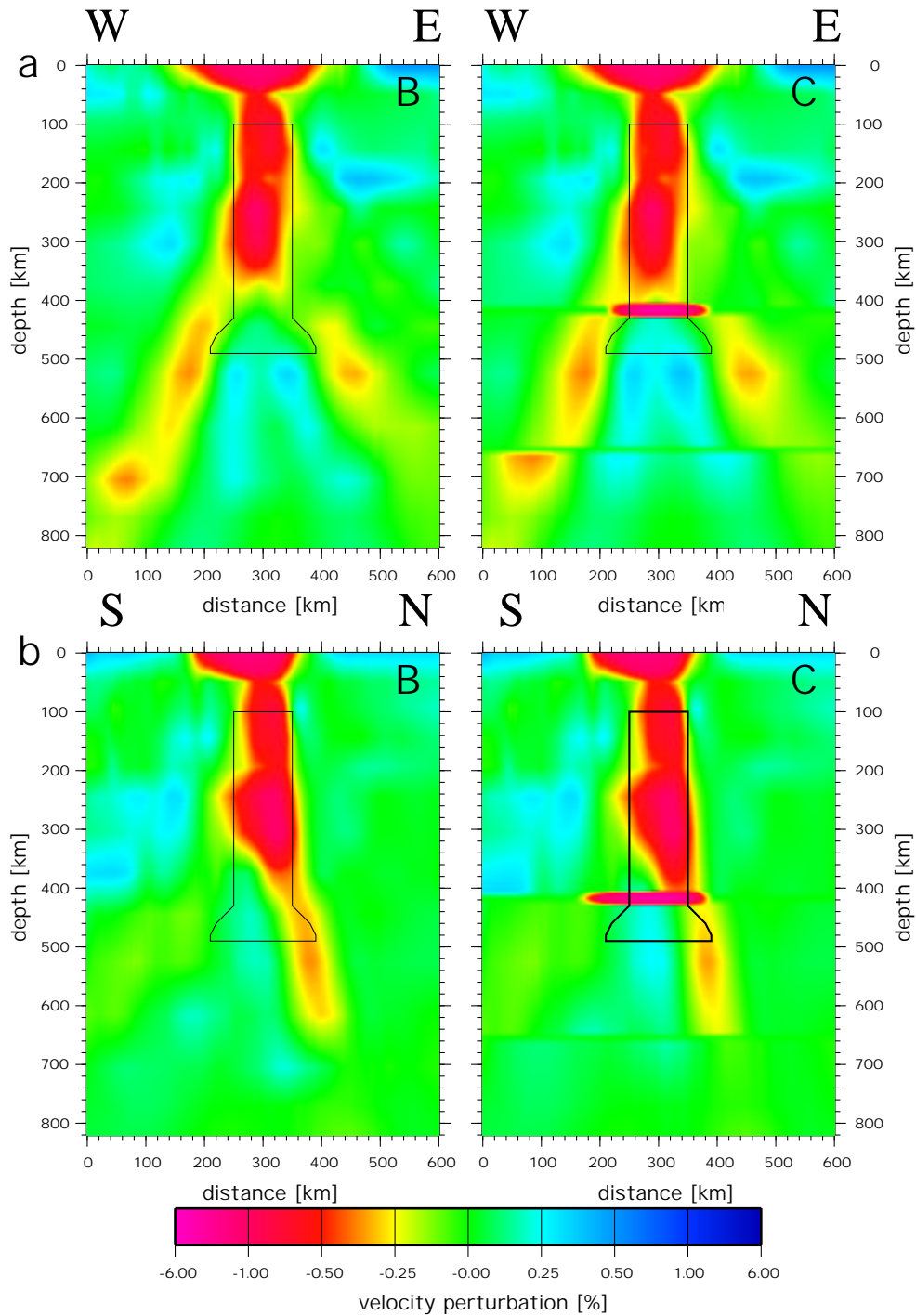
Figure 4.7: Vertical cross sections through tomographic models B (left) and C (right) displaying P-velocity perturbations in percent. The position of the input plume structure is indicated by black lines. a) West-East sections; b) South-North sections. Deviations from the more sharply bounded input plume structure are mainly due to uneven ray coverage and parameterization (see text). The modeled anomaly at crustal depths results from smearing along the subvertical ray paths into poorly resolved regions.
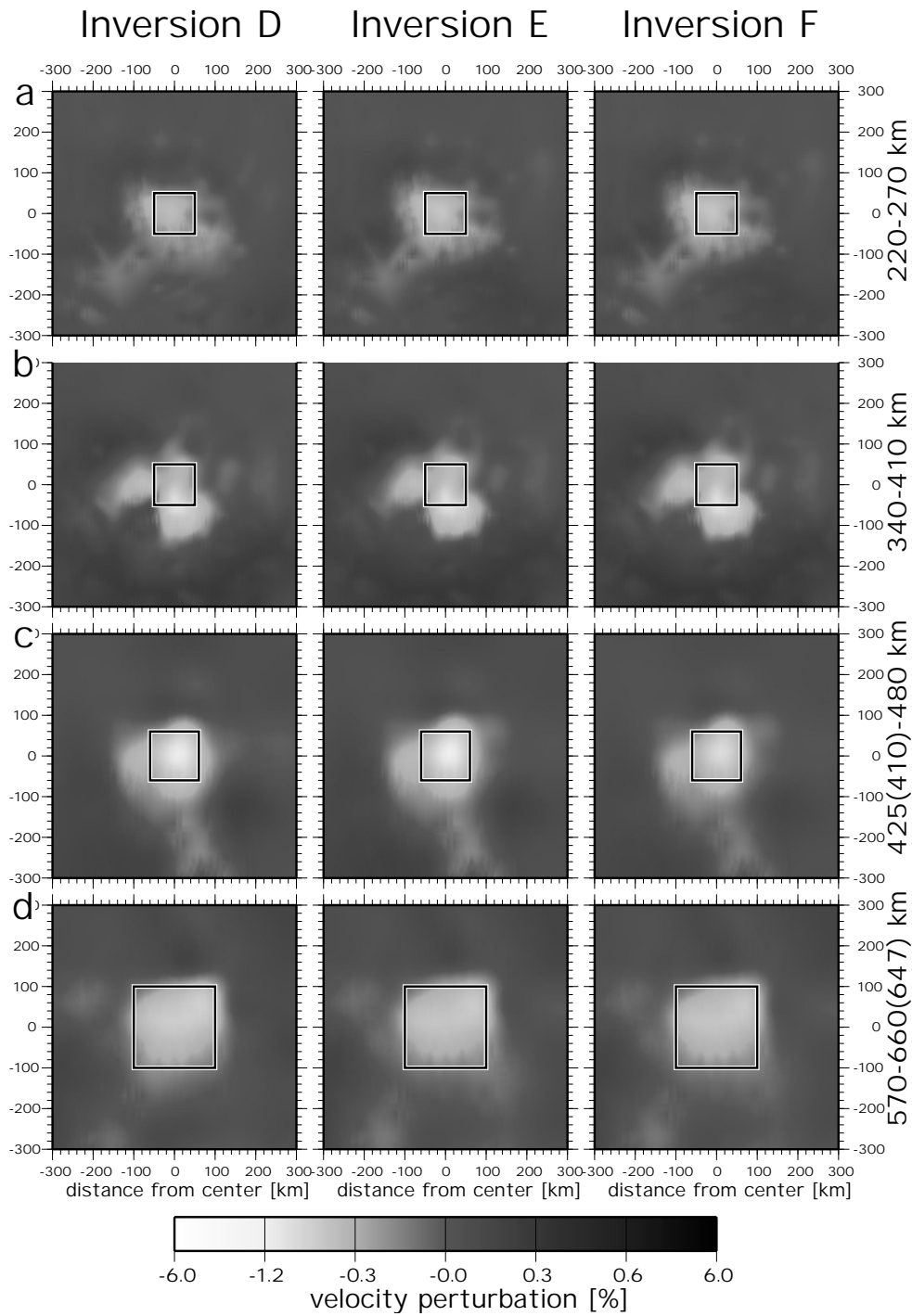
Figure 4.8: Horizontal cross sections through the tomographic models D-F with the lower mantle plume, displaying P-velocity perturbations in percent. The position of the input structure is indicated by a square. a) depth range 220-270 km; b) 340-410 km; c) 410-480 km for A and B, 425-480 km for C; d) 570-660 km for A and B, 570-647 km for C.
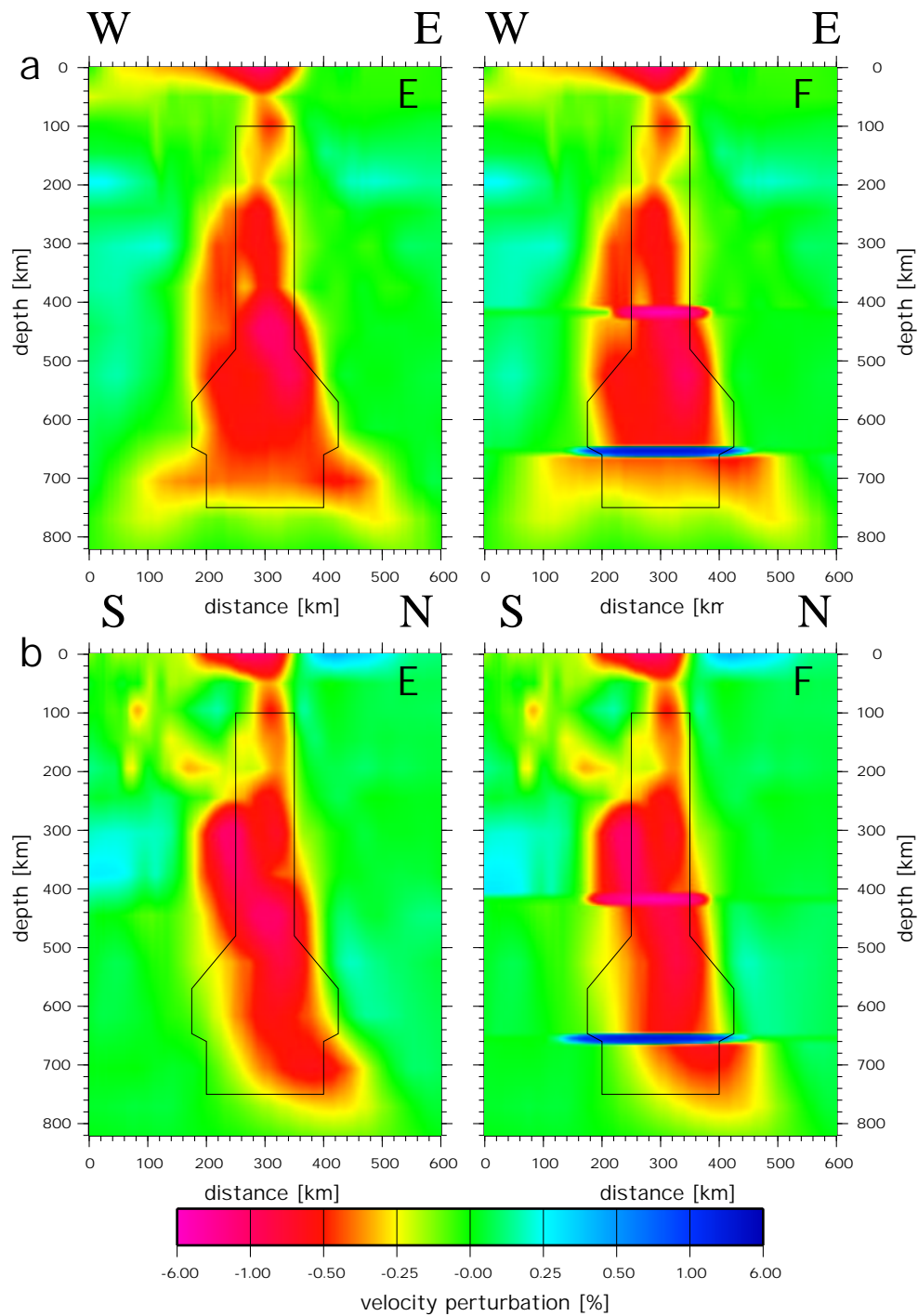
Figure 4.9: Vertical cross sections through tomographic models E (left) and F (right), displaying P-velocity perturbations in percent. The position of the input plume structure is indicated by black lines. a) West-East sections; b) South-North sections.

of the anomaly at shallow depth (above 100 km) the lower end of the upper mantle plume is not correctly recovered. The velocity anomaly of the upper mantle plume is hardly improved by the addition of the a prior information (inversion C) in Figs. 4.6 and 4.7. The constraint that the 410 km discontinuity is lowered to 425 km causes only a 1 % better variance reduction of inversion C compared to B. To compensate the large negative anomaly in the downwarped region at 410 - 425 km depth, below a high-velocity anomaly appears (Fig. 4.6c right) instead of a low-velocity anomaly (-1.5 % to 0 % at 500 km depth). This means that small-scale features with tiny velocity contrast (1 %) in the transition zone are beyond resolution, even when high-quality teleseismic data are used.

In the vertical cross sections (Fig. 4.7) the results from inversions B and C are compared to identify a possible influence of the a priori constraints. In the upper 400 km the images are essentially the same. Even at the 410 km discontinuity, which is fixed in inversion C, the plume edges are not sharper recovered. Below the downwarping of the 410 km discontinuity, which is a strong negative velocity anomaly (-3.6 %), a high-velocity region (+0.3 %) appears to compensate the unresolved deflection.

The horizontal cross sections in Fig. 4.8 compare the results for inversion models D, E, and F. The comparison of D with E shows that the travel-time anomalies due to the deflected discontinuities at 410 and 660 km depth are too small for significant changes in the distribution of the velocity perturbations. Only in the depth range from 570 to 660 km (Fig. 4.8d) a tiny difference is present at the SE edge of the plume which arises from the deflection of the 660 km discontinuity. However, in real applications this feature may be not identified due to its small signal. The application of the a priori information in model F causes only minor modifications of the velocity structure near the transition zone (Fig. 4.9). Directly underneath the 410 km discontinuity at 425 to 480 km depth (Fig. 4.8c) and above the 660 km discontinuity at 570 to 647 km depth (Fig. 4.8d) the input structure is well recovered in its size and velocity contrast. Compared to inversion E there is a 2 % better variance reduction in inversion F (Table 1). This is due to the improved velocity structure around the discontinuities and corresponds to the changes between models E and F which are limited to the transition zone region.

In the vertical cross sections (Fig. 4.9) the results from inversions E and F are compared. In the upper 400 km the images are essentially the same. Note the smearing of structure into the unresolved crustal depth range. Near the discontinuities of the transition zone, which are fixed in inversion F, the plume edges are slightly sharper recovered in F, but the differences are beyond the resolution properties of real applications. The broadening of the plume column in the lower mantle is better resolved in F than in E because of the a priori constraints in the inversion. This is clearly evident in the E-W cross section (Fig. 4.9a). In the N-S cross section (Fig. 4.9b) the lower mantle part of the input structure is not recovered as well as in the E-W cross section due to fewer seismic rays from the south.

4.5   Discussion and conclusions

The synthetic teleseismic experiment and the travel-time tomography described above are able to recover a mantle plume with 1 % to 1.5 % negative velocity perturbation. The horizontal extension of the plumes is well recovered, and the 600 km by 600 km station network is large

enough to prove that the plume is upwelling from the lower mantle and pinches through the transition zone. However, the tomographic results do not resolve the about 15 km vertical deflections of the discontinuities at 410 km and 660 km depth. The images in Figs. 4.6 to 4.9 indicate that the travel-time effect of the deflected areas are too small compared to the much bigger travel-time perturbation of the plume. As consequence an analysis of receiver functions is recommended to find out more details of the transition zone region. Afterwards, a combined inversion of receiver-function results and travel-time data can help to improve slightly the tomographic images. At plumes that possibly penetrate only the 410 km discontinuity and not the 660 km discontinuity (e.g. perhaps Iceland, Foulger et al., 2000) the travel-time effect of the deflected 410 km discontinuity is more obvious, because the opposite travel-time anomaly of the 660 km discontinuity is missing.

Our results have implications for future seismological experiments concerning the design of the networks. First, there is a need for the combined analysis of travel-time residuals and receiver functions. Therefore, the station distribution should be chosen in a way that it is able to measure the data for both types of seismological analyses. Second, tomographic inversion schemes should be able to incorporate as much a priori information as possible. The modeling procedure used for this contribution is recommended to test possible targets of large-scale seismic experiments to optimize the design of a station network, if the approximate distribution of the seismic sources is known.

4.6   Acknowledgments

5 Conclusions

The increasing number of high quality digital mobile seismic instruments and the availability of seismological data via the internet have provided the data base for the derivation of high resolution tomographic images even down into the Earth's lower mantle. This has led to the development of JI-3D, a new method for the iterative joint inversion of teleseismic delaytimes and Bouguer gravity data. The resulting models are the 3D distributions of velocity and density perturbations, relative to standard Earth models, and a 1D model of depth dependent density-velocity correlation parameters. The aim of the inversion method is to achieve inversion results with maximum stability and (spacial) resolution at the same time. Instead of overly regularizing ill posed problems, the philosophy is to optimize the problem itself. This is realized successfully by variably parameterizing the velocity and density model, such that all model parameters are equally well constrained. This results in velocity models with good spacial resolution where the information density is high, whereas in areas where little information can be derived, a parameter rather represents a stable average over a larger region. The block sizes of the density model increase with distance from the observational points, since the influence on the gravity signal decreases. By using a Bayes algorithm, *a priori* information can be incorporated into the inversion. Moreover 3D raytracing (Steck and Prothero (1991)) is applied, since even small velocity changes of e.g. $\pm5\%$ within 10 km distance may influence the ray distribution considerably and wrong raypaths would lead to the misslocalization of the velocity perturbations. The method is applied to a synthetic data set, designed to reveal its advantages and limits.

As expected, the iterative teleseismic delaytime inversion provides good results in well sampled regions of the model space (50 km depth and more) while the upper part (≈ 20 km depth) is affected by vertical smearing, where the resolution of the gravity inversion is optimum. Application of *a priori* information about the uppermost part (≈ 15 km depth) of the model leads to an improved image in the rest of the model.

The full joint inversion of teleseismic delaytimes and Bouguer gravity data does not provide satisfying results, since the smearing into the uppermost layers cannot be reduced. Moreover, the velocity results are deteriorated in the deeper regions of the model space, since the high resolution of the velocity model is destroyed by the correlation to the coarsely parameterized density model. In addition to that, the huge number of parameters makes the inversion hardly manageable in terms of computer power and calculation time. Also the iterative inversion is not possible for the same reason.

Since the main vertical smearing of the delaytime inversion results occurs in the crustal region (≈ 30 km), where the resolution of the Bouguer gravity data is highest, a partial joint inversion is performed. Hence the Bouguer gravity data are highpass filtered, so that they contain only signal that is caused within th uppermost 30 km. Moreover *a priori* information is applied for the uppermost 15 km. Consequently, the joint inversion is only performed in the depth range between 15 km and 30 km. This results in a considerably reduced set of model parameters and a more manageable inversion procedure. The non iterative joint inversion provides good results, however

not significantly improved compared to the pure delaytime inversion with *a priori* information. The velocity and density models corresponding to the iterative partial joint inversion suggest reduced vertical smearing within the upper 50 km which is a clear improvement to the previous inversions. It hast to be mentioned, that all inversion results (except for the full joint inversion) are excellent in greater depths. Finally, it can be concluded, that the method is successful for the pure delaytime inversion, especially if *a priori* information is applied. Partly joint inversion is a promising technique to derive improved inversion models. The applicability of the joint inversion is still limited by computational constraints.

As an application to a real data set, seismic images of the upper mantle below the Quaternary Eifel volcanic fields, Germany, are presented. The data were measured at a dedicated network of more than 200 stations. The results of the non iterative joint inversion show a columnar low P-velocity anomaly in the upper mantle with a lateral contrast of up to 2 %. Since the structure is 100 km wide the corresponding gradient is very small so that non iterative delaytime inversion is applied. The low velocity zone extends to at least 410 km depth and is equivalent to about 150-200 K excess temperature. This clear evidence for a plume below a region of comparatively minor volcanism suggests that deep mantle plumes could be more numerous than commonly assumed. They may often be associated with small volcanic fields or may have no volcanic surface expression at all.

Possible source regions for the genesis of mantle plumes are thermal boundary layers inside the Earth. Therefore, the core-mantle boundary and the transition zone between the lower and the upper mantle are regarded as regions where thermal plumes may start and rise towards the Earth's surface. The positive thermal anomaly associated with mantle plumes causes a negative seismic velocity perturbation and depth changes of the seismic mantle discontinuities. The most important mantle discontinuities in standard Earth models are the temperature- and pressure-sensitive phase boundaries at 410 and 660 km depth. Since the 'Eifel-plume' seems to penetrate the 410 km mantle discontinuity, the influence of deflected discontinuities on teleseismic delaytime inversions is studied. Synthetic teleseismic traveltimes and their relative traveltime residuals are calculated for lower mantle and upper mantle plume models and the forward modeled data are inverted into a 3D velocity perturbation model which well resolves the general plume shape. Because non iterative traveltime tomography alone cannot resolve the velocity perturbations associated with the depth variations of the mantle discontinuities, *a priori* information are added to the inversion problem. Such constraints could be provided by a receiver function analysis with P to S converted phases in real applications. The constrains may add information about the depth variations of the discontinuities. However, the *a priori* information improves only slightly the seismic images, because the plume structure dominates the inversion. As result it is believed, that a combined inversion of seismic traveltime and receiver function analyses is necessary to achieve the best images of the interaction between hot plumes and small-scale mantle structures such as deflections of seismic discontinuities.

6 Appendix

layer n	thickness [km]	layer boundary [km]	node depth [km]	velocity [km/s]	inital least blocks v [km]	final least blocks v [km]	least blocks ρ [km]	number model param. v	number model param. ρ	maximum lateral extension v	maximum lateral extension ρ
1	7	7	0	6,0	10	5	4	37	4156	110	920
2	8	15	14	6,0	20	10	12	38	616	140	960
3	7	22	16	6,0	10	11	21	24	373	160	900
4	8	30	29	6,0	10	10	27	72	189	160	960
5	10	40	31	8,0	15	12	31	29	257	180	920
6	10	50	49	8,0	15	12	40	35	201	200	1130
7	10	60	51	8,0	15	13	49	33	157	240	1270
8	10	70	69	8,0	15	12,5	58	36	105	260	1390
9	10	80	71	8,0	15	14	67	42	137	280	1080
10	10	90	85	8,0	15	12,5	76	46	116	300	1140

Table 6.1: Summary of the parameterization of the velocity (v) and density (ρ) models.
Column 1: number of layer; Columns 2-5: vertical parameterizations; Columns 6 & 7: optimization of the v-model; Column 8: parameterization of the ρ-model; Columns 9 & 10: resulting models in terms of parameters per layer; Columns 11 & 12: maximum lateral extensions of the parameterizations for the v- and ρ-model
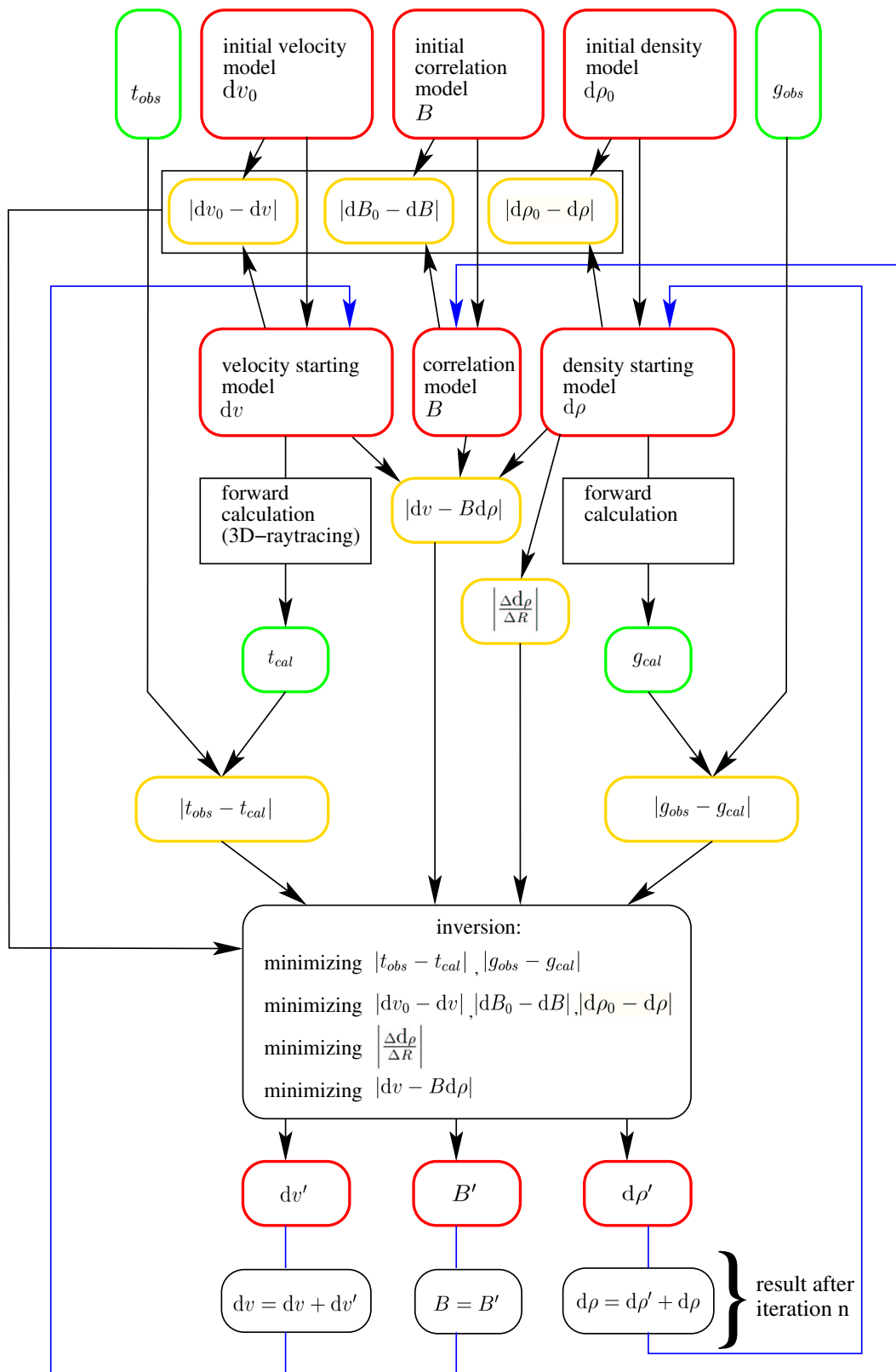
Figure 6.1: *Flowchart of the JI-3D inversion. Data (observed and forward calculated) are indicated by green frames, model parameters correspond to red frames, constraints for the inversion are yellow and calculations and the inversion have black frames.*
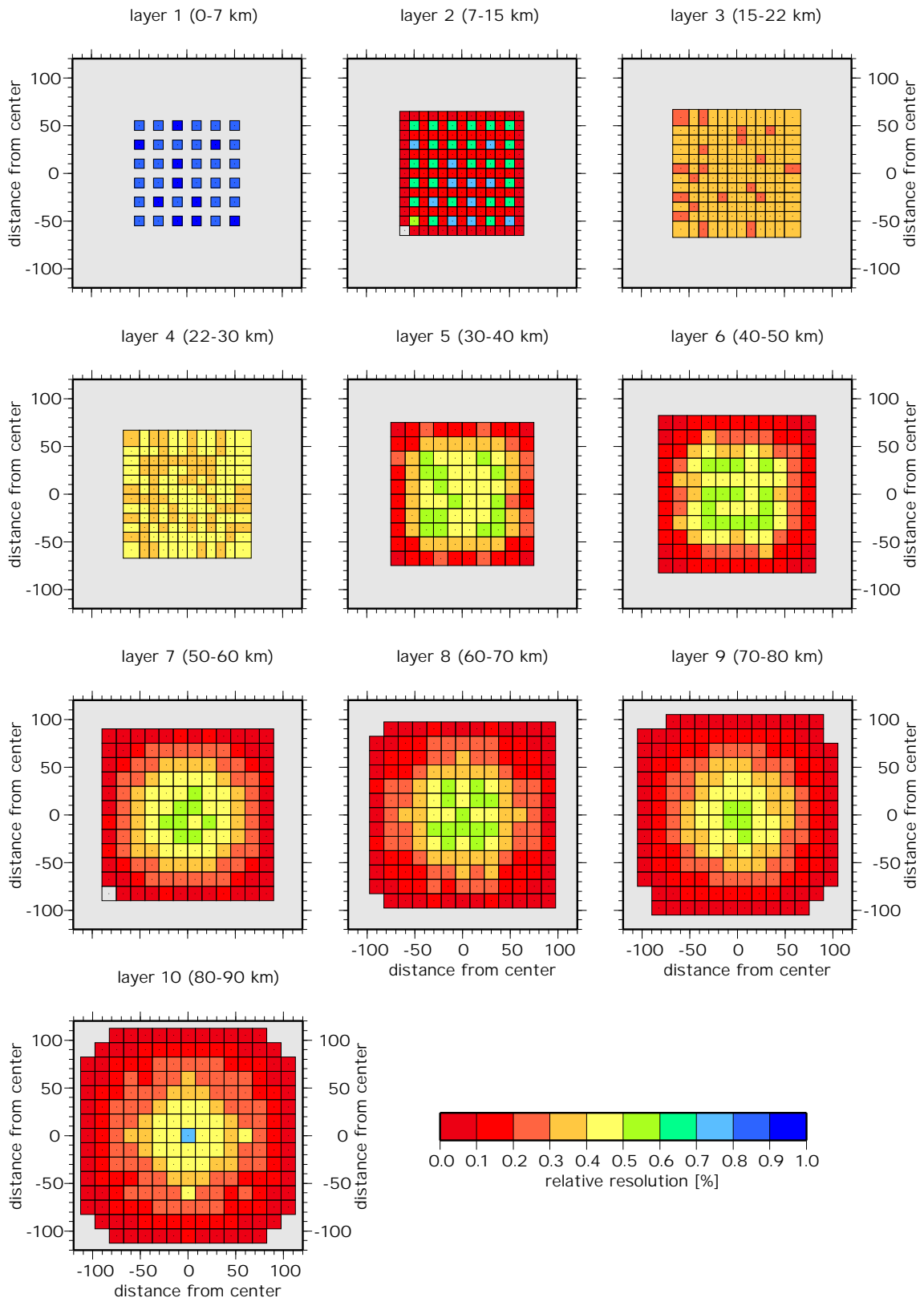
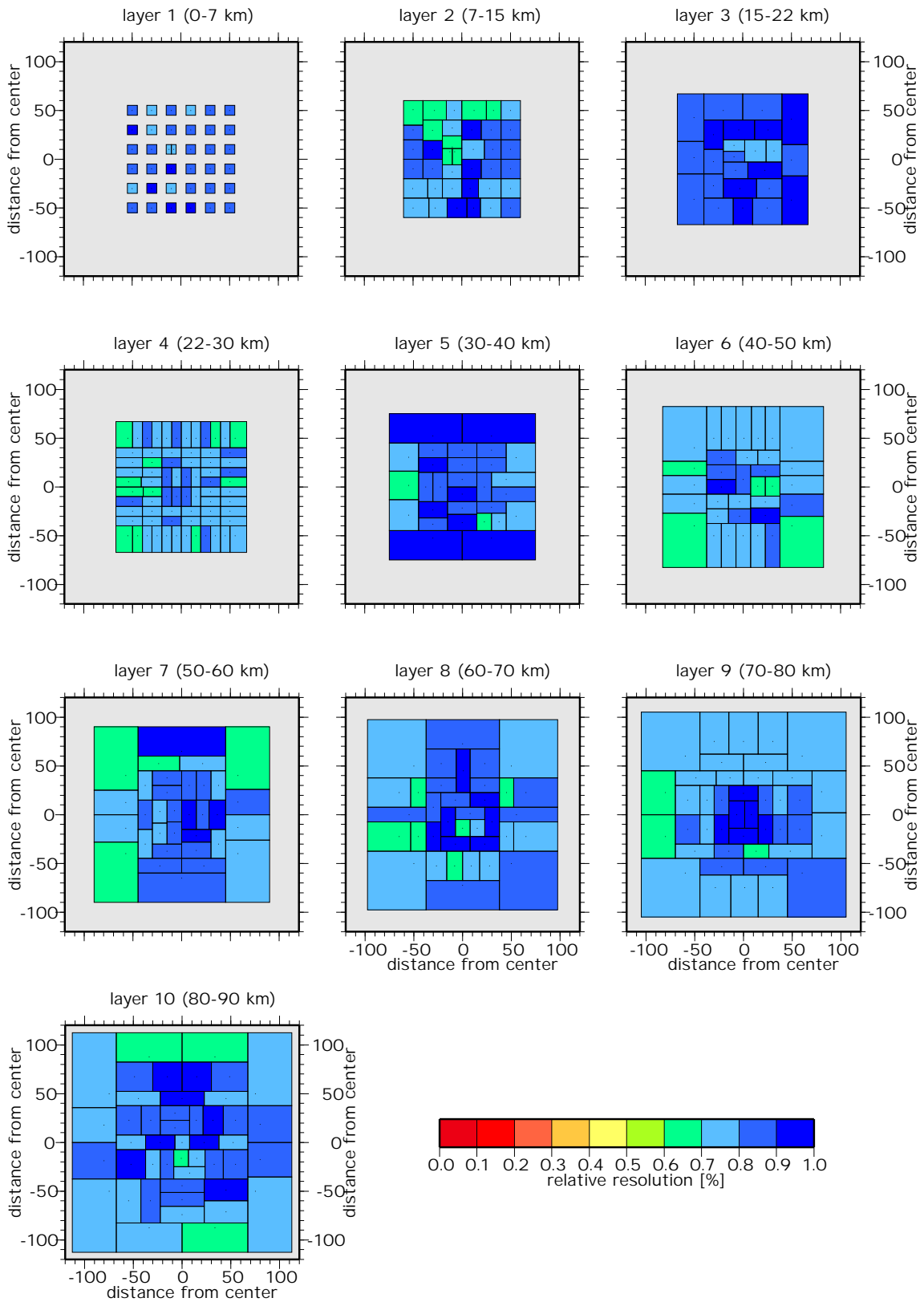Figure 6.2: *initial resolution and parameterization*

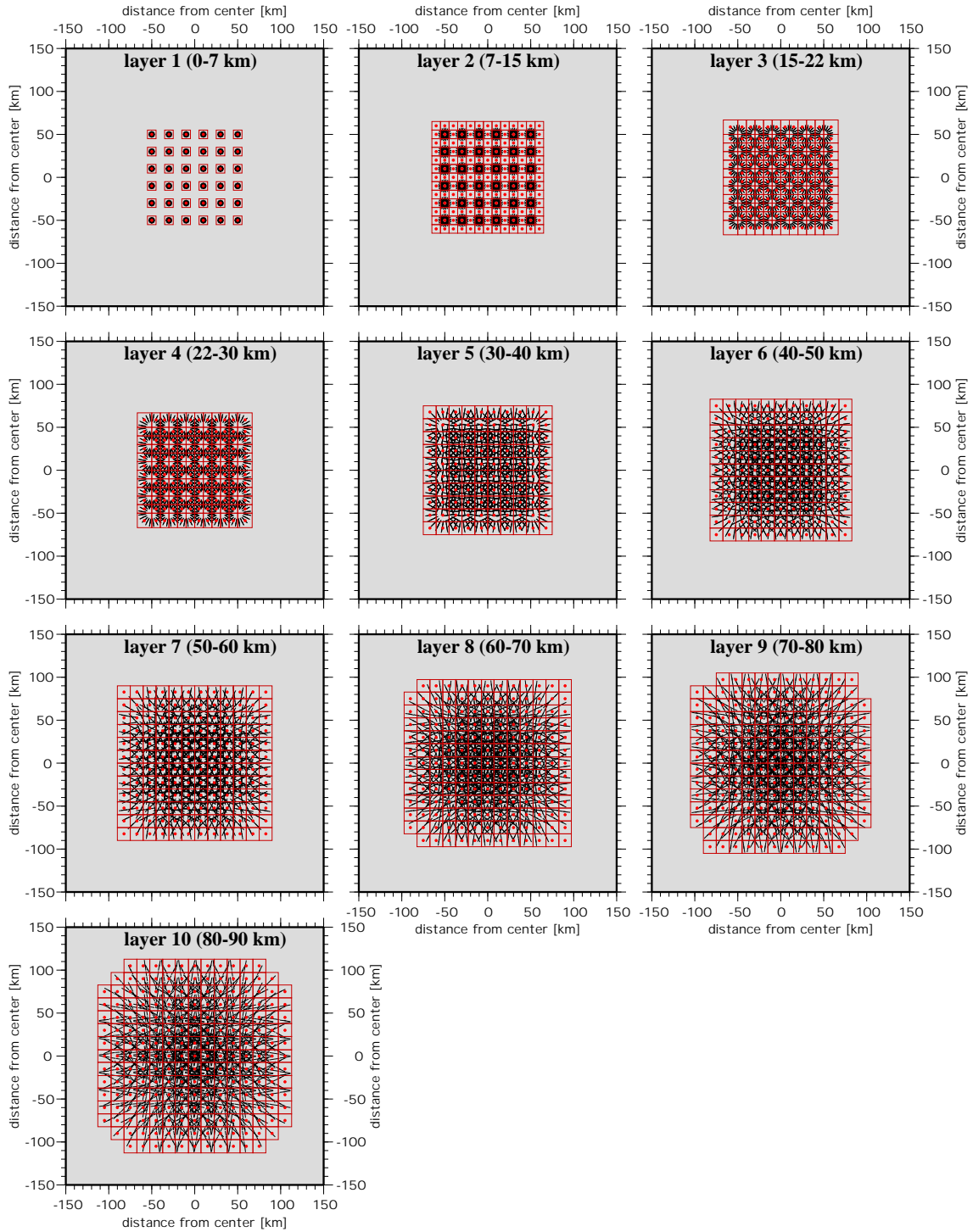Figure 6.3: *final resolution and parameterization*

107

Figure 6.4: *ray distribution*

Bibliography

Aki, K., Christoffersson, A. and Husebye, E., 1977. Determination of the three-dimensional seismic structure of the lithosphere. *J. Geophys. Res.* 82, 277–296.

Aki, K. and Lee, L., 1976. Determination of three dimensional velocity anomalies under a seismic array using first arrival times from local earthquakes. *J. Geophys. Res.* 81, 4381–4399.

Albers, M. and Christensen, U., 1996. The excess temperature of plumes rising from the core-mantle boundary. *Geophys. Res. Lett.* 23, 3567–3570.

Anderson, O., Isaak, D. and Oda, H., 1992. High-temperature elastic constant data on minerals relevant to geophysics. *Rev. Geophys.* 30, 57–90.

Barth, A., 2002. P-Wellen Tomographie des oberen Mantels und der Übergangszone unter Eifel und Zentralmassiv. Master's thesis, Institut für Geophysik, Universität Göttingen.

Berckhemer, H., 1990. Grundlagen der Geophysik. Wissenschaftliche Buchgesellschaft, Darmstadt.

Bina, C. and Helffrich, G., 1994. Phase transition clapeyron slopes and transition zone seismic discontinuity topography. *J. Geophys. Res.* 99, 15853–15860.

Birch, F., 1961. The velocity of compressional waves in rocks to 10 kilobars. *J. Geophys. Res.* 66(7), 2199–2224.

Brunet, D. and Yuen, D., 2000. Mantle plumes pinched in the transition zone. *Earth Planet. Sci. Lett.* 178, 13–27.

Christensen, U., 1998. Dynamic phase boundary topography by latent heat effects. *Earth Planet. Sci. Lett.* 154, 295–306.

Duffy, T. and Anderson, D., 1989. Seismic velocities in mantle minerals and the mineralogy of the upper mantle. *J. Geophys. Res.* 94, 1895–1912.

Duncan, R., Petersen, N. and Hargraves, R., 1972. Mantle plumes, movement of the european plate, and polar wandering. *Nature* 239, 82–85.

Engdahl, E., van der Hilst, R. and Buland, R., 1998. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull. Seism. Soc. Am.* 88, 722–743.

Evans, J. and Achauer, U., 1993. Teleseismic tomography using the ACH method. In H. Iyer and K. Hirahara, editors, *Seismic Tomography, Theory and Practice*, pages 319–360. Chapman and Hall, London.

Evans, J. and Zucca, J., 1988. Active high-resolution seismic tomography of compressional wave velocity and attenuation structure at medicine lake volcano, northern california cascade range. *J. Geophys. Res.* 93, 15015–15036.

Foulger, G. et al., 2000. The seismic anomaly beneath iceland extends down to the mantle transition zone and no deeper. *Geophys. J. Int.* 142, F1–F5.

Goes, S., Spakman, W. and Bijwaard, H., 1999. A lower mantle source for central european volcanism. *Science* 286, 1928–1931.

Gregersen, S. et al., 1999. Important findings expected from europe's largest seismic array. *EOS Transactions AGU* 80, 1&6.

Griesshaber, E., O'Nions, R. and Oxburgh, E., 1992. Helium and carbon isotope systematics in crustal fluids from the eifel, the rhine graben and black forest, f.r.g. *Chemical Geology* 99, 213–235.

Hoernle, K., Zhang, Y.-S. and Graham, D., 1995. Seismic and geochemical evidence for large-scale mantle upwelling beneath the eastern atlantic and western and central europe. *Nature* 99, 213–235.

Iyer, H. and Hirahara, K., editors, 1993. Seismic Tomography, Theory and Practice. Chapman and Hall, London.

Jackson, D. and Matsu'ura, M., 1985. A Bayesian approach to nonlinear inversion. *J. Geophys. Res.* 90, 581–591.

Jordan, M. and Achauer, U., 1999. A new method for the 3-D joint inversion of teleseismic delaytimes and bouguer gravity data with application to the massif central. *EOS Transactions AGU* 80, supplement to no. 46, F696–F697.

Karato, S.-i., 1993. Importance of anelasticity in the interpretation of seismic tomography. *Geophys. Res. Lett.* 20, 1623–1626.

Kennett, B. and Engdahl, E., 1991. Traveltimes for global earthquake location and phase identification. *Geophys. J. Int.* 105, 429–465.

Kennett, B., Engdahl, E. and Buland, R., 1995. Constraints on seismic velocities in the earth from traveltimes. *Geophys. J. Int.* 122, 108–124.

Kertz, W., 1995. Einfhrung in die Geophysik I. Spektrum Akademischer Verlag, Heidelberg.

Lees, J. and VanDecar, J., 1991. Seismic tomography constrained by Bouguer gravity anomalies: Applications in western Washington. *Pageoph* 135, 31–52.

Li, X., Kind, R., Priestley, K., Sobolev, S., Tilmann, F., Yuan, X. and Weber, M., 2000. Mapping the hawaiian plume conduit with converted seismic waves. *Nature* 405, 938–941.

Lippolt, H., 1983. Distribution of volcanic activity in space and time. In Fuchs et al., editors, *Plateau Uplift*, pages 112–120. Springer, Berlin.

Malamud, B. and Turcotte, D., 1999. How many plumes are there? *Earth Planet. Sci. Lett.* 174, 113–124.

Menke, W., 1989. Geophysical data analysis: Discrete inverse theory. Academic Press Inc., San Diego.

Menke, W. and Abbott, D., 1990. Geophysical Theory. Columbia University Press, New York.

Meyer, W. and Stets, J., 1998. Junge tektonik im rheinischen schiefergebirge und ihre quantifizierung. *Z. dt. geol. Ges.* 149, 359–379.

Morgan, W., 1971. Convection plumes in the lower mantle. *Nature* 230, 42–43.

Nataf, H.-C., 2000. Seismic imaging of mantle plumes. *Ann. Rev. Earth Planet. Sci.* 28, 391–417.

Nelder, J. A. and Mead, R., 1985. A simplex method for function minimization. *Computer Journal* 7, 308.

Passier, M. and Snieder, R., 1996. Correlation between shear wave upper mantle structure and tectonic surface expressions: Application to central and southern germany. *J. Geophys. Res.* 101, 25293–25304.

Portmann, M., 2000. Resolutionsvermögen regionaler seismischer tomographie am beispiel des eifel-plume projektes. Master's thesis, Institut für Geophysik und Meteorologie, Universität Köln.

Prothero, W. A., Taylor, W. J. and Eickemeyer, J. A., 1988. A fast, two point, three-dimensional raytracing algorithm using a simple step search method .

Raikes, S. and Bonjer, K.-P., 1983. Large-scale mantle heterogeneity beneath the rhenish massif and its vicinity from teleseismic p-residuals measurements. In Fuchs et al., editors, *Plateau Uplift*, pages 315–331. Springer, Berlin.

Regenauer-Lieb, K., 1998. Dilatant plasticity applied to the alpine collision: ductile void growth in the intraplate area beneath the eifel volcanic field. *J. Geodynamics* 27, 1–21.

Ribe, N. and Christensen, U., 1999. The dynamical origin of hawaiian volcanism. *Earth. Planet. Sci. Lett.* 171, 517–531.

Ritter, J., Achauer, U., Christensen, U. and the Eifel Plume Team, 2000. The teleseismic tomography experiment in the eifel region, central europe: Design and first results. *Seism. Res. Lett.* 71, 437–443.

Ritter, J., Jordan, M., Christensen, U. and Achauer, U., 2001. A mantle plume below the eifel volcanic fields, germany. *Earth Planet. Sci. Lett.* 186, 7–14.

Schmincke, H.-U., Lorenz, V. and Seck, H., 1983. The quaternary eifel volcanic fields. In Fuchs et al., editors, *Plateau Uplift*, pages 139–151. Springer, Berlin.

Shen, Y., Solomon, S., Bjarnason, I. and Wolfe, C., 1998. Seismic evidence for a lower-mantle origin of the iceland plume. *Nature* 395, 62–65.

Smith, W. and Wessel, P., 1990. Gridding with continuous curvature splines in tension. *Geophysics* 55, 293–305.

Sobolev, S. et al., 1997. Upper mantle temperatures and lithosphere-asthenosphere system beneath the french massif central constrained by seismic, gravity, petrologic and thermal observations. *Tectonophysics* 275, 143–164.

Spakman, W. and Bijwaard, H., 1998. Irregular cell parameterization of tomographic problems. *Anales Geoph. Supplement 1* 16, C 28.

Stammler, K., 1993. SeismicHandler - programmable multichannel data handler for interactive and automatic processing of seismological analyses. *Comp. Geosciences* 19, 135–140.

Steck, L. and Prothero, W., 1991. A 3-D raytracer for teleseismic body-wave arrival times. *Bull. Seism. Soc. Am.* 81, 1332–1339.

Tarantola, A. and Valette, B., 1982. Inverse problem = quest for information. *J. Geophys. Res.* 50, 159–170.

Tiberi, C., Diament, M., Dverchre, J., Petit-Mariani, C., Mikhailov, V., Tikhotsky, S. and Achauer, U., in press. Deep structure of the baikal rift zone revealed by joint inversion of gravity and seismology. *J. Geophys. Res.* .

van der Hilst, R., Engdahl, E. and Spakman, W., 1993. Tomographic inversion of p and pp data for aspherical mantle structure below the northwest pacific region. *Geophys. J. Int.* 115, 264–302.

Wedepohl, K. and Baumann, A., 1999. Central european cenozoic plume volcanism with oib characteristics and indications of a lower mantle source. *Contrib. Mineral. Petrol.* 136, 225–239.

Weiland, C., Steck, L., Dawson, P. and Korneev, V., 1995. Nonlinear teleseismic tomography at long valley caldera, using three-dimensional minimum travel time ray tracing. *J. Geophys. Res.* 100, 20379–20390.

Wessel, P. and Smith, W., 1998. New, improved version of generic mapping tools released. *EOS Trans. Amer. Geophys. U.* 79, 579.

White, R., 1993. Melt production rates in mantle plumes. *Phil. Trans. R. Soc. Lond.* A 342, 137–153.

White, R. and McKenzie, D., 1995. Mantle plumes and flood basalts. *J. Geophys. Res.* 100, 17543–17585.

Zeyen, H. and Achauer, U., 1997. Joint inversion of teleseismic delay times and gravity anomaly data for regional structures. In K. Fuchs, editor, *Upper mantle heterogeneities from active and passive seismology*, pages 155–169. Kluwer Academic Publishers, Dordrecht.

Zeyen, H. and Pous, J., 1993. 3-D joint inversion of magnetic and gravimetric data with a prioi information. *Geophys. J. Int.* 112, 244–256.

Lebenslauf

Persönliche Daten

Name:	Michael Jordan
Geburtsdatum:	30. November 1970
Geburtsort:	Miltenberg
Staatsangehrigkeit:	deutsch
Familienstand:	ledig

Schulbildung

1976 - 1980	Grundschule Kleinheubach
1980 - 1990	Johannes-Butzbach-Gymnasium Miltenberg, Abschluß: Abitur

Wehrdienst

1990 - 1991	Luftwaffenausbildungsregiment 1, Goßlar und Infanterieschule Hammelburg

Studium

1991 - 1998	Studium der Geophysik an der Universität Karlsruhe
04 / 1994	Vordiplom Geophysik
11 / 1996 - 5 / 1998	Anfertigung der Diplomarbeit an der Ecole et Observatoire des Sciences de la Terre, Strasbourg, Thema der Arbeit: *„3-D Simultane Inversion von Schwere- und teleseismischen Laufzeitdaten: Weiterentwicklung des Programms INTELE zur Anwendung auf komplexe 3-D seismische Strukturen"*
10 / 1998	Diplomprüfung Geophysik
11 / 1998 - 3 / 2003	Promotion in der Arbeitsgruppe von Prof. U. Christensen. Thema der Arbeit: *„JI-3D - A new approach to high resolution regional seismic tomography: Theory and applications"*

Bibliography

Acknowledgments