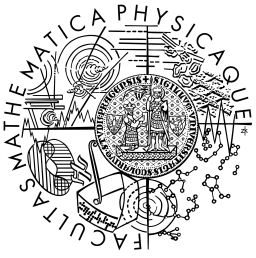

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

DOCTORAL THESIS

Ľubica Valentová

Three-dimensional ambient noise tomography of the Bohemian Massif

Department of Geophysics

Supervisor of the doctoral thesis:	doc. RNDr. František Gallovič, Ph.D.
Study programme:	Physics
Study branch:	Geophysics

Prague 2017

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague November 30, 2017 .....

V prvom rade sa chcem poďakovať školiteľovi doc. RNDr. Františkovi Gallovičovi, Ph.D. za dôsledné vedenie a neskonalé strpenie, a bez ktorého by táto práca nevznikla a nedorástla do súčasnej podoby.

Veľká vďaka patrí dr. Josepovi de la Puente za zasvätenie do SeisSolu, doc. RNDr. Oldřichovi Novotnému, CSc. za oboznámenie a poskytnutie programu VDISP a RNDr. Bohuslavovi Růžkovi, CSc. za poskytnutie spracovaných disperzných kriviek pre našu inverziu.

Ďalej by som sa chcela poďakovať členom Katedry geofyziky za vytvorenie motivačného a priateľského prostredia, a predovšetkým RNDr. Eliške Zábranovej, Ph.D. a RNDr. Jakubovi Velímskému, Ph.D. za to, že mi vytvorili na katedre druhý domov.

Ďalej treba spomenúť mnohých členov FMFI UK v Bratislave, vďaka ktorým bol môj pobyt veľmi poučný.

Poďakovať sa mi treba aj všetkým, ktorí spravujú stanice a poskytli šumové záznamy, najmä sieť CRSN a experimenty PASSEQ a BOHEMA I, II a III.

Vďaka patrí aj všetkým, ktorí tvoria/tvorili ten prvý domov.

Za finančnú podporu ďakujem hlavne projektu QUEST (financovanému vrámci FP7-People, číslo 238007), ďalej projektom SVV-260218 a SVV-260447/2017. Výpočetné prostriedky boli k dispozícii vďaka projektom HPC-EUROPA2 (číslo 228398) a IT4Innovations National Supercomputing Center (LM2015070).

Title: Three-dimensional ambient noise tomography of the Bohemian Massif

Author: Ľubica Valentová

Department: Department of Geophysics

Supervisor: doc. RNDr. František Gallovič, Ph.D., Department of Geophysics

Abstract: We have performed 3D ambient noise tomography of the Bohemian Massif. We invert adopted inter-station dispersion curves of both Love and Rayleigh waves in periods 4–20 s, which were extracted from ambient noise cross-correlations, using a two-step approach. In the first step, the inter-station dispersion curves are localized for each period into the so-called dispersion maps. To account for finite-frequency effects, gradient method employing Fréchet kernels is used. Assuming membrane wave approximation of the surface wave propagation at each period, the kernels were calculated using the adjoint method. To reduce the effect of data noise, the kernels were regularized by Gaussian smoothing. The proper level of regularization is assessed on synthetic tests. In the second step, the phase-velocity dispersion maps are inverted into a 3D S-wave velocity model using the Bayesian approach. The posterior probability density function describing the solution is sampled by more than one million models obtained by Monte-Carlo approach (parallel tempering). The calculated variance of the model shows that the well resolved part corresponds to the upper crust (i.e., upper 20 km). The mean velocity model contains mainly large scale structures that show good correlation with the main geologic domains of the Bohemian Massif. Furthermore, in the well-studied area of Western Bohemia our model agrees well with S-wave velocity models published by other authors.

Keywords: ambient noise tomography, adjoint inversion, Bayesian inversion, Bohemian Massif

Název práce: Trojrozměrná tomografie Českého masivu ze seismického šumu

Autor: Ľubica Valentová

Katedra: Katedra geofyziky

Vedoucí disertační práce: doc. RNDr. František Gallovič, Ph.D.

Abstrakt: Vykonali sme 3D tomografiu Českého masívu vychádzajúcu z prevzatých disperzných kriviek povrchových vĺn v rozsahu periód 4-20 s, ktoré boli získané kroskoreláciou seizmického šumu medzi stanicami. V tomografickej inverzii využívame dvojkrokový postup. V prvom kroku sa disperzné krivky medzi stanicami lokalizujú pre každú periódu zvlášť pomocou gradientovej metódy zahŕňajúcej konečno-frekvenčné efekty pomocou Fréchetovych kernelov; výsledkom sú tzv. disperzné mapy. Na výpočet kernelov sa použila adjungovaná metóda s predpokladom membránovej aproximácie šírenia povrchových vĺn pre jednotlivé periódy. Aby sme potlačili šum v dátach, regularizujeme Fréchetove kernely hladením Gaussovou funkciou, ktorej šírku stanovujeme na základe syntetických testov. V druhom kroku sa disperzné mapy fázových rýchlostí invertujú pomocou Bayesovského prístupu do 3D modelu rýchlostí S-vĺn. Riešenie obrátenej úlohy, definované pomocou aposteriórnej hustoty pravdepodobnosti, je reprezentované viac ako jedným miliónom modelov, ktoré boli získané pomocou metódy Monte Carlo (konkrétne paralelným temperovaním). Vypočítaná variancia modelových parametrov ukazuje, že dobre rozlíšená časť zodpovedá vrchnej kôre (t.j. horných 20 km). Výsledný štatisticky priemerný rýchlostný model vykazuje najmä dlhovlnné štruktúry, ktoré korelujú s hlavnými geologickými oblasťami Českého masívu. Okrem toho sa náš model zhoduje s publikovanými modelmi rýchlostí S-vĺn pre dobre preskúmanú oblasť západných Čiech.

Klíčová slova: tomografia zo seizmického šumu, adjungovaná metóda, Bayesovská inverzia, Český masív

Contents

Introduction

The term "seismic tomography" covers a wide range of geophysical inverse problems that use data derived from seismic records to infer the subsurface structure of the Earth. The result of seismic tomographic problem is represented by seismic velocity model, occasionally other elastic parameters or density. Basically, the solution of the inverse problem is given by two different approaches – deterministic in which the solution is represented by one model best fitting the data while meeting some apriori condition (also called regularization); or stochastic when the solution is defined as posterior probability density function on model parameters (Tarantola (2005)). The posterior probability density function is inferred as apriori information combined with conditional probability from data measurement in the so-called Bayesian framework. The problem is usually solved using Monte Carlo algorithm generating model samples following the posterior probability distribution. Despite the fact that probabilistic approach is considered more general than the single model solution (usually corresponding to the maximal probability), deterministic approach has been favored for many years mainly due to computational reasons.

The first seismic models of the Earth were derived as early as at the beginning of the 20th century, spherical symmetry of the Earth was assumed leading to radial models with main discontinuities at core-mantle boundary and later at inner core boundary (Gutenberg and Richter (1939); Jeffreys and Bullen (1940)). The first works on seismic tomography in traditional sense (i.e., employing first arrival times of seismic waves combined with ray theoretical calculation) originated much later, in the 1970s: on regional scale Aki and Lee (1976); Aki et al. (1977) developed the so-called ACH method, and on global scale Dziewonski et al. (1977) imaged velocity structure in the Earth's mantle for low spherical degree. In broader sense, the seismic tomography includes also inversion of Earth's free oscillations spectra combined with normal mode theory, inversion of surface wave dispersion curves to infer vertical velocity model or different inversions to infer seismic attenuation. As a result of signigicant improvement in both instrumental development and data processing techniqes as well as the increasing computational power, the seismic tomographic methods have undergone a great progress. Present-day applications aim to employ the most of the recorded data in the so-called full-waveform inversion (Fichtner et al. (2013); Tape et al. (2010); Virieux and Operto (2009); Bozdağ et al. (2011)), which requires calculation of the full elastodynamic equation.

Although usually the input data are body wave traveltimes, the most prominent signal on seismograms comes from the surface waves. The studies utilizing the surface wave data originate in the 1950s when the matrix method of calculating surface wave dispersion curves in layered media was developed (Thomson

(1950); Haskell (1953)). To obtain the dispersion data independent on the source, the dispersion curves are measured at two stations aligned on a profile with the epicenter and their differences are then used to obtain a mean vertical model between the stations (the so-called two-station method). Nowadays, the surface wave inversion is performed taking one of the following paths: the surface waveform is inverted directly into the 3D S-wave velocity model using gradient (or Fréchet) kernels (e.g., calculated assuming Born approximation, Snieder (1988); Zhao et al. (2000); Zhou et al. (2006)), or by the so-called two-step approach: in the first step the dispersion measurements are localized for each period and a set of phase velocity maps is constructed; in the second step the dispersion curves extracted from the maps in each point are inverted into a vertical model, finally to be assembled into a 3D model (Nataf et al. (1986); Montagner (1986)).

Most of the seismic tomographic apllications utilize data from close or distant earthquakes. The resulting data coverage thus depends not only on the station distribution of the studied domain but also on the distribution of sources. This may represent a great disadvantage in regional studies of seismically not active regions. For regional tomography, either data coming from the teleseismic events are used which are longer period data and crustal corrections are required. Another possibility to improve the data coverage is to use controlled sources. An important breakthrough was discovery that the information about the structure may be extracted from ambient seismic noise. In particular by crosscorrelating random/diffuse wavefields at two points, Green's function is obtained (Shapiro and Campillo (2004); Campillo and Paul (2003)). Thus long-term ambient noise measurements at different seismic stations are used and the data coverage and resolution of the tomographic inversion depends only on the station distribution. Moreover, the strongest signal in the ambient noise Green's functions is observed for surface waves at periods below 20s – thus the data is usually complimentary to the earthquake data. Besides data coverage another advantages of the noise data are, for example, the possibility to employ same methods as for tomographic inversions of the earthquake data or removing the effects of sources. The tomographic inversions employing ambient noise surface wave data (Ritzwoller et al. (2011); Saygin and Kennett (2012); Lin et al. (2013*b*)) and also the ones with successful extraction of body wave data (Nakata et al. (2015)) have been performed and the resulting models agree well with geological interpretation. Furthermore, the temporal changes in the ambient noise data have also been utilized in the so-called 4D tomography, e.g. under a volcano (Brenguier et al. (2007)).

In this thesis, we employ ambient noise dispersion curves in a two-step surface wave inversion. In the first step, the adopted inter-station dispersion curves are inverted for each period into a set of phase-velocity dispersion maps using 2D adjoint method (Gauthier et al. (1986); Tromp et al. (2005); Fichtner et al. (2006)). The adjoint method is a gradient method with calculation of the finite-frequency Fréchet kernels. To employ the adjoint method, the standard membrane wave approximation is assumed for the surface wave propagation at each frequency (Tanimoto (1990); Peter et al. (2007)). In the second step, the phase-velocity dispersion maps are inverted into a 3D S-wave velocity model in a Bayesian framework. In this part, Monte Carlo algorithm is combined with the dispersion curve calculation in layered media using matrix method (Novotný (1999)).

In both steps, nonlinear methods are employed which may lead to nonunique solutions. Therefore, regularization is employed. It should be noted that, besides explicit regularization (e.g., by smoothing of the resulting model), the implicit regularization (e.g., choice of the model parametrization) is also involved. The regularization setting has an important influence on the resulting model and one must take great care in its choice. For this task, the synthetic tests are usually of great service. Therefore, we discuss them in larger extent.

We solve the ambient noise tomography in the region of the Bohemian Massif, Variscan orogen with complex structure and history (see overview in Matte (2001); Franke (2000)). It is comprised of several main tectonic domains (in particular Saxothuringian, Teplá-Barrandian, Moldanubian, Brunia and Sudetes) whose origin was traced to Neoproterozoic. The individual domains were part of subduction processes leading to continental collision but were subject to different development and thus show different properties. For instance, the Moldanubian domain consists of medium to high metamorphic rocks whereas the adjacent Teplá-Barrandian domain shows rather low metamorphism. Numerous granitoid intrusions have been emplaced during the orogeny, for example, magmatic arc of the Saxothuringian subduction beneath the Teplá-Barrandian domain (Central Bohemian Plutonic Complex) or in the Moldanubian domain as a result of increased heat flow (Central Moldanubian Plutonic Complex). After the Variscan orogeny, the elevated orogen was subject to erosion. Most of the eroded material was transferred to the north to form the Polish Basin leaving the Bohemian Massif rather exposed. During the Alpine orogeny, several of the domain boundaries were reactivated which led to formation of younger sedimentary basins (e.g., Central Bohemian Cretaceous Basin). Boundary between the Saxothuringian and the Teplá-Barrandian domain is active even nowadays in the so-called Eger Rift which exhibits regular earthquake swarms and CO_2 emanations (Fischer et al. (2014)). Although many of the orogenic processes are known from geologic studies, the subsurface seismic properties of the Bohemian Massif from ambient noise data may give additional insight into its development and structure.

The dissertation is organized as follows. Chapter 1 introduces the adjoint method, its derivation for seismic tomographic problem, its basic properties and finally its performance in the synthetic checkerboard tests. Since the checkerboard tests do not serve as a sufficient represantative for the real-data inversion, the synthetic tests employing a simple smooth model and complex heterogeneous model are performed in Chapter 2 (also published as Valentová et al. (2015)). With the help of these tests, the regularization parameters are chosen to be employed in the inversion of inter-station dispersion curves into phase velocity maps. In Chapter 3, the Bayesian inversion of the dispersion maps into 3D S-wave velocity model is introduced and demonstrated using 1D synthetic tests. Chapter 4 (published as Valentová et al. (2017)) displays the results of both steps – phase velocity maps as well as the 3D S-wave velocity model of the Bohemian Massif. The geologic interpretation with the help of several 2D cross-sections corresponding to the reflection and refraction profiles is also presented. In Chapter 5, we additionally discuss the results of the inversion by comparing the synthetic data calculated in the resulting models with the input data as well as the resulting 1D S-wave velocity models with those by other authors.

Chapter 1

Adjoint tomography

For many years, the seismic tomographic studies have been based on the ray theory: the traveltime of seismic wave between the source and the receiver depends only on the velocity distribution along the ray. Although the ray theory represents the high-frequency appoximation of the elastodynamic wave equation, it has turned out to be sufficient enough to explain the fundamental information in the measured seismograms. The method has played an important part in uncovering the Earth's interior: for example, detecting the most important interfaces (core-mantle boundary, inner-core boundary, the Mohorovičić discontinuity), obtaining the radial Earth models and even 3D models of the mantle mapping the subducting plates in the lower mantle (van der Hilst et al. (1997); Romanowicz (2003)). These findigs represent important constraints for geodynamic application (e.g., numerical modelling of mantle convection), which gives insight into the processes occuring hundreds of kilometers beneath the Earth's surface.

Despite the numerous advantages of the ray methods, it was apparent that when employing the lower frequency data (i.e. longer-wavelength waves), the concept of infinitely thin rays decribing the propagation of seismic waves may become inadequate. Therefore, a more general approach that accounts for the volumetric dependency of wave propagation on seismic properties of media (similar to Fresnel zone in optics) is essential. This is represented by finite-frequency or sensitivity kernels reflecting the effect of each point on the wave propagation (Marquering et al. (1999); Dahlen et al. (2000); Tromp et al. (2005)). The so-called finite-frequency tomography employing the sensitivity kernels is then able to invert not only the first arrivals but the whole waveforms. Furthermore, the data coverage as well as the resolution is improved. The finite-frequency tomography was thus successful to image, for example, a mantle plume beneath Hawaii (Montelli et al. (2004)).

The sensitivity kernel is related to a wavefield perturbation with respect to the perturbation of a model parameter. However, to calculate directly the perturbation with respect to every model parameter may represent a computational challenge. One of the possible approaches to obtain the sensitivity kernel is the so-called adjoint method. The adjoint method computes the sensitivity kernel by performing only 2 calculations: one forward calculation of the wavefield propagation from source to receiver and one adjoint calculation (in our application this refers to the backward propagation of the wavefield from receiver to the sources). By combining these two wavefields, the sensitivity kernel of the model parameter

may be obtained. The sensitivity kernels are then used in the iterative improvement of the model parameters, i.e. in the gradient method of the tomographic inversion.

The utilization of adjoint method in seismic tomography is rather unintuitive and its derivation is thus difficult to grasp. Therefore, this chapter is devised to introduce the method step-by-step. In Section 1.1, basic derivation of classical traveltime ray tomography is reminded. In Section 1.2, a very simple example of finite-frequency traveltime tomography, that is adjoint method for membrane wave problem, is presented. This approach is later used in Chapter 2 for real-data tomographic inversion of surface waves. To provide the reader with a more formal and general introduction of the adjoint approach, in Section 1.3 it is derived for arbitrary linear differential operator and misfit. In addition, the adjoint method is specified for problem of 3D elastic wave propagation. The basic properties of sensitivity kernels for various misfits and frequency ranges are shown in Section 1.4. In Section 1.6, the implementation of the adjoint method in gradient tomographic inversion is described and employed in synthetic checkerboard tests.

1.1 Traveltime ray tomography

Let us remind the basic theory of the classical traveltime ray tomography, which is later generalized for finite frequencies. The main objective of the classical traveltime tomography is to minimize the traveltime misfit:

$$\chi = \frac{1}{2}\sum_i \delta T_i^2, \tag{1.1}$$

where (further we omit the subscript i) δT is the traveltime residual defined as the difference between measured traveltime T_{obs} and theoretical traveltime calculated for some reference model in the sense of the first arrival of high-frequency seismic waves,

$$\delta T = T_{\text{obs}} - T_{\text{ref}}. \tag{1.2}$$

In the classical ray theory, one assumes that the traveltime depends on the velocity distribution $c(\mathbf{x})$ only along the infinitely thin ray, leading to

$$\delta T = \int_{\text{ray}} \frac{1}{c(\mathbf{x})}\mathrm{d}s - \int_{\text{ray}_{\text{ref}}} \frac{1}{c_{\text{ref}}(\mathbf{x})}\mathrm{d}s. \tag{1.3}$$

Next, one assumes that the traveltime residual is sufficiently small, caused only by small perturbations of the velocity δc with respect to the reference model, i.e. $c(\mathbf{x}) = c_{\text{ref}}(\mathbf{x}) + \delta c(\mathbf{x})$. Hence, the real ray is approximated by the reference one, resulting in

$$\delta T = \int_{\text{ray}_{\text{ref}}} \left(\frac{1}{c_{\text{ref}}(\mathbf{x}) + \delta c(\mathbf{x})} - \frac{1}{c_{\text{ref}}(\mathbf{x})} \right) \mathrm{d}s, \tag{1.4}$$

or

$$\delta T = \int_{\text{ray}_{\text{ref}}} \frac{-\delta c(\mathbf{x})}{c_{\text{ref}}^2(\mathbf{x}) \left(1 + \frac{\delta c(\mathbf{x})}{c_{\text{ref}}(\mathbf{x})}\right)} \mathrm{d}s, \tag{1.5}$$

which is linearized in the denominator

$$\delta T = \int_{\mathrm{ray}_{\mathrm{ref}}} \left(-\frac{\delta c(\mathbf{x})}{c^2_{\mathrm{ref}}(\mathbf{x})} + o(\delta c^2(\mathbf{x})) \right) \mathrm{d}s. \tag{1.6}$$

Finally, the terms $o(\delta c^2(\mathrm{x}))$ are omitted to obtain the linear relation between traveltime residual δT and velocity perturbation $\delta c(\mathbf{x})$ along the ray

$$\delta T = \int_{\mathrm{ray}_{\mathrm{ref}}} -\frac{\delta c(\mathbf{x})}{c^2_{\mathrm{ref}}(\mathbf{x})} \mathrm{d}s. \tag{1.7}$$

Intuitively, this relationship would be generalized by assuming dependence of the traveltime residual on the velocity perturbations in the whole 3D domain introducing a weighting function $K(\mathbf{x})$

$$\delta T = \int_V K(\mathbf{x}) \delta c(\mathbf{x}) \mathrm{d}V. \tag{1.8}$$

The function $K(\mathbf{x})$ is usually called (sensitivity) kernel.

It is clear, that for the classical ray tomography, the kernel takes form

$$K(\mathbf{x}) = -\frac{1}{c^2_{\mathrm{ref}}(\mathbf{x})} \delta(\mathbf{x} - \mathbf{x}_{\mathrm{ray}_{\mathrm{ref}}}), \tag{1.9}$$

where $\delta(\mathbf{x} - \mathbf{x}_{\mathrm{ray}_{\mathrm{ref}}})$ represents Dirac delta function along the reference ray.

1.2 Adjoint method for simple example

We now derive the adjoint method for kernel calculation for a simple example of membrane wave (scalar wave in elastic isotropic inhomogeneous media). Here, we follow the derivation by Fichtner et al. (2006). Alternative derivation employing the Lagrange multipliers method was done by Liu and Tromp (2006), see also Section 1.3.

Firstly, assume the cross-correlation traveltime misfit

$$\chi = \frac{1}{2} \sum_i \Delta T_i^2, \tag{1.10}$$

where (subscript i is further omitted without loss of generality) ΔT is the traveltime obtained from cross-correlation of the observed ($u_{\mathrm{obs}}(t)$) and synthetic waveform ($u(t)$), i.e.

$$\Delta T = \arg\max_t \int u(\tau) u_{\mathrm{obs}}(t+\tau) d\tau. \tag{1.11}$$

Furthermore, the wavefield u satisfies the equation for membrane wave,

$$\ddot{u}(\mathbf{x}, t) - \nabla \cdot \left(c^2(\mathbf{x}) \nabla u(\mathbf{x}, t) \right) = f(\mathbf{x}, t) \tag{1.12}$$

for all $\mathbf{x} \in \Omega$ and all t, where $c(\mathbf{x})$ is the velocity of the elastic wave.

The initial and boundary conditions are assumed, for simplicity as

$$\begin{aligned} u(\mathbf{x},t_0) &= 0, && (1.13)\\ \dot{u}(\mathbf{x},t_0) &= 0, && (1.14)\\ \text{and } u(\mathbf{x},t) &= 0, \quad \text{for } \mathbf{x}\in\partial\Omega && (1.15)\end{aligned}$$

Eq. (1.12) can be expressed alternatively as

$$\int_\Omega\int_t [\ddot{u}(\mathbf{x},t) - \nabla\cdot\left(c^2(\mathbf{x})\nabla u(\mathbf{x},t)\right) - f(\mathbf{x},t)]\phi(\mathbf{x},t)\mathrm{d}V\mathrm{d}t = 0 \tag{1.16}$$

for arbitrary $\phi(\mathbf{x},t)$.

Condition (1.16) may be added to the misfit calculation for one crosscorrelation traveltime ΔT:

$$\chi = \frac{1}{2}\Delta T^2 + \int_\Omega\int_t [\ddot{u}(\mathbf{x},t) - \nabla\cdot\left(c^2(\mathbf{x})\nabla u(\mathbf{x},t)\right) - f(\mathbf{x},t)]\phi(\mathbf{x},t)\mathrm{d}V\mathrm{d}t, \tag{1.17}$$

or also

$$\begin{aligned}\chi = \int_\Omega \Bigg[&\frac{1}{2}\Delta T^2\delta(\mathbf{x}-\mathbf{x}_{\mathrm{rec}}) + \\ &+ \int_t [\ddot{u}(\mathbf{x},t) - \nabla\cdot\left(c^2(\mathbf{x})\nabla u(\mathbf{x},t)\right) - f(\mathbf{x},t)]\phi(\mathbf{x},t)\mathrm{d}t\Bigg]\mathrm{d}V, \end{aligned}\tag{1.18}$$

where $\delta(\mathbf{x}-\mathbf{x}_{\mathrm{rec}})$ is delta function in the reciever.

The variational derivate of the misfit with respect to the model parameter $c(\mathbf{x})$ is

$$\begin{aligned}\delta_c\chi = \int_\Omega \Bigg\{&\Delta T\delta_c(\Delta T)\delta(\mathbf{x}-\mathbf{x}_{\mathrm{rec}}) + \\ &+ \int_t [\delta_c\ddot{u}(\mathbf{x},t) - \nabla\cdot(2c(\mathbf{x})\delta c\nabla u(\mathbf{x},t)) - \nabla\cdot\left(c^2(\mathbf{x})\nabla\delta_c u(\mathbf{x},t)\right)]\phi(\mathbf{x},t)\mathrm{d}t\Bigg\}\mathrm{d}V. \end{aligned}\tag{1.19}$$

In the first term of (1.19), the cross-correlation traveltime perturbation may be expressed as (derived later in Section 1.4.1)

$$\delta_c(\Delta T) = -\frac{\int_t \dot{u}_{\mathrm{obs}}(\mathbf{x},t+\Delta T)\delta_c u(\mathbf{x},t)\mathrm{d}t}{\int_t \dot{u}(\mathbf{x},\tau)\dot{u}_{\mathrm{obs}}(\mathbf{x},\tau+\Delta T)\mathrm{d}\tau}. \tag{1.20}$$

The denominator in (1.20) represents wavefield normalization, we will denote it N.

Putting (1.20) into (1.19), the misfit derivative is

$$\begin{aligned}\delta_c\chi = \int_\Omega\int_t \Bigg[&-\Delta T\frac{1}{N}\dot{u}_{\mathrm{obs}}(\mathbf{x},t+\Delta T)\delta_c u(\mathbf{x},t)\delta(\mathbf{x}-\mathbf{x}_{\mathrm{rec}}) + \\ &+ [\delta_c\ddot{u}(\mathbf{x},t) - \nabla\cdot(2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)) - \nabla\cdot\left(c^2(\mathbf{x})\nabla\delta_c u(\mathbf{x},t)\right)]\phi(\mathbf{x},t)\Bigg]\mathrm{d}t\mathrm{d}V. \end{aligned}\tag{1.21}$$

Now, we would like to join all terms containing the variation of wavefield $\delta_c u(\mathbf{x}, t)$. However, except for the first term of (1.21), $\delta_c u$ is subject to temporal or spatial differentiation and some calculus is necessary.

In the second term of (1.21), per-partes integration in time is applied twice

$$\int_t \delta_c \ddot{u}(\mathbf{x}, t)\phi(\mathbf{x}, t)\mathrm{d}t = \Big[\delta_c \dot{u}(\mathbf{x}, t)\phi(\mathbf{x}, t)\Big]_{t_0}^{t_1} - \Big[\delta_c u(\mathbf{x}, t)\dot{\phi}(\mathbf{x}, t)\Big]_{t_0}^{t_1} + \int_t \delta_c u(\mathbf{x}, t)\ddot{\phi}(\mathbf{x}, t)\mathrm{d}t. \tag{1.22}$$

For $t = t_0$, the initial conditions (1.13)–(1.14) for $\delta_c u$ may be applied. To remove the terms for final t_1, terminal conditions for ϕ are defined:

$$\phi(\mathbf{x}, t_1) = 0, \tag{1.23}$$
$$\dot{\phi}(\mathbf{x}, t_1) = 0, \tag{1.24}$$

yielding

$$\int_t \delta_c \ddot{u}(\mathbf{x}, t)\phi(\mathbf{x}, t)\mathrm{d}t = \int_t \delta_c u(\mathbf{x}, t)\ddot{\phi}(\mathbf{x}, t)\mathrm{d}t \tag{1.25}$$

For the last term of (1.21), firstly rule for divergence of product is assumed:

$$\nabla \cdot \left(c^2(\mathbf{x})\nabla\delta_c u(\mathbf{x}, t)\right)\phi(\mathbf{x}, t) = \nabla \cdot \left(c^2(\mathbf{x})\phi(\mathbf{x}, t)\nabla\delta_c u(\mathbf{x}, t)\right) - \\ - c^2(\mathbf{x})\nabla\phi(\mathbf{x}, t) \cdot \nabla\delta_c u(\mathbf{x}, t), \tag{1.26}$$

and similarly again

$$\nabla \cdot \left(c^2(\mathbf{x})\nabla\delta_c u(\mathbf{x}, t)\right)\phi(\mathbf{x}, t) = \nabla \cdot \left(c^2(\mathbf{x})\phi(\mathbf{x}, t)\nabla\delta_c u(\mathbf{x}, t)\right) - \\ - \nabla \cdot \left(c^2(\mathbf{x})\nabla\phi(\mathbf{x}, t)\delta_c u(\mathbf{x}, t)\right) + \nabla \cdot \left(c^2(\mathbf{x})\nabla\phi(\mathbf{x}, t)\right)\delta_c u(\mathbf{x}, t) \tag{1.27}$$

When integrating (1.27) over Ω, Gaussian theorem is applied for the first two terms

$$\int_\Omega \nabla \cdot \left(c^2(\mathbf{x})\nabla\delta_c u(\mathbf{x}, t)\right)\phi(\mathbf{x}, t)\mathrm{d}V = \int_{\partial\Omega} \mathbf{n} \cdot \left(c^2(\mathbf{x})\phi(\mathbf{x}, t)\nabla\delta_c u(\mathbf{x}, t)\right)\mathrm{d}S - \\ - \int_{\partial\Omega} \mathbf{n} \cdot \left(c^2(\mathbf{x})\nabla\phi(\mathbf{x}, t)\delta_c u(\mathbf{x}, t)\right)\mathrm{d}S + \int_\Omega \nabla \cdot \left(c^2(\mathbf{x})\nabla\phi(\mathbf{x}, t)\right)\delta_c u(\mathbf{x}, t)\mathrm{d}V. \tag{1.28}$$

The second term of (1.28) disappears due to perturbation of the (constant, zero) boundary condition (1.15) for wavefield u. The first term disappears, if we require the boundary condition for $\phi(\mathbf{x}, t)$ as

$$\phi(\mathbf{x}, t) = 0, \quad \text{for } \mathbf{x} \in \partial\Omega. \tag{1.29}$$

Then, (1.28) reduces to

$$\int_\Omega \nabla \cdot \left(c^2(\mathbf{x})\nabla\delta_c u(\mathbf{x}, t)\right)\phi(\mathbf{x}, t)\mathrm{d}V = \int_\Omega \nabla \cdot \left(c^2(\mathbf{x})\nabla\phi(\mathbf{x}, t)\right)\delta_c u(\mathbf{x}, t)\mathrm{d}V. \tag{1.30}$$

Note, that similarly to these operations, in (1.21) the term containing δc may be expressed as

$$\begin{aligned}
&\int_\Omega \nabla\cdot\left(2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\right)\phi(\mathbf{x},t)\mathrm{d}V = \\
&= \int_\Omega \left[\nabla\cdot\left(2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\phi(\mathbf{x},t)\right) - 2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\cdot\nabla\phi(\mathbf{x},t)\right]\mathrm{d}V = \\
&= \int_{\partial\Omega} \mathbf{n}\cdot\left(2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\phi(\mathbf{x},t)\right)\mathrm{d}S - \int_\Omega 2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\cdot\nabla\phi(\mathbf{x},t)\mathrm{d}V
\end{aligned} \tag{1.31}$$

so applying the boundary condition (1.29) for $\phi(\mathbf{x},t)$, we get

$$\int_\Omega \nabla\cdot\left(2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\right)\phi(\mathbf{x},t)\mathrm{d}V = -\int_\Omega 2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\cdot\nabla\phi(\mathbf{x},t)\mathrm{d}V. \tag{1.32}$$

Combining all terms (1.25), (1.30) and (1.32) in (1.21) and joining terms containing $\delta_c u$, we obtain

$$\begin{aligned}
\delta_c\chi = &\int_\Omega\int_t 2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\cdot\nabla\phi(\mathbf{x},t) + \\
&+\left[\frac{-\Delta T}{N}\dot{u}_{\mathrm{obs}}(\mathbf{x},t+\Delta T)\delta(\mathbf{x}-\mathbf{x}_{\mathrm{rec}}) + \ddot{\phi}(\mathbf{x},t) - \nabla\cdot\left(c^2(\mathbf{x})\nabla\phi(\mathbf{x},t)\right)\right]\delta_c u(\mathbf{x},t)\mathrm{d}V\,\mathrm{d}t.
\end{aligned} \tag{1.33}$$

So far, function $\phi(\mathbf{x},t)$ was arbitrary except for zero terminal and boundary conditions above. However, if we consider $\phi(\mathbf{x},t)$ fulfilling

$$\ddot{\phi}(\mathbf{x},t) - \nabla\cdot\left(c^2(\mathbf{x})\nabla\phi(\mathbf{x},t)\right) = \frac{\Delta T}{N}\dot{u}_{\mathrm{obs}}(\mathbf{x},t+\Delta T)\delta(\mathbf{x}-\mathbf{x}_{\mathrm{rec}}), \tag{1.34}$$

then the misfit derivative is reduced to

$$\delta_c\chi = \int_\Omega\int_t 2c(\mathbf{x})\delta c(\mathbf{x})\nabla u(\mathbf{x},t)\cdot\nabla\phi(\mathbf{x},t). \tag{1.35}$$

Eq. (1.34) for $\phi(\mathbf{x},t)$ is called adjoint equation, $\phi(\mathbf{x},t)$ is then called adjoint wavefield with source term

$$f^\dagger(\mathrm{x},t) = \frac{\Delta T}{N}\dot{u}_{\mathrm{obs}}(\mathbf{x},t+\Delta T)\delta(\mathbf{x}-\mathbf{x}_{\mathrm{rec}}), \tag{1.36}$$

so-called adjoint source. Note that for the case of membrane waves, the adjoint equation is identical to the original one (i.e., the problem is self-adjoint).

The sensitivity kernel for the cross-correlation traveltime is expressed from

$$\delta_c\chi = \Delta T\delta_c(\Delta T) = \Delta T\int_\Omega K(\mathbf{x})\delta c(\mathbf{x})\mathrm{d}V. \tag{1.37}$$

Comparing (1.35) and (1.37) we get

$$K(\mathbf{x}) = \int_t 2c(\mathbf{x})\nabla u(\mathbf{x},t)\cdot\nabla\phi(x,t)\mathrm{d}t. \tag{1.38}$$

with $\phi(x,t)$ calculated using $\Delta T = 1$ in (1.36).

As expected, the sensitivity kernel (1.38) depends on spatial coordinates $\mathbf{x}$ and is nontrivial combination of the forward and adjoint wavefields. This makes it less intuitive than simple ray sensitivity kernel (1.9).

1.3 General derivation of the adjoint method

For the sake of completness, here we show derivation of the adjoint method of misfit minimization for arbitrary linear differential operator following Lagrange multiplier approach as in Liu and Tromp (2006).

We start from the misfit minimization assuming residuals D_i obtained between observed and synthetic wavefields $\boldsymbol{u}_i(\mathbf{x},t)$

$$\chi = \frac{1}{2}\sum_i D\big(\boldsymbol{u}_i(\mathbf{x}_i,t)\big), \quad \text{or} \tag{1.39}$$

$$= \frac{1}{2}\sum_i \int_\Omega D\big(\boldsymbol{u}_i(\mathbf{x},t)\big)\delta(\mathbf{x}-\mathbf{x}_i)\mathrm{d}V. \tag{1.40}$$

The residual D may represent residual of either traveltime measurement in L2 norm or directly the wavefield differences or other complex properties derived from the wavefield (e.g., time-frequency misfits, Fichtner et al. (2008); Kristeková et al. (2009)).

The wavefield $\boldsymbol{u}(\mathbf{x},t)$ for each measurement satisfies general condition

$$\boldsymbol{L}_\mathbf{m}(\boldsymbol{u}(\mathbf{x},t)) = \boldsymbol{f}(\mathbf{x},t) \tag{1.41}$$

for all $\mathbf{x}$ and t. Here $\boldsymbol{L}_\mathbf{m}$ represents linear differential operator on the wavefield $\boldsymbol{u}(\mathbf{x},t)$, which depends on model parameters $\mathbf{m}(\mathbf{x})$, and $\boldsymbol{f}(\mathbf{x},t)$ is the source term. Suppose, that $\boldsymbol{L}_\mathbf{m}$ is also continuous, i.e. bounded.

To minimize/maximize a functional which is also subject to another contraint(s), the method of Lagrange multiplier is employed. Firstly, the Lagrangian is constructed

$$\mathcal{L} = \frac{1}{2}\sum_i \Bigg\{ \int_\Omega D\big(\boldsymbol{u}_i(\mathbf{x})\big)\delta(\mathbf{x}-\mathbf{x}_i)\mathrm{d}V + \\ + \int_\Omega\int_t \boldsymbol{\lambda}_i(\mathbf{x},t)\cdot\left[\boldsymbol{L}_\mathbf{m}(\boldsymbol{u}_i(\mathbf{x},t)) - \boldsymbol{f}_i(\mathbf{x},t)\right]\mathrm{d}t\mathrm{d}V\Bigg\}. \tag{1.42}$$

The (variational) derivative of the Lagrangian with respect to the model parameters $\mathbf{m}(\mathbf{x},t)$

$$\delta_\mathbf{m}\mathcal{L} = \sum_i \Bigg\{ \int_\Omega \delta_\mathbf{m} D(\boldsymbol{u}_i(\mathbf{x},t))\delta(\mathbf{x}-\mathbf{x}_i)\mathrm{d}V + \\ + \int_\Omega\int_t \boldsymbol{\lambda}_i\cdot\left[\boldsymbol{L}_\mathbf{m}(\delta_\mathbf{m}\boldsymbol{u}_i(\mathbf{x},t)) + \delta_\mathbf{m}\boldsymbol{L}_\mathbf{m}(\boldsymbol{u}_i(\mathbf{x},t))\right]\mathrm{d}t\mathrm{d}V\Bigg\}. \tag{1.43}$$

Using the definition of adjoint operator $\boldsymbol{L}^\dagger$ (considering inner product in form $\langle .,. \rangle = \int_\Omega\int_t ..\mathrm{d}t\mathrm{d}V$):

$$\langle \boldsymbol{L}_\mathbf{m}(\boldsymbol{u}_i(\mathbf{x},t)), \boldsymbol{\lambda}_i(\mathbf{x},t)\rangle = \langle \boldsymbol{u}_i(\mathbf{x},t), \boldsymbol{L}^\dagger_\mathbf{m}(\boldsymbol{\lambda}_i(\mathbf{x},t))\rangle, \tag{1.44}$$

and in addition requiring that variation of D may be expressed as

$$\delta_\mathbf{m} D(\boldsymbol{u}_i(\mathbf{x},t)) = \int_t \boldsymbol{d}\big(\boldsymbol{u}_i(\mathbf{x},t)\big)\cdot\delta_\mathbf{m}\boldsymbol{u}_i(\mathbf{x},t)\mathrm{d}t, \tag{1.45}$$

it is possible to rewrite (1.43) as

$$\delta_{\mathbf{m}}\mathcal{L} = \sum_i \left\{ \int_\Omega \int_t \left[\boldsymbol{d}(\boldsymbol{u}_i(\mathbf{x},t))\delta(\mathbf{x}-\mathbf{x}_i) + \boldsymbol{L}^\dagger_{\mathbf{m}}(\boldsymbol{\lambda}_i(\mathbf{x},t)) \right] \cdot \delta_{\mathbf{m}}\boldsymbol{u}_i(\mathbf{x},t) + \right.$$
$$\left. + \int_\Omega \int_t \boldsymbol{\lambda}_i(\mathbf{x},t) \cdot \delta_{\mathbf{m}}\boldsymbol{L}_{\mathbf{m}}(\boldsymbol{u}_i(\mathbf{x},t))\mathrm{d}t\mathrm{d}V \right\}. \quad (1.46)$$

If for each i, equation

$$\boldsymbol{L}^\dagger_{\mathbf{m}}(\boldsymbol{\lambda}_i(\mathbf{x},t)) = -\boldsymbol{d}(\boldsymbol{u}(\mathbf{x},t))\delta(\mathbf{x}-\mathbf{x}_i) \quad (1.47)$$

is fulfilled, then the variation of Lagrangian (1.46) reduces to

$$\delta_{\mathbf{m}}\mathcal{L} = \sum_i \int_\Omega \int_t \boldsymbol{\lambda}_i(\mathbf{x},t) \cdot \delta_{\mathbf{m}}\boldsymbol{L}_{\mathbf{m}}(\boldsymbol{u}_i(\mathbf{x},t))\mathrm{d}t\mathrm{d}V. \quad (1.48)$$

Since (1.47) employs the adjoint operator to $\boldsymbol{L}_{\mathbf{m}}$, function $\boldsymbol{\lambda}_i(\mathbf{x},t)$ is called the adjoint wavefield. The term on the right-hand side of (1.47) represents the adjoint source

$$\boldsymbol{f}^\dagger_i(\mathbf{x},t) = -\boldsymbol{d}(\boldsymbol{u}_i(\mathbf{x},t))\delta(\mathbf{x}-\mathbf{x}_i). \quad (1.49)$$

Note, that the adjoint source is a point source located in the receiver's position and its time function depends on the misfit definition.

Furthermore, let us assume that operator $\boldsymbol{L}_{\mathbf{m}}$ is linear in model parameters $\mathbf{m}$, then $\delta_{\mathbf{m}}\boldsymbol{L}_{\mathbf{m}} = \boldsymbol{L}_{\delta\mathbf{m}}$. We can define linear operator $\boldsymbol{A}$ as

$$\boldsymbol{L}_{\delta\mathbf{m}}(\boldsymbol{u}_i(\mathbf{x},t)) = \boldsymbol{A}_{\boldsymbol{u}_i}(\delta\mathbf{m}(\mathbf{x})), \quad (1.50)$$

find its adjoint $\boldsymbol{A}^\dagger$ and rewrite the derivative of Lagrangian (1.48) as

$$\delta_{\mathbf{m}}\mathcal{L} = \sum_i \int_\Omega \int_t \boldsymbol{A}^\dagger_{\boldsymbol{u}_i}(\boldsymbol{\lambda}_i(\mathbf{x},t))\delta\mathbf{m}(\mathbf{x})\mathrm{d}t\mathrm{d}V. \quad (1.51)$$

This gives the relation for the sensitivity kernel of the ith measurement

$$(K_{\mathbf{m}})_i(\mathbf{x}) = \int_t \boldsymbol{A}^\dagger_{\boldsymbol{u}_i}(\boldsymbol{\lambda}_i(\mathbf{x},t))\mathrm{d}t. \quad (1.52)$$

Operator $\boldsymbol{A}^\dagger$ is applied to the adjoint wavefield $\boldsymbol{\lambda}_i(\mathbf{x},t)$ and depends on the forward wavefield $\boldsymbol{u}_i(\mathbf{x},t)$.

1.3.1 Adjoint method for elastic wave propagation

We derive the adjoint operator for problem of elastic wave propagation in 3D heterogeneous media:

$$\rho(\mathbf{x})\ddot{\boldsymbol{u}}(\mathbf{x},t) - \nabla \cdot \big(\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{u}(\mathbf{x},t)\big) = \boldsymbol{f}(\mathbf{x},t), \quad (1.53)$$

where $\rho(\mathbf{x})$ is density and $\mathbf{C}(\mathbf{x})$ is the 4th order tensor of elastic parameters. We assume initial and boundary conditions

$$\begin{aligned} \boldsymbol{u}(\mathbf{x}, t_0) &= 0, && (1.54)\\ \dot{\boldsymbol{u}}(\mathbf{x}, t_0) &= 0, \quad \text{and} && (1.55)\\ \mathbf{n}(\mathbf{x}) \cdot \boldsymbol{\sigma}(\mathbf{x}, t) &= 0, \quad \text{for } \mathbf{x} \in \partial\Omega && (1.56) \end{aligned}$$

where $\boldsymbol{\sigma}(\mathbf{x}, t) = \mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{u}(\mathbf{x}, t)$ is the stress, and $\mathbf{n}(\mathbf{x})$ is normal to $\partial\Omega$.

The linear operator is thus defined as

$$\boldsymbol{L} = \rho(\mathbf{x}) \frac{\partial^2}{\partial t^2} + \nabla \cdot \mathbf{C}(\mathbf{x}) : \nabla. \tag{1.57}$$

In this case, the model parameters are represented by density and elastic parameters

$$\mathbf{m}(\mathbf{x}) = (\rho(\mathbf{x}), \mathbf{C}(\mathbf{x})). \tag{1.58}$$

We now search for the adjoint operator $\boldsymbol{L}^\dagger$

$$\langle \boldsymbol{L}(\boldsymbol{u}(\mathbf{x}, t)), \boldsymbol{\lambda}(\mathbf{x}, t) \rangle = \langle \boldsymbol{u}(\mathbf{x}, t), \boldsymbol{L}^\dagger(\boldsymbol{\lambda}(\mathbf{x}, t)) \rangle \tag{1.59}$$

where

$$\begin{aligned} &\langle \boldsymbol{L}(\boldsymbol{u}(\mathbf{x}, t)), \boldsymbol{\lambda}(\mathbf{x}, t) \rangle = \\ &\qquad = \int_\Omega \int_t \left[\rho(\mathbf{x}) \ddot{\boldsymbol{u}}(\mathbf{x}, t) - \nabla \cdot \big(\mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{u}(\mathbf{x}, t) \big) \right] \cdot \boldsymbol{\lambda}(\mathbf{x}, t) \mathrm{d}t \mathrm{d}V. \end{aligned} \tag{1.60}$$

The first term of (1.60) is solved using double per-partes integration

$$\begin{aligned} \int_\Omega \int_t \rho(\mathbf{x}) \ddot{\boldsymbol{u}}(\mathbf{x}, t) \cdot \boldsymbol{\lambda}(\mathbf{x}, t) \mathrm{d}t \mathrm{d}V = \int_\Omega \rho(\mathbf{x}) \Big\{ \big[\dot{\boldsymbol{u}}(\mathbf{x}, t) \cdot \boldsymbol{\lambda}(\mathbf{x}, t) \big]_{t_0}^{t_1} - \\ - \big[\boldsymbol{u}(\mathbf{x}, t) \cdot \dot{\boldsymbol{\lambda}}(\mathbf{x}, t) \big]_{t_0}^{t_1} + \int_t \boldsymbol{u}(\mathbf{x}, t) \cdot \ddot{\boldsymbol{\lambda}}(\mathbf{x}, t) \mathrm{d}t \Big\} \mathrm{d}V. \end{aligned} \tag{1.61}$$

Now the initial conditions (1.54)–(1.55) for $\boldsymbol{u}$ are applied. Furthermore, when we require the terminal conditions for $\boldsymbol{\lambda}$ in form:

$$\begin{aligned} \boldsymbol{\lambda}(\mathbf{x}, t_1) &= 0, && (1.62)\\ \dot{\boldsymbol{\lambda}}(\mathbf{x}, t_1) &= 0, && (1.63) \end{aligned}$$

(1.61) simplifies to

$$\int_\Omega \int_t \rho(\mathbf{x}) \ddot{\boldsymbol{u}}(\mathbf{x}, t) \cdot \boldsymbol{\lambda}(\mathbf{x}, t) \mathrm{d}t \mathrm{d}V = \int_\Omega \int_t \rho(\mathbf{x}) \boldsymbol{u}(\mathbf{x}, t) \cdot \ddot{\boldsymbol{\lambda}}(\mathbf{x}, t) \mathrm{d}t \mathrm{d}V. \tag{1.64}$$

Considering the second term in (1.60), we use identities based on divergence of product and symmetries of $\mathbf{C}(\mathbf{x})$, obtaining

$$\begin{aligned} \nabla \cdot \big(\mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{u}(\mathbf{x}, t) \big) \cdot \boldsymbol{\lambda}(\mathbf{x}, t) = &\nabla \cdot \big(\mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{u}(\mathbf{x}, t) \cdot \boldsymbol{\lambda}(\mathbf{x}, t) \big) - \\ &- \mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{u}(\mathbf{x}, t) : \nabla \boldsymbol{\lambda}(\mathbf{x}, t) \\ = \nabla \cdot \big(\mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{u}(\mathbf{x}, t) \cdot \boldsymbol{\lambda}(\mathbf{x}, t) \big) &- \nabla \cdot \big(\mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{\lambda}(\mathbf{x}, t) \cdot \boldsymbol{u}(\mathbf{x}, t) \big) + \\ &+ \nabla \cdot \big(\mathbf{C}(\mathbf{x}) : \nabla \boldsymbol{\lambda}(\mathbf{x}, t) \big) \cdot \boldsymbol{u}(\mathbf{x}, t). \end{aligned} \tag{1.65}$$

Substituting (1.65) in the integral of (1.60) and applying Gaussian theorem, we get

$$\int_t\int_\Omega \nabla\cdot\big(\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{u}(\mathbf{x},t)\big)\cdot\boldsymbol{\lambda}(\mathbf{x},t)\mathrm{d}V\mathrm{d}t = \int_t\Big\{\int_{\partial\Omega}\mathbf{n}(\mathbf{x})\cdot\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{u}(\mathbf{x},t)\cdot\boldsymbol{\lambda}(\mathbf{x},t)\mathrm{d}S - \\ -\int_{\partial\Omega}\mathbf{n}(\mathbf{x})\cdot\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{\lambda}(\mathbf{x},t)\cdot\boldsymbol{u}(\mathbf{x},t)\mathrm{d}S + \int_\Omega\nabla\cdot\big(\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{\lambda}(\mathbf{x},t)\big)\cdot\boldsymbol{u}(\mathbf{x},t)\mathrm{d}V\Big\}\mathrm{d}t. \tag{1.66}$$

Assuming the boundary conditions (1.56), the first term is zero. To cancel out the second term, the boundary conditions are set similarly for $\boldsymbol{\lambda}(\mathbf{x},t)$:

$$\mathbf{n}(\mathbf{x})\cdot\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{\lambda}(\mathbf{x},t) = 0 \quad \text{for } \mathbf{x}\in\partial\Omega. \tag{1.67}$$

Thus, we have derived the second part of the adjoint operator

$$\int_\Omega\int_t\nabla\cdot\big(\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{u}(\mathbf{x},t)\big)\cdot\boldsymbol{\lambda}(\mathbf{x},t)\mathrm{d}t\mathrm{d}V = \int_\Omega\int_t\nabla\cdot\big(\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{\lambda}(\mathbf{x},t)\big)\cdot\boldsymbol{u}(\mathbf{x},t)\mathrm{d}t\mathrm{d}V. \tag{1.68}$$

Summing up terms in equations (1.64) and (1.68)),

$$\int_\Omega\int_t\big[\rho(\mathbf{x})\ddot{\boldsymbol{u}}(\mathbf{x},t) - \nabla\cdot\big(\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{u}(\mathbf{x},t)\big)\big]\cdot\boldsymbol{\lambda}(\mathbf{x},t)\mathrm{d}t\mathrm{d}V = \\ = \int_\Omega\int_t\Big[\rho(\mathbf{x})\cdot\ddot{\boldsymbol{\lambda}}(\mathbf{x},t) - \nabla\cdot\big(\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{\lambda}(\mathbf{x},t)\big)\Big]\cdot\boldsymbol{u}(\mathbf{x},t)\mathrm{d}t\mathrm{d}V, \tag{1.69}$$

we have found the adjoint operator to the operator of elastic wave equation

$$\boldsymbol{L}^\dagger = \rho(\mathbf{x})\frac{\partial^2}{\partial t^2} + \nabla\cdot\mathbf{C}(\mathbf{x}) : \nabla. \tag{1.70}$$

with terminal and boundary conditions (1.62),(1.63) and (1.67). The adjoint operator is identical as the original operator $\boldsymbol{L}$, that is the operator of the elastic wave propagation is self-adjoint.

Furthermore, operator $\boldsymbol{L}$ is linear in model parameters $\rho(\mathbf{x})$ and $\mathbf{C}(\mathbf{x})$, leading to the variation of Lagrangian (for one measurement) in form

$$\delta_{\mathbf{m}}\mathcal{L} = \langle\boldsymbol{L}_{\delta\mathbf{m}}(\boldsymbol{u}(\mathbf{x},t)),\boldsymbol{\lambda}(\mathbf{x},t)\rangle \\ = \int_\Omega\int_t\big[\delta\rho(\mathbf{x})\ddot{\boldsymbol{u}}(\mathbf{x},t) - \nabla\cdot\big(\delta\mathbf{C}(\mathbf{x}) : \nabla\boldsymbol{u}(\mathbf{x},t)\big)\big]\cdot\boldsymbol{\lambda}(\mathbf{x},t)\mathrm{d}t\mathrm{d}V. \tag{1.71}$$

The kernel for density $\rho(\mathbf{x})$ is obtained directly as

$$\begin{aligned} K_\rho(\mathbf{x}) &= \int_t\ddot{\boldsymbol{u}}(\mathbf{x},t)\cdot\boldsymbol{\lambda}(\mathbf{x},t)\mathrm{d}t \\ &= -\int_t\dot{\boldsymbol{u}}(\mathbf{x},t)\cdot\dot{\boldsymbol{\lambda}}(\mathbf{x},t)\mathrm{d}t. \end{aligned} \tag{1.72}$$

To find the kernel for $\mathbf{C}(\mathbf{x})$, similar operations to finding the adjoint operator to $\boldsymbol{L}$ should be performed (applying rule for divergence of product and Gaussian theorem, assuming boundary conditions) and we obtain

$$\mathbf{K}_{\mathbf{C}}(\mathbf{x}) = \int_t\nabla\boldsymbol{u}(\mathbf{x},t)\nabla\boldsymbol{\lambda}(\mathbf{x},t)\mathrm{d}t. \tag{1.73}$$

1.3.2 Adjoint method for wave propagation in elastic isotropic media

In isotropic case, the tensor of elastic parameters $\mathbf{C}(\mathbf{x})$ depends only on two parameters, for example the Lamé parameters $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$

$$C_{klmn}(\mathbf{x}) = \lambda(\mathbf{x})\delta_{kl}\delta_{mn} + \mu(\mathbf{x})(\delta_{km}\delta_{ln} + \delta_{kn}\delta_{lm}). \tag{1.74}$$

When applied into the definition of the sensitivity kernel for $\mathbf{C}$ for one measurement (1.73), one easily obtains kernels for Lamé parameters:

$$\begin{aligned} K_\lambda(\mathbf{x}) &= K_{C_{klmn}}(\mathbf{x})\delta_{kl}\delta_{mn} && (1.75)\\ &= \int_t u_{k,l}(\mathbf{x},t)\lambda_{m,n}(\mathbf{x},t)\delta_{kl}\delta_{mn}\mathrm{d}t = \int_t u_{k,k}(\mathbf{x},t)\lambda_{m,m}(\mathbf{x},t)\mathrm{d}t && (1.76)\\ &= \int_t (\nabla\cdot\boldsymbol{u}(\mathbf{x},t))(\nabla\cdot\boldsymbol{\lambda}(\mathbf{x},t))\mathrm{d}t, && (1.77) \end{aligned}$$

and

$$\begin{aligned} K_\mu(\mathbf{x}) &= K_{C_{klmn}}(\mathbf{x})(\delta_{km}\delta_{ln} + \delta_{kn}\delta_{lm}) = 2K_{C_{klmn}}(\mathbf{x})\delta_{km}\delta_{ln} && (1.78)\\ &= \int_t 2u_{k,l}(\mathbf{x},t)\lambda_{m,n}(\mathbf{x},t)\delta_{km}\delta_{ln}\mathrm{d}t = \int_t 2u_{k,n}(\mathbf{x},t)\lambda_{k,n}(\mathbf{x},t)\mathrm{d}t && (1.79)\\ &= \int_t 2(\nabla\boldsymbol{u}(\mathbf{x},t)) : (\nabla\boldsymbol{\lambda}(\mathbf{x},t))\mathrm{d}t. && (1.80) \end{aligned}$$

where in the first step we have used symmetric property of $\mathbf{C}$.

1.3.3 Adjoint method for membrane wave

The membrane wave problem is a special case of isotropic elastic wave problem assuming 2D (in general inhomogeneous) media (representing a membrane) and scalar wavefield u which is perpendicular to the 2D membrane:

$$L = \rho(\mathbf{x})\ddot{u}(\mathbf{x},t) - \nabla\cdot\big(\mu(\mathbf{x})\nabla u(\mathbf{x},t)\big) \tag{1.81}$$

The sensitivity kernel is easily obtained for parameter μ using scalar wavefields in (1.80):

$$K_\mu(\mathbf{x}) = \int_t 2\nabla u(\mathbf{x},t)\cdot\nabla\lambda(\mathbf{x},t)\mathrm{d}t \tag{1.82}$$

Hereafter, this expression is employed.

1.4 Sensitivity kernels - examples and properties

Here, we show several examples of the sensitivity kernels calculated by 2D adjoint method (i.e., assuming membrane wave problem) and discuss their most important properties.

In membrane wave equation (1.81) we further assume constant density ρ. The forward source used here is a point source with Ricker wavelet time function:

$$f(\mathbf{x},t) = \delta(\mathbf{x}-\mathbf{x}_{\text{src}}) \left[\frac{1}{2} - (\pi f_p t)^2\right] e^{-(\pi f_p t)^2}, \tag{1.83}$$

with f_p representing the dominant frequency of the wavelet.

Moreover, in these simple examples we assume homogenous media $\mu(\mathbf{x}) = \mu$ and the perturbation in media generating the perturbed wavefield is also homogenous, i.e. $\delta\mu(\mathbf{x}) = \delta\mu$. The sensitivity kernels are calculated according to (1.82), with adjoint wavefield generated by adjoint source depending on misfit according to (1.49).

In the following, we show the sensitivity kernels assuming different misfit types or frequency content. For the calculation, we have employed the 2D adjoint version of software *SeisSol*. One forward calculation for the 20s Ricker wavelet takes approximately 1 CPU-hour.

1.4.1 Sensitivity kernels for various misfits

Reminding Section 1.3, the dependency of the sensitivity kernels on the type of the employed misfit (i.e. type of measurement D) is not straightforward but is resulting from the adjoint wavefield as generated by the adjoint source:

$$f^{\dagger}(\mathbf{x},t) = -d(u(\mathbf{x},t))\delta(\mathbf{x}-\mathbf{x}_{\text{rec}}), \tag{1.84}$$

where $\mathbf{x}_{\text{rec}}$ is location of the receiver and function $d(u(\mathbf{x},t))$ representing the adjoint source time function comes from expression

$$\delta_\mu D = \int_t d(u(\mathbf{x},t))\delta_\mu u(\mathbf{x},t)\mathrm{d}t. \tag{1.85}$$

Here, we show derivation of the adjoint sources and show examples of sensitivity kernels for some of traditionally employed misfit measurements. For sensitivity kernels of other misfit types, the reader is referred to Tromp et al. (2005); Fichtner et al. (2008); Chen et al. (2010); Bozdağ et al. (2011).

Full waveform misfit

The full waveform misfit for a measurement in $\mathbf{x}_{\text{rec}}$ is defined as

$$\chi = \frac{1}{2}\int_\Omega \int_t \left[u(\mathbf{x},t) - u_{\text{obs}}(\mathbf{x},t)\right]^2 \delta(\mathbf{x}-\mathbf{x}_{\text{rec}})\mathrm{d}t\mathrm{d}V. \tag{1.86}$$

The variation of the misfit with respect to model parameter $\mu(\mathbf{x})$ leads to

$$\delta_\mu\chi = \int_\Omega \int_t \left[u(\mathbf{x},t) - u_{\text{obs}}(\mathbf{x},t)\right] \delta(\mathbf{x}-\mathbf{x}_{\text{rec}})\delta_\mu u(\mathbf{x},t)\mathrm{d}t\mathrm{d}V, \tag{1.87}$$

leading straightforward to adjoint source in form

$$f^{\dagger}(\mathbf{x},t) = -\left[u(\mathbf{x},t) - u_{\text{obs}}(\mathbf{x},t)\right] \delta(\mathbf{x}-\mathbf{x}_{\text{rec}}). \tag{1.88}$$

Cross-correlation traveltime misfit

As already mentioned above (see Section 1.2), the cross-correlation traveltime misfit is defined

$$\chi = \frac{1}{2}\Delta T^2, \tag{1.89}$$

where the traveltime residual ΔT is obtained using cross-correlation function between observed wavefield $u_{\text{obs}}(\mathbf{x}, t)$ and calculated wavefield $u(\mathbf{x}, t)$

$$C(\mathbf{x}, t) = \int_t u(\mathbf{x}, \tau) u_{\text{obs}}(\mathbf{x}, t + \tau) \mathrm{d}\tau. \tag{1.90}$$

as

$$\Delta T = \arg\max_t \int_\Omega C(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_{\text{rec}}) \mathrm{d}V \tag{1.91}$$

Now, the derivative of ΔT with respect to μ should be expressed using $\delta_\mu u$:

$$\delta_\mu D = \Delta T \delta_\mu(\Delta T) = \int_t d(u(\mathbf{x}, t)) \delta_\mu u(\mathbf{x}, t) \mathrm{d}t. \tag{1.92}$$

Firstly, we differentiate the cross-correlation function $C(\mathbf{x}, t)$ and employ that for $t = \Delta T$, $C(\mathbf{x}, t)$ attains its maximum:

$$\left.\frac{\partial C(\mathbf{x}, t)}{\partial t}\right|_{t=\Delta T} = \int_t u(\mathbf{x}, \tau) \dot{u}_{\text{obs}}(\mathbf{x}, \Delta T + \tau) \mathrm{d}\tau = 0. \tag{1.93}$$

This equation is now differentiated with respect to the model parameter μ. Note that now not only the calculated wavefield $u(\mathbf{x}, t)$ but also the position of cross-correlation maxima ΔT depends on the model parameter:

$$\int_t \left[\delta_\mu u(\mathbf{x}, \tau) \dot{u}_{\text{obs}}(\mathbf{x}, \tau + \Delta T) + u(\mathbf{x}, \tau) \ddot{u}_{\text{obs}}(\mathbf{x}, \tau + \Delta T) \delta_\mu(\Delta T)\right] \mathrm{d}t = 0. \tag{1.94}$$

From (1.94), the variation in the traveltime residual may be expressed

$$\delta_\mu(\Delta T) = -\frac{\int_t \dot{u}_{\text{obs}}(\mathbf{x}, \tau + \Delta T) \delta_\mu u(\mathbf{x}, \tau) \mathrm{d}\tau}{\int_t u(\mathbf{x}, \tau) \ddot{u}_{\text{obs}}(\mathbf{x}, \tau + \Delta T) \mathrm{d}\tau}. \tag{1.95}$$

Comparing equations (1.92) and (1.95), we find

$$d(u(\mathbf{x}, t)) = -\Delta T \frac{\dot{u}_{\text{obs}}(\mathbf{x}, t + \Delta T_i)}{\int_t u(\mathbf{x}, \tau) \ddot{u}_{\text{obs}}(\mathbf{x}, \tau + \Delta T_i) \mathrm{d}\tau} \tag{1.96}$$

leading to adjoint source in form

$$f^\dagger(\mathbf{x}, t) = \Delta T \frac{\dot{u}_{\text{obs}}(\mathbf{x}, t + \Delta T)}{\int_t u(\mathbf{x}, \tau) \ddot{u}_{\text{obs}}(\mathbf{x}, \tau + \Delta T) \mathrm{d}\tau} \delta(\mathbf{x} - \mathbf{x}_{\text{rec}}). \tag{1.97}$$

Note that (1.97) deviates from the more traditional form of normalized calculated velocity weighed by the traveltime residual, as used and derived for instance by Dahlen et al. (2000); Marquering et al. (1999):

$$f^\dagger(\mathbf{x}, t) = -\Delta T \frac{1}{\|\dot{u}(\mathbf{x})\|^2} \dot{u}(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_{\text{rec}}). \tag{1.98}$$

To obtain (1.98), one assumes in (1.97) that the observed and calculated waveforms are similar, differing only in time shift ΔT, i.e. $\dot{u}_{\text{obs}}(\mathbf{x}, t + \Delta T) \approx \dot{u}(\mathbf{x}, t)$. Moreover, the integration per partes in the denominator is performed (assuming zero waveforms at the beginning and end of time interval): $\int_t u(\mathbf{x}, \tau) \ddot{u}_{\text{obs}}(\mathbf{x}, \tau + \Delta T) \mathrm{d}\tau \approx -\int_t \left[\dot{u}(\mathbf{x}, \tau)\right]^2 \mathrm{d}\tau = -\|\dot{u}(\mathbf{x})\|^2$

Envelope cross-correlation traveltime misfit

Although this type of misfit is not generally used, it may be employed for the inversion of group wave traveltimes when membrane wave approximation of surface wave propagation is assumed.

The misfit would be defined as

$$\chi = \frac{1}{2}\Delta T_{\text{env}}^2, \tag{1.99}$$

where ΔT_{env} is calculated from the cross-correlation of the signal envelopes $e(\mathbf{x}, t)$:

$$\Delta T_{\text{env}} = \arg\max_t \int_t \int_\Omega e(\mathbf{x}, \tau) e_{\text{obs}}(\mathbf{x}, t+\tau) \delta(\mathbf{x} - \mathbf{x}_{\text{rec}}) \mathrm{d}V \mathrm{d}\tau, \tag{1.100}$$

with the signal envelopes $e(\mathbf{x}, t)$ defined as

$$e(\mathbf{x}, t) = \sqrt{u(\mathbf{x}, t)^2 + \mathcal{H}(u(\mathbf{x}, t))^2}, \tag{1.101}$$

where $\mathcal{H}$ denotes the Hilbert transform.

As shown in (1.95), the variation using the cross-correlation function leads to

$$\delta_\mu(\Delta T_{\text{env}}) = -\frac{\int_t \dot{e}_{\text{obs}}(\mathbf{x}, \tau + \Delta T_{\text{env}}) \delta_\mu e(\mathbf{x}, \tau) \mathrm{d}\tau}{\int_t e(\mathbf{x}, \tau) \ddot{e}_{\text{obs}}(\mathbf{x}, \tau + \Delta T_{\text{env}}) \mathrm{d}\tau}. \tag{1.102}$$

We further need to express the variation of the envelope $\delta_\mu e(\mathbf{x}, t)$ using the variation of the waveform $\delta_\mu u(\mathbf{x}, t)$:

$$\begin{aligned} \delta_\mu e(\mathbf{x}, t) &= \frac{\partial e}{\partial u} \delta_\mu u(\mathbf{x}, t) \\ &= \frac{1}{e(\mathbf{x}, t)} \left[u(\mathbf{x}, t) \delta_\mu u(\mathbf{x}, t) + \mathcal{H}(u(\mathbf{x}, t)) \mathcal{H}(\delta_\mu u(\mathbf{x}, t)) \right]. \end{aligned} \tag{1.103}$$

Now we need to extract $\delta_\mu u(\mathbf{x}, t)$ from the Hilber transform:

$$\begin{aligned} &\int \frac{\dot{e}_{\text{obs}}(\mathbf{x}, \tau + \Delta T_{\text{env}})}{e(\mathbf{x}, \tau)} \mathcal{H}(u(\mathbf{x}, \tau)) \mathcal{H}(\delta_\mu u(\mathbf{x}, \tau)) d\tau = \\ &\quad = \int \frac{\dot{e}_{\text{obs}}(\mathbf{x}, \tau + \Delta T_{\text{env}})}{e(\mathbf{x}, \tau)} \mathcal{H}(u(\mathbf{x}, \tau)) \left(\int \frac{\delta_\mu u(\mathbf{x}, t)}{\pi(\tau - t)} dt \right) d\tau \\ &\quad = \iint \frac{\dot{e}_{\text{obs}}(\mathbf{x}, \tau + \Delta T_{\text{env}})}{e(\mathbf{x}, \tau)} \mathcal{H}(u(\mathbf{x}, t)) \frac{1}{\pi(\tau - t)} d\tau \delta_\mu u(\mathbf{x}, t) dt \\ &\quad = \int -\mathcal{H} \left(\frac{\dot{e}_{\text{obs}}(\mathbf{x}, t + \Delta T_{\text{env}})}{e(\mathbf{x}, t)} \mathcal{H}(u(\mathbf{x}, t)) \right) \delta_\mu u(\mathbf{x}, \tau) d\tau, \end{aligned} \tag{1.104}$$

where we have interchanged the two time integrals (valid operation assuming that the result exists).

To sum up all terms, the adjoint source is in form

$$\begin{aligned} f^\dagger(\mathbf{x}, t) = \frac{\Delta T_{\text{env}} \delta(\mathbf{x} - \mathbf{x}_{\text{rec}})}{\int \dot{e}(\mathbf{x}, \Delta T_{\text{env}} + \tau) \dot{e}_{\text{obs}}(\mathbf{x}, \tau) d\tau} \Bigg[&\frac{\dot{e}_{\text{obs}}(\mathbf{x}, t + \Delta T_{env}) u(\mathbf{x}, t)}{e(\mathbf{x}, t)} - \\ &- \mathcal{H} \left(\frac{\dot{e}_{\text{obs}}(\mathbf{x}, t + \Delta T_{\text{env}}) \mathcal{H}(u(\mathbf{x}, t))}{e(\mathbf{x}, t)} \right) \Bigg]. \end{aligned} \tag{1.105}$$

Moreover, assuming (similarly as for the cross-correlation of waveforms) that the envelopes $e(\mathbf{x},t)$ and $e_{\text{obs}}(\mathbf{x},t)$ are similar and differ only by the time shift ΔT_{env}, the adjoint sources are simplified to

$$f^{\dagger}(\mathbf{x},t) = -\frac{\Delta T_{\text{env}}\delta(\mathbf{x}-\mathbf{x}_{\text{rec}})}{\|\dot{e}(\mathbf{x})\|^2}\left[\frac{\dot{e}(\mathbf{x},t)u(\mathbf{x},t)}{e(\mathbf{x},t)} - \mathcal{H}\left(\frac{\dot{e}(\mathbf{x},t)\mathcal{H}(u(\mathbf{x},t))}{e(\mathbf{x},t)}\right)\right]. \quad (1.106)$$

One notices that in the denominator the envelope might be $e(\mathbf{x},t)=0$. This happens when both $u(\mathbf{x},t)$ and its Hilbert tranform $\mathcal{H}(u(\mathbf{x},t))$ are zero. When inspecting the numerator, the envelope is multiplied by the waveform or its Hilbert tranform. The estimations lead to the whole term $\frac{\dot{e}u}{e}$ in (1.105) or (1.106) to be $O(u(\mathbf{x},t))$ for all t and is thus bounded.

Fig. 1.1 shows the example of the normalized adjoint sources for the three different types of misfits assumed above. Although the definitions for all three types are quite diverse, the time functions of the adjoint sources in Fig. 1.1 are similar – especially for the waveform and their crosscorrelation traveltime misfit. Omitting the amplitudes, they actually differ by a small time shift. This is presumably the result of the simple setting in the example (homogenous media), where the difference between the perturbed and unperturbed waveform leads only to the time shift. The most distinct of them is the envelope cross-correlation traveltime misfit with more complex behavior resulting from the combination of the two terms in (1.105) or (1.106).

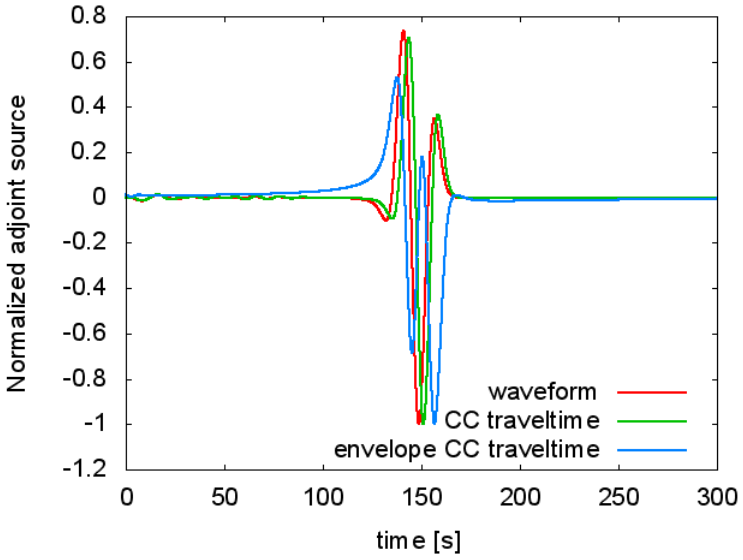

Figure 1.1: Normalized adjoint sources for different types of misfits (CC stands for cross-correlation).

The sensitivity kernels corresponding to the above-mentioned misfit types are shown in Fig. 1.2. The kernels have quite complex form and resemble the Fresnel zones in optics. Moreover, they show very high amplitudes in the source and the receiver. All kernels are similarly wide reflecting the same frequency content. The waveform kernel shows positive and negative alternating (Fresnel) zones while the

traveltime kernels show one main (positive) Fresnel zone surrounded by weaker and thinner (negative) higher Fresnel zones. In the central part of the kernels along the ray, the waveform kernel shows reverse amplitude compared to the traveltime kernels. The traveltime kernels show weaker amplitude in their central part. This phenomenon is known in 3D as banana-doughnut paradox, where the cross-correlation traveltime sensitivity kernel is zero along the ray path (Dahlen et al., 2000; Hung et al., 2000). Note, that both (envelope and signal cross-correlation) traveltime kernels are very similar. The envelope traveltime kernel has only higher amplitudes (probably it is a result of positive superposition of the wavefields generated by the adjoint sources containing the two terms). One may, therefore, in case of the simple waveform employ in the inversion the simpler form of the traveltime kernel (i.e., signal cross-correlation kernel).

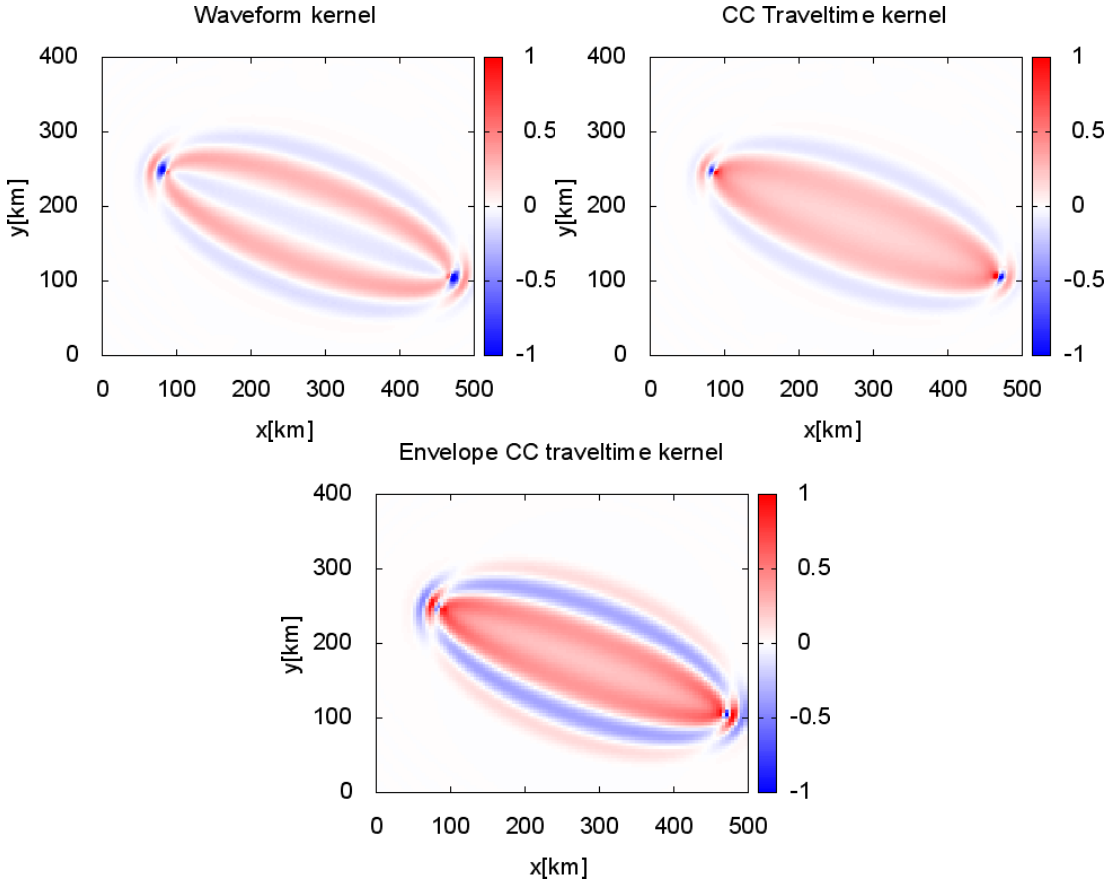

Figure 1.2: Sensitivity kernels for 3 different types of misfits (CC stands for cross-correlation).

In conclusion, the waveform misfit takes into account both phase and amplitudes of signals. However, if one is not interested in amplitudes or the data consist of traveltime measurements only, the cross-correlation traveltime misfit can be employed. We prefer the utilization of the cross-correlation traveltime misfit over the first-arrival traveltime misfit as it better accounts for the frequency content of the data. However, if the observed waveform is very complex, the traveltime misfit is inapplicable and the waveform misfit may give questionable results. In this case, use of a more advanced misfit (e.g., time-frequency misfit) is desirable (Bozdağ et al., 2011; Fichtner et al., 2008).

1.4.2 Frequency-dependence of the sensitivity kernels

To examine the dependency of the sensitivity kernels on the frequency content of the data, we consider different dominant frequency of the Ricker wavelet in particular 20 s, 10 s and 5 s. The wavelet and their spectra are shown in Fig. 1.3.

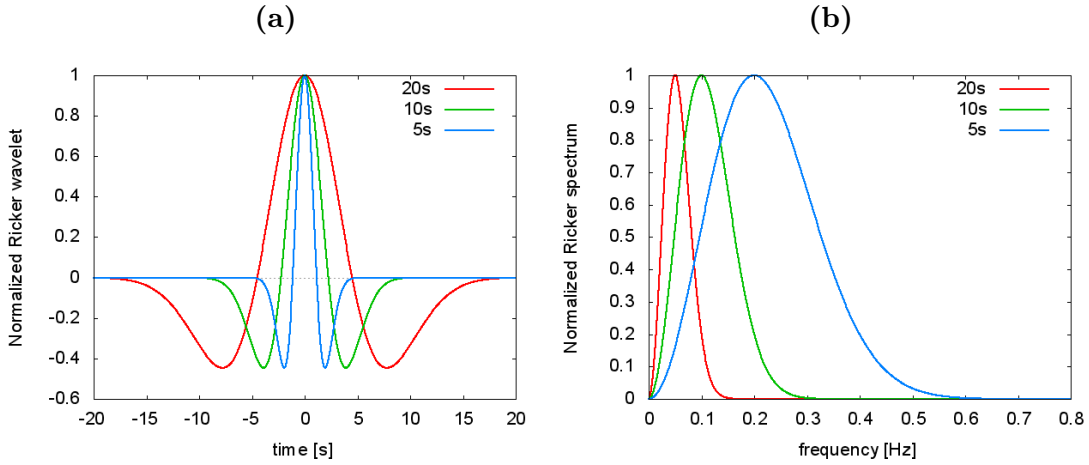

Figure 1.3: Ricker wavelet function for different dominant frequencies (a) and their normalized spectra (b).

The cross-correlation traveltime sensitivity kernels for each frequency are shown in Fig. 1.4. From Fig. 1.4, we can deduce that when employing the higher frequency data (i.e., low periods), the sensitivity kernels are narrower. This agrees with the intuitive concept of the sensitivity kernel which should for infinitely high frequency reduce to infinitely thin line – the ray.

1.4.3 Sensitivity kernel – Gaussian smoothing

The sensitivity kernel contains singularity inherited from the wavefields $u(\mathbf{x}, t)$ and $u^{\dagger}(\mathbf{x}, t)$, located in the point source and adjoint point source, as already seen in Fig. 1.2 and 1.4.

In the adjoint tomographic inversion, the individual kernels are summed into the misfit gradient kernel, which is used, for example, in the steepest descent method to update the model parameters. If the kernels are unmodified, the model would change mainly in the vicinity of the singularities, that is where the sources and the receivers are located. To avoid this issue, the kernels are usually preconditioned by various means. One of the simpler and thus commonly employed kernel regularization is smoothing the kernels, for example by convolution with a Gaussian function. This approach was also implemented in the *Seissol2D*, therefore it is demonstrated here.

Fig. 1.5 shows cross-correlation traveltime sensitivity kernels smoothed by 2D Gaussian of different widths, in particular 30, 60 and 100 km. The sensitivity kernels in the example were calculated for 20 s Ricker wavelet assuming homogenous media of 3km/s, which gives dominant wavelength of 60 km. Thus the assumed size of the Gaussian represents either undersmoothing, ideal smoothing or oversmoothing of the problem (see also Chapter 2) .

Already the smoothing using 30 km wide Gaussian significantly eliminates the kernel's singularities while preserving its main features – the second Fresnel zone

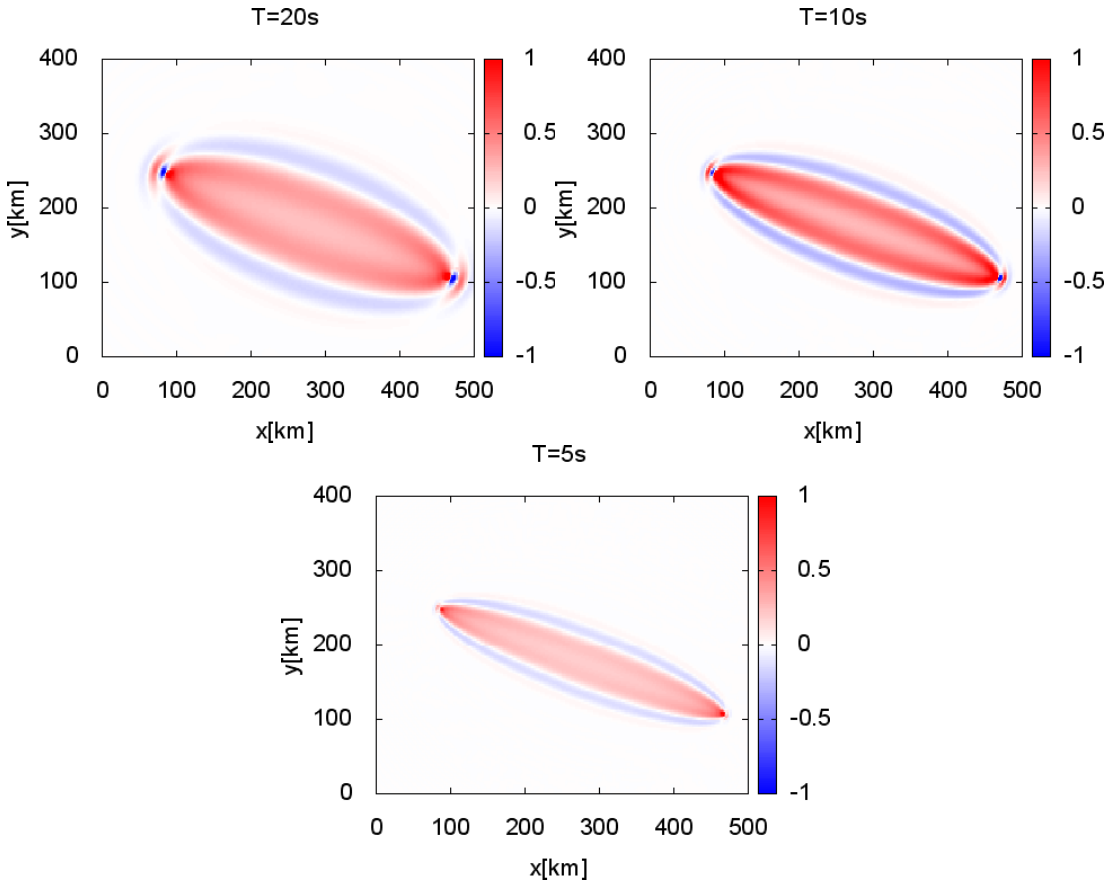

Figure 1.4: Traveltime (cross-correlation) sensitivity kernels with different frequency of Ricker wavelet source.

and weaker sensitivity in the central part. With the increasing the smoothing width, the first Fresnel zone is more homogeneous and the second Fresnel zone diminishes. In the case of oversmoothing by the 100 km wide Gaussian, the traveltime sensitivity kernel is completely blurred in the first Fresnel zone.

In Fig. 1.6 we show the waveform sensitivity kernels for Gaussian smoothing 30 km and 100 km. As in the previous case, the smoothing of 30 km wide Gaussian maintains the basic properties of the kernel and removes the singularities. On the other hand, the oversmoothing leads to oversimplified kernel containing only one Fresnel zone, which extends over all positive and negative Fresnel zones of the original unsmoothed kernel (see also Fig. 1.2). In this case, the blurred kernel shows decrease in amplitudes in the central part which is a relic of the negative sensitivity in the original kernel.

The natural choice of the smoothing size might be considered lower than the dominant wavelength, that is the one which reduces the singularities but preserves the main features of the original kernel. In the presented examples this corresponds to the smoothing by 30 km wide Gaussian. However, real data application (which is considered later in Chapters 2 and 4) requires significantly larger values of the smoothing width to reduce the effect of the data noise. In particular, for the 20 s Love wave group traveltimes with the estimated noise level the optimal smoothing width of the Gaussian was 100 km.

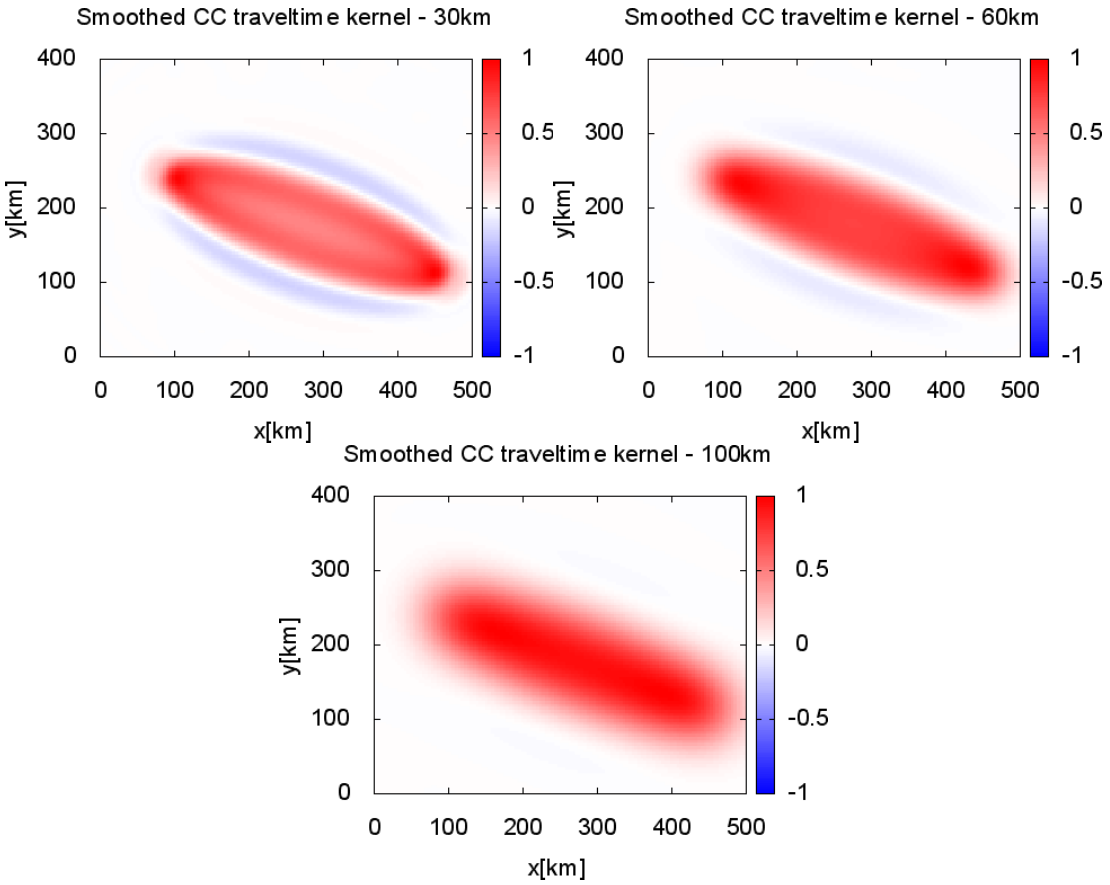

Figure 1.5: Cross-correlation (CC) traveltime sensitivity kernels convolved with Gaussian bell of different widths: 30 km, 60 km and 100 km.

1.5 Gradient based misfit minimization and implementation

Gradient method is a common approach in misfit minimization for nonlinear problems. The method is iterative, the new model is found by modifying the current model in the so-called descent direction $\boldsymbol{\gamma}_n(\mathbf{x})$:

$$\mathbf{m}_{n+1}(\mathbf{x}) = \mathbf{m}_n(\mathbf{x}) + C_n \boldsymbol{\gamma}_n(\mathbf{x}) \tag{1.107}$$

where C_n is a positive value, usually denoted as the stepsize. The process is repeated until the misfit minimum is satisfactorily achieved.

To calculate the descent direction, the misfit gradient (steepest ascent direction) is employed. If the descent direction is calculated as negative of the gradient, the method is known as the steepest descent. However, the steepest descent method may by converging to the minimum of the misfit slowly as the steepest descent direction may apply only locally. Therefore, more advanced methods to calculate the descent directions are applied: Newton method which employs not only the misfit gradient but also the second derivative – Hessian, or conjugate gradient method which calculates the descent direction using all previous directions.

According to (1.51) and (1.52), we have expressed the gradient of the Lagrangian using the sensitivity kernels $(K_{\mathbf{m}})_i(\mathbf{x})$ calculated by the adjoint method.

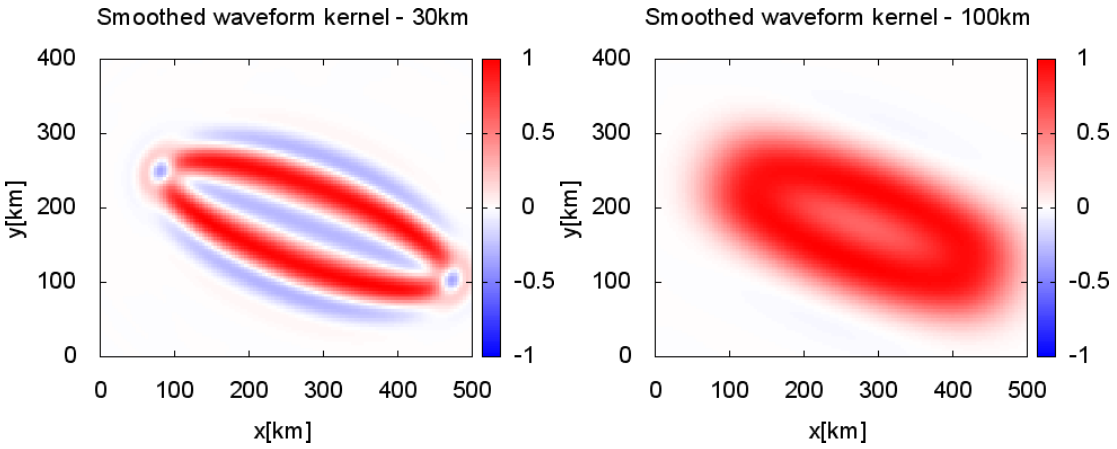

Figure 1.6: Waveform sensitivity kernels convolved with Gaussian bell of different widths: 30km and 100km.

For the steepest descent method, (1.107) then reads

$$\mathbf{m}_{n+1}(\mathbf{x}) = \mathbf{m}_n(\mathbf{x}) - C_n \sum_i (K_{\mathbf{m}_n})_i(\mathbf{x}). \tag{1.108}$$

However, in practice conjugate gradient approach is preferred as faster converging method,

Minimization employing the adjoint calculation was implemented into the 2D version of *SeisSol* (http://www.seissol.org/). We have further developed this implementation (originally prepared by Josep de la Puente) to be suitable for our application.

The forward calculation is based on arbitrary high-order accurate derivative discontinuous Galerkin method (ADER-DG, Käser and Dumbser (2006); Dumbser and Käser (2006); Käser et al. (2007); Dumbser et al. (2007); de la Puente et al. (2007)). We use unstructured triangular mesh, where the size of the elements depends on the dominant frequency (wavelength) of the propagating waves – the lower frequency (i.e., the higher period) the larger is the size of the element. Similarly, the temporal sampling is adjusted. In particular, for the 20s membrane wave, the size of the mesh elements is 10 km and the time sampling is 0.1 s.

The kernels are calculated on regular 2D grid; its spacing depends likewise on the dominant period of the calculated waves. To calculate the sensitivity kernels the procedure is the following: the forward 2D wavefield is calculated and stored in the computer memory for the 2D kernel grid in selected time intervals. To save the memory space, the time sampling step of the stored wavefield is larger than the time step of the forward calculation, i.e., several tens to hundred timesteps are not stored. After the forward calculation, the adjoint sources are determined using the residuals between the synthetics and the data and the adjoint calculation is performed. During this calculation, adjoint wavefield is combined with the stored forward wavefield for the corresponding time steps on the 2D regular grid and integrated into the sensitivity kernel.

The calculated kernels are then summed into the misfit gradient kernel. The sensitivity kernels derived in Section 1.3 correspond to each data residual, which leads to high computational cost: (number of measurements)$\times$(2 simulations). To reduce the computational demand, Tromp et al. (2005) proposed to use the

so-called event sensitivity kernels, where all kernels from a given forward source are calculated at once: the forward wavefield is saved on the kernel grid, the adjoint sources are computed for all receivers and then the adjoint wavefield is backpropagated from all the adjoint sources together. During the latter calculation, the event sensitivity kernel is being evaluated. The computational cost is (number of forward sources)$\times$(2 simulations), which presents significant reduction in case of large number of receivers – a common situation in seismic tomography.

The sensitivity kernels or their sum representing the gradient kernel are eventually smoothed out by convolution with 2D Gaussian function of chosen width.

The gradient method strongly depends on the choice of the starting model. Since we have no detailed apriori information on the starting model, we use homogeneous starting model with velocity obtained from average dispersion curve of the studied area. We believe that this general model is less prone to be stuck in local minimum of the misfit.

The new model is calculated iteratively from the previous one using the conjugate gradient method, the scheme employed in our application was either Polak-Ribiére or Hestenes-Stiefel (see e.g., Tarantola (2005)). We have tested both methods, they give similar results and have good properties when the old and new gradients are significantly different. After determining the descent direction, the stepsize C_n in descent direction $\boldsymbol{\gamma}_n(\mathbf{x})$ is decided: several test steps are performed in the descent direction, the misfit in the testing models is calculated and quadratic formula is fitted to the stepsize-misfit dependence. The minimum obtained from the fitted parabola is assigned as the stepsize C_n. For this, one would need to calculate the misfit for all measurements in several testing models, which presents additional computational burden. Therefore, the misfit for the test steps is calculated only for a subset of forward sources which are manually chosen to achieve good coverage over the studied area.

1.6 Synthetic checkerboard tests

The synthetic tests were carried out to examine the performance of the adjoint inversion by $SeisSol2D$. Here we present the results of the checkerboard tests for 2D membrane wave problem. The checkerboard tests, apart from their (mis)use in real data application, may provide insight into the general properties of the inversion.

The source–station configuration is taken from the real data application (see Chapter 2). Therein, the 2D membrane wave problem is used to approximate the surface wave propagation. The real data traveltimes are measured using cross-correlations between the station pairs. Thus, some of the stations act also as sources in the calculation. The source–station configuration is shown in Fig. 1.7.

The synthetic sources are characterized by Ricker wavelet time function with dominant frequency 20 s. The average velocity in media is assumed 3 km/s, i.e. the wavelength of the waves is $\sim$ 60 km. The synthetic waveform data are kept noise-less. The checkerboard tests show mainly the data coverage of the domain. Moreover, we have performed several checkerboard tests with different size of the heterogeneities to estimate also the resolving power of the long-wavelength dataset.

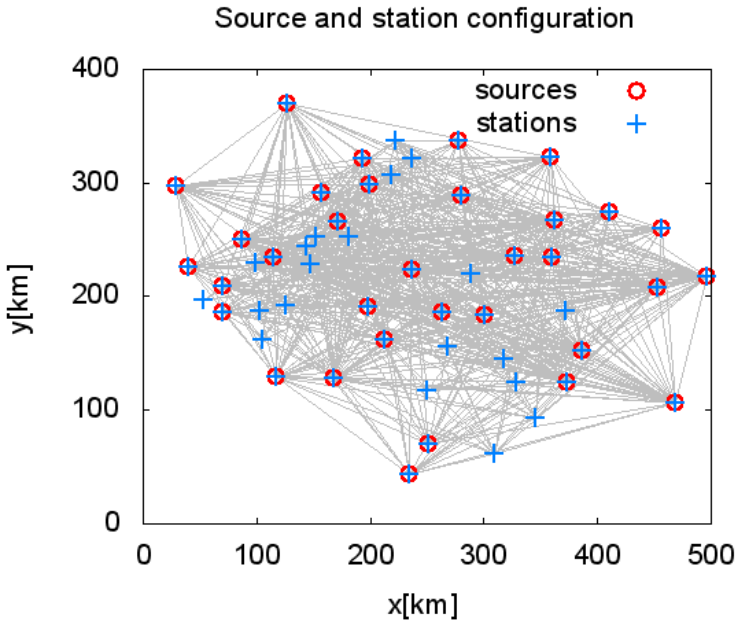

Figure 1.7: Configuration of stations for the checkerboard tests. The stations acting in the calculation as sources are marked by red circles. The station-pairs with traveltime measurement are connected by lines.

The inversion employs the cross-correlation traveltime sensitivity kernels for the membrane wave approximation (see Section 1.2). The individual kernels are combined into the gradient kernel, which is used in the conjugate gradient method (in particular Polak-Ribiére scheme). To diminish the kernels' singularities, the kernels are convolved with the Gaussian bell (5, 10 or 30 km wide). The iterations are stopped when the model does not change significantly; the number of the iteration steps varies between 5 and 8.

1.6.1 Checkerboard test with long-wavelength structure

In the first test (CB test I), the reference/target model is composed of relatively large structures – the heterogeneities are 100 km apart and smoothed, their amplitude reaches 5% (see Fig. 1.8 left). Thus the wavelength of the anomalies is larger than the wavelength of the waves used for the inversion. The calculated misfit gradient was smoothed using 10 km wide Gaussian bell to reduce the singularities without artificially increasing the kernel size.

The inverted model (Fig. 1.8 right) generally recovers the original structures in areas where the data coverage is good. Minor discrepancies are resulting from the uneven station coverage and concentrate in the areas where the data coverage is rather sparse (see Fig. 1.7). In addition, the recovered amplitudes are generally smaller. The wavelength of the structures is larger than the wavelength of the waves used, thus the model appears to be well resolved.

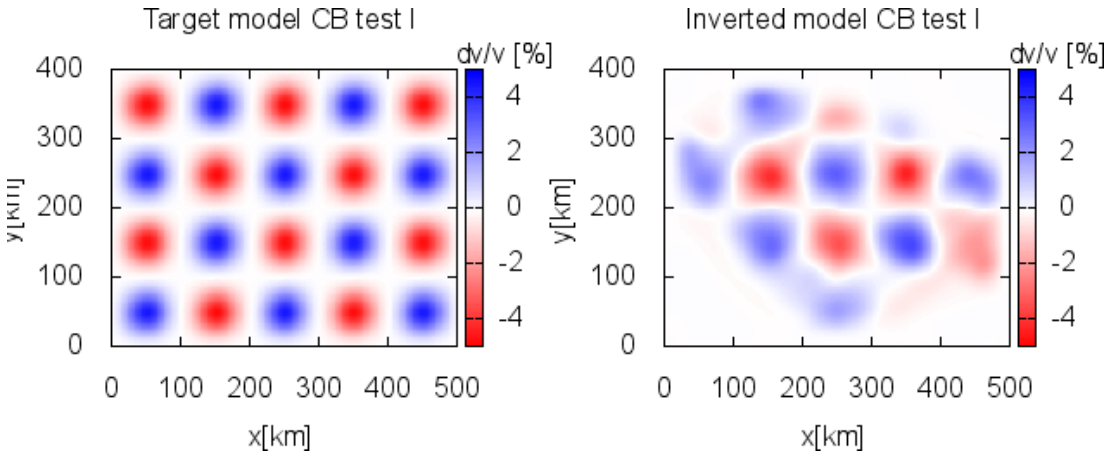

Figure 1.8: Checkerboard test I (long-wavelength structures): left – target model, right – inverted model.

1.6.2 Checkerboard test with short-wavelength structures

The second test (CB test II) is carried out for the checkerboard model with small-scale structures but keeping the large distance between them (see Fig. 1.9a). The wavelength of the waves employed for the inversion is much larger than the size of the anomalies.

Fig. 1.9b shows the result of the inversion assuming two different widths of the smoothing Gaussian (5 and 30 km) to demonstrate its effect on the resulting model. Both results in Fig. 1.9b show similar large-scale structures. In addition, they resemble the result for the long-wavelength structure (see Fig. 1.8), representing a bias caused by the finite-frequency effect of the employed waveforms. This confirms that the imaged structures are primarily affected by a) the source-station configuration, b) the wavelength of the employed waves.

The differences between the two results in Fig. 1.9 arise on smaller-scales – the model obtained with 5 km wide Gaussian smoothing appears to have much more (though artificial) details. Furthermore, the maximal amplitudes of the heterogeneities are higher. The decrease in the amplitudes due to the regularization by means of model smoothing is a well-known phenomenon.

In conclusion, we should bear in mind that the smaller-than-wavelength structures are difficult to recover correctly and are inherently smoothed out by the long-wavelength waves into larger scale structures with decreased amplitudes.

1.6.3 Checkerboard test with mixed structures

The last presented checkerboard test (CB test III) is composed of a combination of the previous reference models. The reference model thus contains both large-scale and short-scale structures (see Fig. 1.10a). The results of the inversion assuming again two different degrees of smoothing are shown in Fig. 1.10b.

The inverted models agree with the reference model and with each other in the large-scale structures. The small-scale structures were smoothed out by the long-wavelength waves and are, therefore, not perceptible in the resulting models. Another problem in this case are the smallest-scale structures (peaks) present in the model for the smoothing of 5 km. They correlate with the position of the

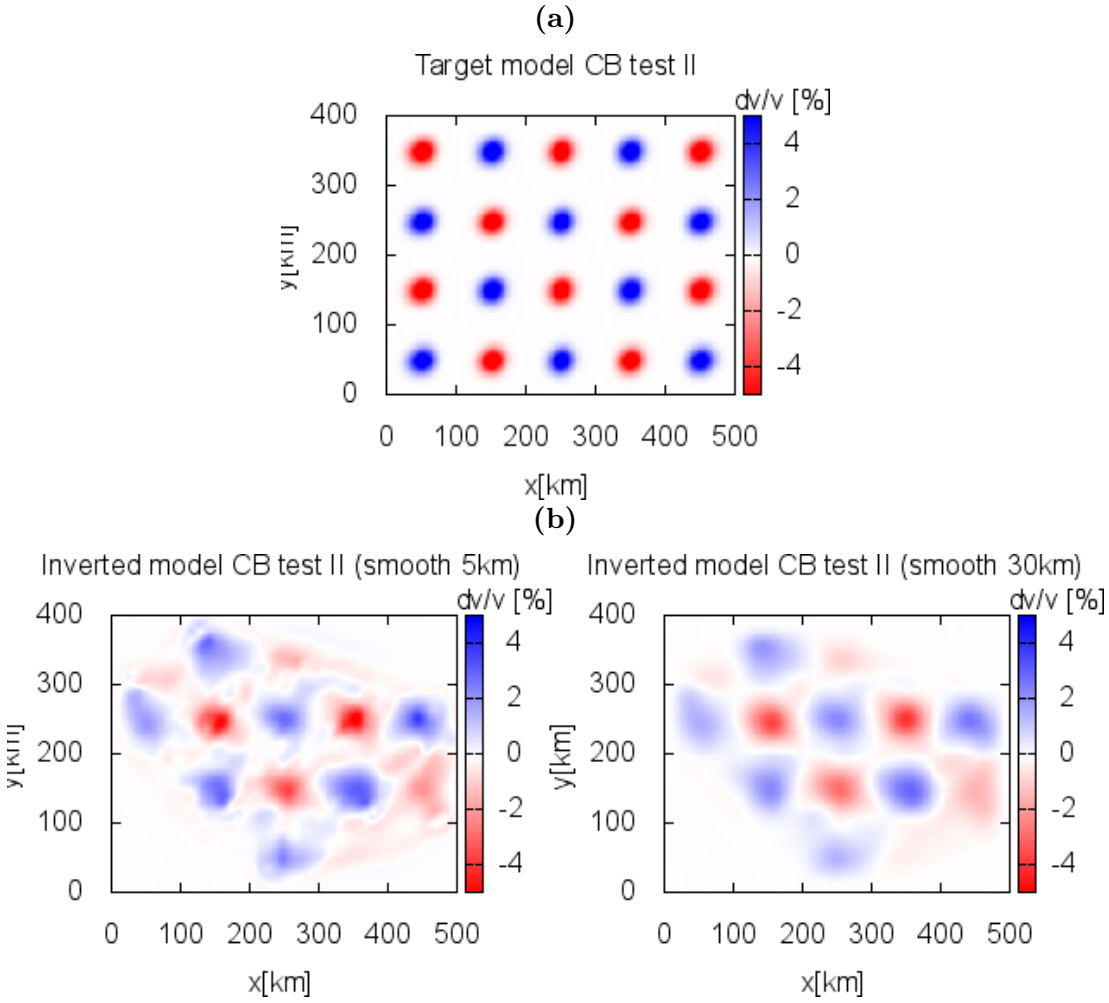

Figure 1.9: Checkerboard test II (shorter-wavelength structures): a) target model, b) inverted models using 5 km (left) and 30 km (right) Gaussian smoothing.

stations (see Fig. 1.7) where the sensitivity kernel singularities are located. The resulting model containing these distinct artifacts is considered underregularized and, for the next application utilizing the long-period data, the regularization by larger Gaussian smoothing function should be employed.

1.7 Conclusion

In this chapter, we have summarized the state-of-the-art adjoint method that is with increasing favor applied in the seismic tomography. Although intuitively it may be found as the generalization of the classical ray method, the theory behind it is more complicated. The adjoint inversion method is based on iterative improvement of the model using the gradient approach. The gradient employing the volumetric sensitivity kernels is calculated using a forward wavefield and an adjoint wavefield corresponding to the adjoint operator of the original problem.

The main advantage of the adjoint method is that the elastodynamic equation may be employed without further approximations. Therefore, the only restric-

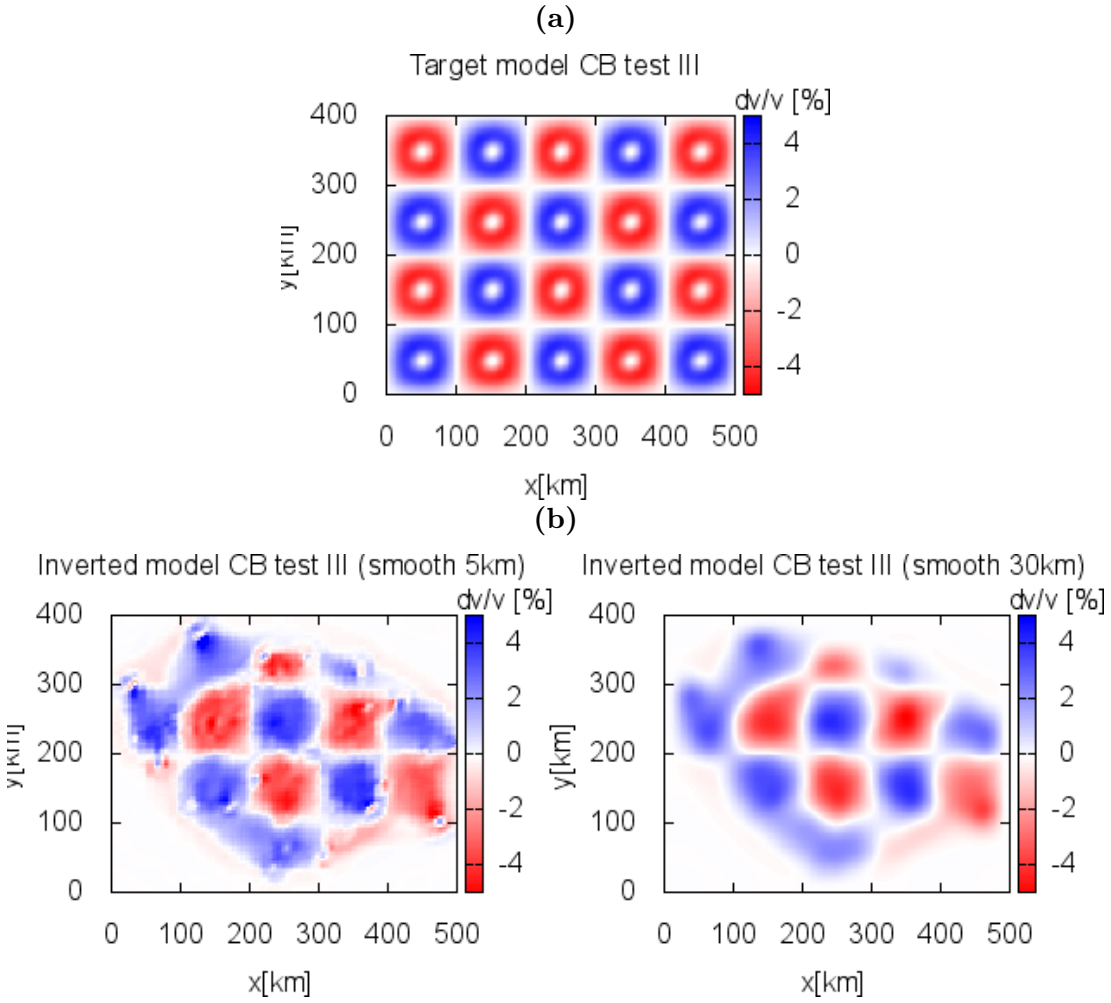

Figure 1.10: Checkerboard test III (both long and short-wavelength structures): a) target model, b) inverted models using 5 km (left) and 30 km (right) Gaussian smoothing.

tions are placed by the computational resources of the calculation of the seismic wave propagation and the reliability of the utilized data.

Although the adjoint tomography appears to be a powerful tool, it is not without drawbacks. The most significant one is the inevitable presence of the singularities in the sensitivity kernels (as presented in Section 1.4). This is usually treated by various regularization techniques, from which we have presented one employing the Gaussian smoothing of the sensitivity kernels.

Both strengths and weaknesses of the method were presented in this chapter by means of the synthetic checkerboard tests. However, for real data application the usage of the checkerboard tests as a tool for estimating the resolution power of the inversion is strongly criticized (Lévêque et al. (1993); Rawlinson and Spakman (2016)). In Chapter 2, which deals with real data inversion using adjoint method, we introduce synthetic tests employing two distinct reference models to assess the resolving power as well as the optimal regularization for the provided dataset.

Chapter 2

Choice of regularization in adjoint tomography based on two-dimensional synthetic tests

Published in Geophysical Journal International,
Volume 202, Issue 2, 787–799, August 2015, doi:10.1093/gji/ggv183

Lubica Valentová [1,2],František Gallovič[1],Bohuslav Růžek[3],Josep de la Puente[4],Peter Moczo [2]

Abstract We present synthetic tests of 2-D adjoint tomography of surface wave traveltimes obtained by the ambient noise cross-correlation analysis across the Czech Republic. The data coverage may be considered perfect for tomography due to the density of the station distribution. Nevertheless, artefacts in the inferred velocity models arising from the data noise may be still observed when weak regularization (Gaussian smoothing of the misfit gradient) or too many iterations are considered. To examine the effect of the regularization and iteration number on the performance of the tomography in more detail we performed extensive synthetic tests. Instead of the typically used (although criticized) checkerboard test, we propose to carry out the tests with two different target models – simple smooth and complex realistic model. The first test reveals the sensitivity of the result on the data noise, while the second helps to analyze the resolving power of the data set. For various noise and Gaussian smoothing levels, we analyzed the convergence towards (or divergence from) the target model with increasing number of iterations. Based on the tests we identified the optimal regularization, which we then employed in the inversion of 16 and 20 s Love-wave group traveltimes.

Keywords: Numerical approximations and analysis, Tomography, Seismic tomography, Europe

[1]Department of Geophysics, Faculty of Mathematics and Physics, Charles University in Prague, V Holešovičkách 2, 18000 Prague, Czech Republic

[2]Department of Astronomy, Physics of the Earth and Meteorology, Faculty of Mathematics, Physics, and Informatics, Comenius University in Bratislava, Mlynská dolina F1, 84248 Bratislava, Slovakia

[3]Institute of Geophysics, Academy of Sciences of the Czech Republic, Boční II/1401, 14131 Prague, Czech Republic

[4]Computer Applications in Science & Engineering, Barcelona Supercomputing Centre, Carrer del Gran Capitá 2-4, Barcelona, Spain

2.1 Introduction

With increasing computational power, the seismic tomography based on the so-called adjoint calculation of sensitivity kernels in 3-D models is becoming a common tool for improving our knowledge of the Earth's structure (e.g., Gauthier et al. (1986); Tromp et al. (2005); Liu and Tromp (2006); Fichtner et al. (2006), etc.). The greatest advantage of the adjoint method is the usage of numerical solution of the elastodynamic equation (EDE), meaning the least amount of simplifications (approximations) in the forward problem. To obtain the sensitivity kernel one needs only two calculations solving the EDE, which makes the method computationally feasible. Inversion is then performed by an iterative procedure of improving model parameters based on the kernel calculation and misfit between data and synthetics. Adjoint tomographic calculations usually employ full waveform information, that is surface and body waves by means of traveltimes (e.g., Tape et al. (2010)) or instantaneous phase misfits obtained by the time-frequency analysis (e.g., Fichtner et al. (2009); Colli et al. (2013); Rickers et al. (2013); Fichtner et al. (2013)). Both amplitude and phase misfits were used in the adjoint tomographic inversion, for example, of North-Atlantic and Europe by Zhu et al. (2013). The adjoint method is closely related to the scattering-integral method (Zhao et al. (2005); Chen et al. (2007*a*)), which is under certain conditions (e.g., large number of sources) even more efficient (Chen et al. (2007*b*)).

In this study, the adjoint tomography is combined with traveltime measurements originating from the ambient-noise cross-correlations. It has been shown that by the cross-correlation of diffuse wave fields between two receiver points, the Green's function between receivers may be extracted (e.g., Weaver and Lobkis (2002)). Many studies have been devoted to the extraction of Greens' functions, see, for example, Shapiro and Campillo (2004); Bensen et al. (2007).

The ambient-noise based Greens' functions are usually dominated by surface waves. Furthermore, there were also successful efforts to extract body waves (e.g., Gouédard et al. (2008); Zhan et al. (2010); Poli et al. (2012*b*,*a*); Lin et al. (2013*a*); Lin and Tsai (2013); Boué et al. (2013)).

Nevertheless, in most cases the application of the ambient noise measurement is focused on the surface wave tomography considering only vertical component of the Rayleigh waves. Rayleigh wave group or phase velocity maps of different regions from ambient noise data have already been retrieved using tomographic methods, for example, Shapiro et al. (2005); Yang et al. (2007); Moschetti et al. (2007); Verbeke et al. (2012). Love wave tomography was performed by, for example, Bensen et al. (2008).

Surface wave tomography aiming to retrieve 2-D surface wave velocity maps from ambient noise data usually employs the ray methods for traveltime calculation (e.g., Barmin et al. (2001)). Several surface wave tomography studies that compared the results obtained using the ray tomography with the Gaussian smoothing constraint with those using the finite-frequency kernels from the scattering theory (Born approximation) have found no significant improvement with the finite-frequency method (e.g., Sieminski et al. (2004); Boschi (2006); Zhou et al. (2005)). However, other authors (e.g., Ritzwoller et al. (2002); Yang and Forsyth (2006)) have claimed significant improvement when using the finite-frequency sensitivity kernels. Peter et al. (2009) used the membrane wave approx-

imation instead of Born approximation to simulate finite-frequency surface wave propagation. They demonstrated significant improvement by comparing with the ray tomography results in synthetic tests in case of short-wavelength heterogeneities, but only marginal improvement for real data applications. Trampert and Spetzler (2006) ascribe the contradictory results between the different authors to the importance of regularization effects in both the finite-frequency and ray approaches.

Recently, first attempts to perform full 3-D adjoint tomography employing ambient noise data have been made by, for example, Xu et al. (2013); Chen et al. (2014); Gao and Shen (2014). However, they come with large computational expense, making any extensive synthetic testing difficult. Less expensive approach is the more traditional one mentioned above, where one first inverts for 2-D surface wave velocity maps at distinct periods, which are then interpreted in terms of a 3-D structure.

Here we employ the iterative 2-D adjoint tomography method with the membrane wave approximation. Despite the method provides 2-D finite-frequency sensitivity kernels, an additional smoothing of a kernel is commonly applied for reducing singularities (peaks) in the position of point sources occurring in the EDE, for example, Tape et al. (2007, 2010); Peter et al. (2011). Generally, the strength of smoothing depends on the level of noise present in the data. However, especially in the adjoint tomographic studies it is usually chosen 'ad-hoc' and its impact on the inversion results is not properly analysed.

Another technical problem occurring in the iterative methods is the choice of the number of iteration steps. Both – amount of smoothing and the number of iterations – are affecting quality of the resulting model. For example, application of the smoothing function with a small width or too many iterations lead to rather complex models, which merely translate the noise present in data into artefacts in the model. On the other hand, oversmoothing or stopping the iteration process too early results in oversimplified models, generally loosing a large amount of information in the data.

In this paper, we perform a surface wave adjoint tomography employing the Love group traveltimes obtained from the ambient noise cross-correlations across the Czech Republic. In order to identify the proper smoothing strength and the optimal number of iteration steps, we perform synthetic tests with data corrupted by noise estimated from a real data set. Instead of using standard (albeit criticized) checkerboard test, we propose synthetic tests for two target models: a simple smooth model and a model with small-scale heterogeneities. Using the first model we investigate occurrence of artefacts due to data noise. Using the second model we test reliability of inferring small-scale structures. Combining all synthetic tests we achieve the best relation between the quality of the obtained model and the regularization in terms of (i) the amount of the smoothing and (ii) number of iterations.

After understanding and discussing both the limits and benefits of our method, we apply our methodology to the real data set, that is the 16 and 20 s Love-wave group traveltimes in the Czech Republic.

2.2 Data

The data consist of surface wave traveltimes obtained using ambient noise cross-correlations. The ambient noise data were recorded at stations located in the Czech Republic as well as in the adjacent border regions. The 54 stations (see Fig. 2.1a) include permanent stations of the Czech Regional Seismological Network (CRSN) and Virtual European Broadband Seismic Network (VEBSN), and temporary stations, which operated within passive experiments PASSEQ (Wilde-Piórko et al. (2008)), BOHEMA I, II and III (Plomerová et al. (2003); Babuška et al. (2005)). The stations were equipped with broad-band sensors, most of them with STS-2, few with Guralp. Three components were recorded continuously with the sampling frequency 20 Hz. More details on the processing noise data for obtaining surface wave traveltimes can be found in Appendix A.1 and in the paper by Růžek et al. (2012).

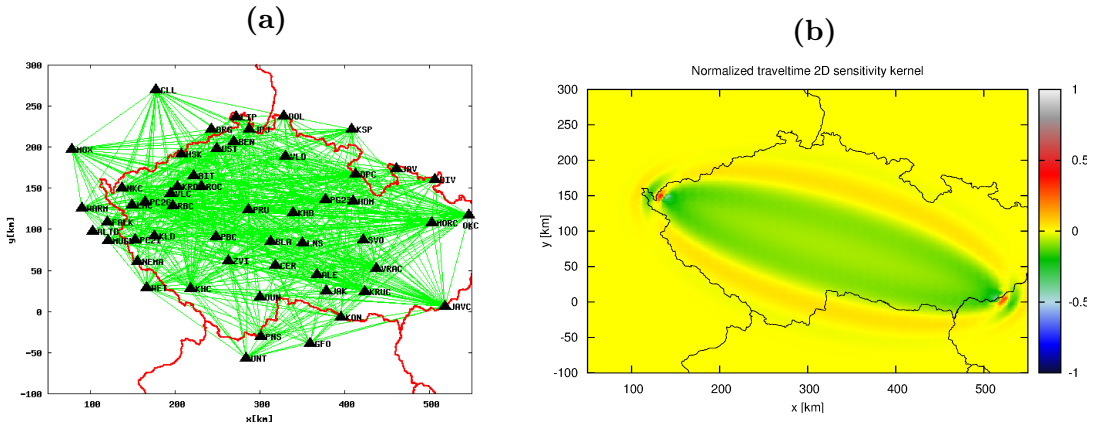

Figure 2.1: (a) Positions of stations used in the ambient noise processing. The stations are connected with lines along which traveltimes of the 20 s Love waves for the station pairs were obtained. The figure demonstrates the almost perfect data coverage of the studied domain. (b) Normalized traveltime sensitivity kernel for membrane waves with dominant period 20 s between two selected stations.

Because not all stations were in operation simultaneously, the maximum number of station pairs with available ambient noise data is only 819. Altogether 5525 Love-wave dispersion data were picked in the period range of 2–20 s. The traveltime values corresponding to a station-pair dispersion curve at a given period serve as input data in our inverse problem. In this work, we utilize only the longest wavelength Love-wave data corresponding to the periods of 20 and 16 s. Fig. 2.1a shows the 20 s Love-wave data coverage. Each of the 568 lines connects two stations with estimated traveltime value. Fig. 2.1b shows an example of a 2-D sensitivity kernel corresponding to the 20 s Love wave between two stations indicating an areal extent of the sensitivity of the waves used.

2.3 Method

2.3.1 Forward problem

The data are the cross-correlation traveltimes at 16 and 20 s (representing rather long periods in crustal studies). According to Fig. 2.1b, the corresponding waves have a wide-area sensitivity to the structural model. Therefore, it is desirable that the method employed for forward calculations can model the finite-frequency effects. Thus, we apply the membrane-wave approximation to model the surface wave propagation (e.g., Tanimoto (1990); Peter et al. (2007); Tape et al. (2007)). With 2-D modelling, the approach enables extensive synthetic tests which are not feasible in 3-D. We favour this over the ray method because it enables us to use structural models with strong small-scale heterogeneities which pose a serious problem for the ray-tracing methods.

In particular, when considering only Love waves, the membrane-wave approximation leads to solving scalar wave equation in 2-D. Furthermore, we assume homogeneous density distribution with arbitrarily chosen value. The perturbations in the group velocities are then interpreted using perturbations in parameter μ appearing in the membrane-wave equation (e.g., Tape et al. (2007)). We use Ricker wavelet (centred at a considered period) as a source-time function. It represents any waveform of a given frequency content. Therefore, since we work with the group surface wave traveltimes at periods of 20 and 16 s, the inferred velocity maps (see below) correspond to group surface wave velocity heterogeneities at the corresponding periods.

2.3.2 Inverse problem

Here we invert the traveltime residuals at each period independently (noise cross-correlation signal was bandpass-filtered for each period separately). The misfit is defined as the L2 norm of the weighted cross-correlation traveltime residuals ΔT_i,

$$\chi = \frac{1}{2} \sum_i h_i \Delta T_i^2. \tag{2.1}$$

The sum is taken over all measured traveltime residuals (station–station pairs) at the given period, h_i represents the traveltime weight. In our application, we assign $h_i = 0.5$ if the receiver acts both as source and receiver (to prevent from having duplicate data), otherwise $h_i = 1$. The traveltime residual ΔT_i between the synthetic seismogram u_i and observed seismogram u_i^0 is given as the time of their cross-correlation maximum:

$$\Delta T_i = \arg\max_t \int u_i(\tau) u_i^0(t+\tau) d\tau. \tag{2.2}$$

When interpreting the complete cross-correlation waveform, the correct approach to incorporate source in the adjoint calculation would follow Tromp et al. (2010). However, since our data consist of noise cross-correlation traveltime values only (i.e., complete observed waveforms are not used), we use the following simplified approach. We create an 'observed' waveform considering a point source with

Ricker-wavelet time function in a homogeneous medium with velocity equal to the station-station distance divided by the measured traveltime. The traveltime of this waveform then corresponds to the traveltime obtained from the noise cross-correlations. Ricker-wavelet source is also employed to generate synthetic waveforms in the adjoint inversion.

The objective of the inverse problem is to find model parameters for which misfit χ is minimal. This is accomplished using the conjugate gradient method for which the misfit gradient must be evaluated. The misfit gradient in direction δm is calculated using the Fréchet derivative kernels K_i, defined as

$$\delta_m \chi = \sum_i h_i \Delta T_i \int_V K_i \, \delta m \, dV. \tag{2.3}$$

We apply the adjoint method to calculate kernels K_i. In case of membrane waves, the kernel for the parameter μ corresponding to the given traveltime residual ΔT_i is given by (e.g., Fichtner et al. (2006))

$$K_i = \int_t (\nabla u_i) \cdot (\nabla u_i^\dagger) dt, \tag{2.4}$$

where u_i represents the forward wavefield and $u_i^\dagger$ the so-called adjoint wavefield. The adjoint wavefield $u_i^\dagger$ is calculated by back-propagating the wavefield from the adjoint sources $f_i^\dagger$ (Luo and Schuster (1991))

$$f_i^\dagger = -\frac{\dot{u}_i}{\int \dot{u}_i{}^2 dt} \delta(x - x_i), \tag{2.5}$$

where x_i stands for the position of the receiver. This means that the adjoint source is a point source located at the receiver's position with the time function given by the normalized synthetic velocity $\dot{u}_i$ from the forward calculation.

The adjoint method was implemented into software package SeisSol2D. The forward calculation is carried out by the Discontinuous Galerkin method with the Arbitrary High Order Time Derivatives (ADER-DG) on unstructured meshes (Käser and Dumbser (2006); Dumbser and Käser (2006); Käser et al. (2007); Dumbser et al. (2007); de la Puente et al. (2007)). The model parameters follow the same triangular computational mesh given by the numerical solver of the wave equation. Model parameters are constant in the elements. The total number of elements is over 7000.

2.3.3 Iteration scheme and regularization

According to eq. (2.3), the misfit gradient is obtained as a sum of kernels multiplied by $\Delta T_i h_i$. The conjugate gradient method (specifically Polak–Ribiére scheme) is applied for iteratively improving the model.

As commonly applied in the adjoint methods, the misfit gradient is convolved with the Gaussian bell. This efficiently removes the waveform singularities at the sources and receivers – see Fig. 2.1b, where an example of the kernel for 20 s waves is shown. No other advanced preconditioning, for example, the source subspace

projection (Tape et al. (2009)) or approximate Hessian (Chen et al. (2007*b*)), was applied for the sake of simplicity.

The step size taken in the descent direction is calculated as a minimum of a parabola fitted to at least three test models obtained by perturbing the model in the descent direction.

2.4 Synthetic tests

Using the synthetic tests we investigate two effects on the convergence towards a correct model: the effect of (i) noise level added to synthetics and (ii) spatial Gaussian smoothing of the gradient.

To quantify model improvement, we define the model misfit as the L2 norm of the difference between the obtained model m_{n} at iteration n and target (i.e., true) model m_{targ} normalized by the L2 norm of the initial model m_0:

$$\zeta = \frac{||m_{\mathrm{n}} - m_{\mathrm{targ}}||}{||m_0||} \cdot 100\% \tag{2.6}$$

Usually the curve has a local minimum. We denote the model corresponding to the minimum as the optimal model. Note that the model misfit is unknown in real applications because we do not know the true model.

In our synthetic tests we use two models: Model I represents a smooth structure, Model II contains strong small-scale heterogeneities. Both models originate from the model obtained by the preliminary inversion of the 20 s Love-wave data. Note that our models are closer to reality than, for example, the checkerboard model (standardly used in the seismic tomography despite criticism, e.g., by Lévêque et al. (1993)).

The data coverage, that is the source-receiver configuration used in the synthetic tests, is the same as in the case of the real 20/16 s Love-wave data inversion. This may additionally help to distinguish which areas show stable results and are not so much affected by the errors in the data or insufficient data coverage. The analysis of the areas with stable results is useful when interpreting results of real data application. We also perform an additional test with a modified station distribution.

2.4.1 Target models

The tests are performed using two different target models:

- Model I: smooth model with small maximum amplitudes of the heterogeneity.
- Model II: complex model with pronounced small-scale structures and large maximum amplitudes.

The models are based on the real group velocity model in the studied region. Model I was created by smoothing Model II by convolution with the spatial Gaussian bell of 200 km width. This suppresses the small-scale structures in Model II and reduces the amplitudes of heterogeneities. Model II was obtained by

the adjoint inversion of the 20 s Love group traveltimes with an initial parameter setting. It is suitable to test the inversion for the short-wavelength structures in Test II (see Fig. 2.2b).

The two target models used to generate synthetic data for the tests are shown in Fig. 2.2. To reduce the time demand of the computations, not all receivers are used as sources. The stations acting simultaneously as the point source in forward calculations are marked with a circle. Note that the corresponding source–station pairs are downweighted by $h_i = 0.5$ in the data misfit; see Section 2.3, eq. (2.1).

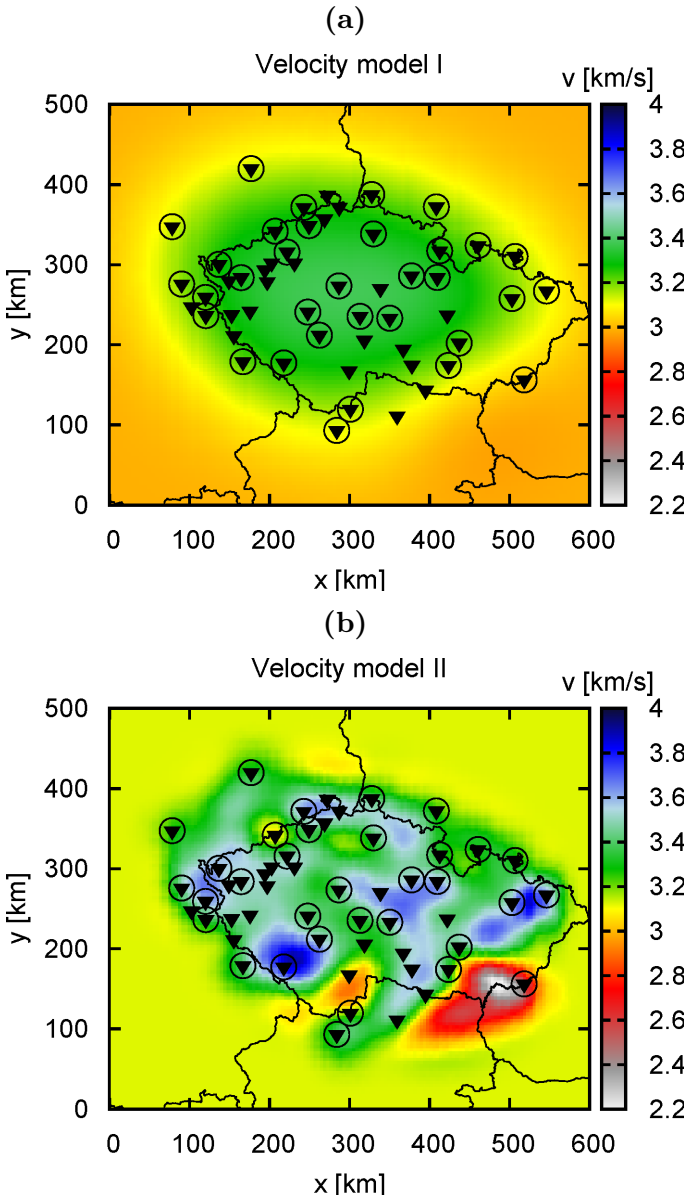

Figure 2.2: Velocity (target) models I (a) and II (b) for synthetic tests using 20 s data. Stations and sources are shown by inverted triangles and circles, respectively.

2.4.2 Synthetic data errors and regularization

We employ the 16 and 20 s Love-wave group traveltimes to estimate data error to be used in the synthetic tests. Because they are determined by almost the same structure, the differences between the corresponding values do not originate from the differences between the physical models but from the errors of the traveltime estimation.

Fig. 2.3a shows differences between the 20 and 16 s Love-wave group traveltimes as functions of distance. The absence of an obvious correlation with distance supports our assumption on the origin of the differences. The histogram of the differences between the 20 and 16 s Love-wave group traveltimes is shown in Fig. 2.3b. The histogram is fitted by a Gaussian distribution centred close to 0 with $\sigma_t = 1.53$. This additionally confirms that the differences between the data sets are random, originating rather from the data processing than from properties of the real structure.

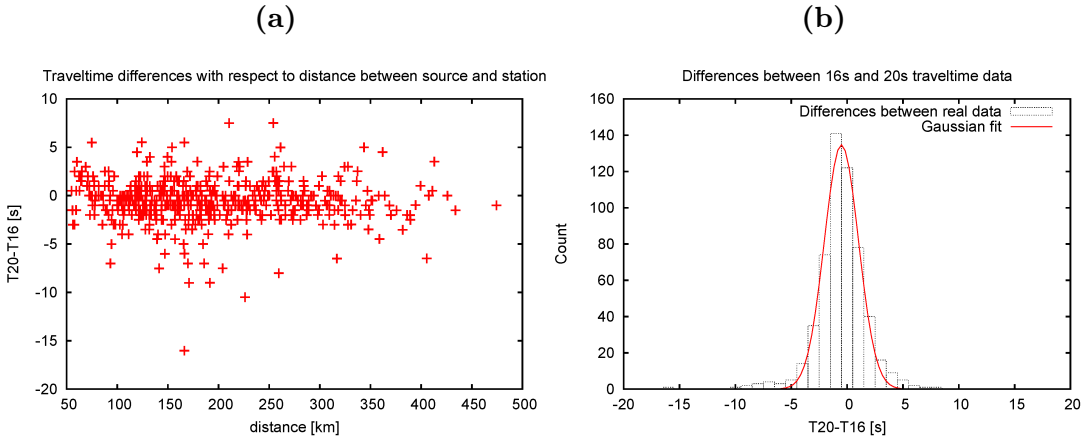

Figure 2.3: Differences between the 20 and 16 s Love-wave group traveltimes plotted in terms of (a) station-pair distance and (b) differences displayed as a histogram and fitted by the Gaussian distribution.

In the following tests, we use three values of the noise level specified in terms of the standard deviation. One of them corresponds to the real noise-level σ_t. The other two levels are chosen for analysing the effect of smaller and larger data noise considering $1/3\sigma_t$ and $3/2\sigma_t$, respectively. The Gaussian distributions for all three noise levels are shown in Fig. 2.4. The 'accurate' synthetic seismograms calculated for models shown in Fig. 2.2 are shifted by a value generated randomly from the corresponding Gaussian distribution. The waveforms themselves are not perturbed.

We smooth the calculated misfit gradients by means of convolution with a 2-D isotropic Gaussian function. We consider three widths (denoted as σ_x), 50, 100 and 150 km, representing different strengths of the smoothing. The smallest width corresponds to the wavelength of the 20 s data (or little less). This might be considered the natural choice because it prevents the smaller-than-wavelength structures without oversmoothing. The other two smoothing levels represent two different degrees of over-regularization. With increasing level of smoothing the smaller-scale heterogeneities should be suppressed and the resulting model is expected to contain less detailed structure.

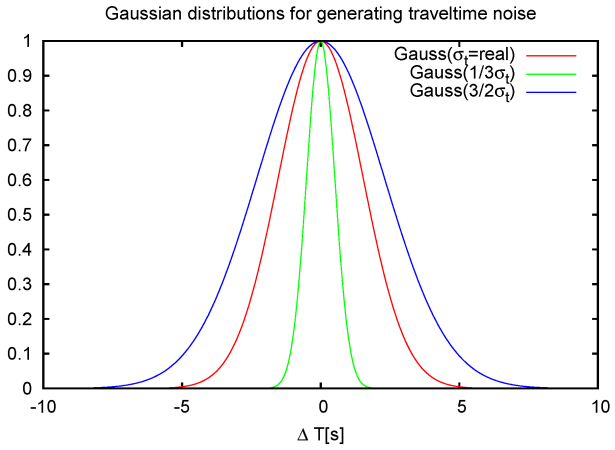

Figure 2.4: Gaussian functions used for generating traveltime noise in the synthetic tests: red, green and blue curves correspond to the Gaussian distributions with the standard deviations σ_t (from real data), $1/3\sigma_t$ and $3/2\sigma_t$, respectively.

2.4.3 Test I - inversion of the long-wavelength structures

The results for Test I are shown in the left column of Fig. 2.5. The individual rows show results for the three gradient smoothing levels σ_x (50, 100 and 150 km) with distinct colours and symbols. The decrease of data misfit (traveltime residual RMS) for the three noise levels is plotted in grey using the respective symbols. Fig. 2.6 shows several model examples at selected iteration steps and for different levels of the gradient smoothing σ_x. Note that the areas with negligible model update are masked.

For the weakest smoothing $\sigma_x = 50$ km (Fig. 2.5a), the optimal model was achieved in 1–3 iteration steps for all noise levels. Fig. 2.6a shows the optimal model (iteration 3) for noise level σ_t and smoothing $\sigma_x = 50$ km. In the case of the lowest noise level the results remain stable with further iterations and artefacts do not appear (Fig. 2.6c). For other noise levels, the model misfit starts to increase from the optimum. This is due to emergence of small-scale false structures originating from the noise in the data. The stronger the noise, the more pronounced heterogeneities are obtained (compare Figs. 2.6b and 2.6d), which results in the increase of the model misfit.

With the increasing level of the gradient smoothing, the results stabilize in the optimum after a larger number of iteration steps (Figs. 2.5c and 2.5e). Fig. 2.6e shows the model example for $\sigma_x = 100$ km and data noise σ_t obtained after the seventh iteration. The best results in Test I were achieved when the strongest smoothing ($\sigma_x = 150$ km) was applied, even for the highest level of noise. This is due to the fact that the target model is very smooth. The example model obtained after the seventh iteration for noise level σ_t is shown in Fig. 2.6f. Note that the long-wavelength structures in all the inferred models in Fig. 2.6 are similar because they are obtained in the initial iterations.

To sum up Test I, if data are corrupted by low noise level, the results are stable and do not depend on the amount of regularization (i.e., smoothing level and number of iterations). If the noise level is higher, the smoothing corresponding to the wavelengths used, that is $\sigma_x = 50$ km, is mostly inadequate (Fig. 2.6d). The inversion results in model with false structures despite the almost perfect

station coverage. Therefore, a certain degree of over-regularization is necessary for obtaining a more stable result. However, in the case of very high data noise, it may be still insufficient and the inversion should be stopped after just a few iterations. Assuming that noise in the real data is close to the noise σ_t estimated from differences between 20 and 16 s Love-wave data, the most reliable results are obtained with smoothing width of at least 100 km and up to the sixth iteration.

2.4.4 Test II - inversion of a realistic structure

The evolution of the model misfit and data misfit during iterations of Test II is shown in the right column of Fig. 2.5. Gradient smoothing σ_x increases in Fig. 2.5 from the top to bottom. For illustration, Fig. 2.7 shows several examples of models obtained during the inversions.

Common characteristic of both tests is the overall behaviour of the model misfit with iterations: Except for the cases with the lowest noise level, the divergence from the target model (expressed by the increase in the model misfit) emerges after achieving the optimum. The main difference from Test I is the generally higher number of iterations needed to achieve the optimum, namely 6–8. In other words, since the target model contains smaller-scale structures than that of Test I, more iterations are needed in order to obtain the main features of Model II. Figs. 2.7a and 2.7b show models obtained after iteration 3 and 7 (optimal), respectively, considering the same parameters (σ_t noise level and $\sigma_x = 50$ km). The inferred structures differ mainly in the value of the maximum amplitudes: the inversion is not able to reveal smaller-scale structures with correct amplitudes in the initial phase of the inversion.

To test solely the effect of smoothing, we have also performed a noise-free test (not shown here). For all σ_x the amplitudes were increasing with increasing iterations, still the correct values were not achieved. This was clearly controlled by the smoothing constraint that generally blurs the structures. In particular, for smoothing $\sigma_x = 50$km the amplitudes of the heterogeneities obtained by the inversion of the noise-free and noisy data are similar only during the first several iteration steps. In later steps, the artefacts of noisy data occur and inversion of the noise-free data performs much better. Obviously, when the gradient smoothing is increased, the results stabilize close to the optimum for more iterations, see Fig. 2.5d. For the smoothing of 100 km and 150 km, the amplitudes of the heterogeneities are similar for all iterations, both for the noise-free and noisy data inversion.

The problem arises with application of the strongest smoothing of 150 km, when the recovered model is bound to contain only very long wavelength structures, see Fig. 2.7f. This is observed as the higher value of the model misfit in the optimum as compared with other cases, Fig. 2.5f.

It is important to note that none of the smallest scale structures of Model II were correctly resolved for any of the cases considered, not even considering noise-free data. These structures emerge during very late iteration steps (much later after the optimum is achieved) and are, therefore, most vulnerable to adverse effects of the data noise. This should be always taken into account when the model with too many details is found.

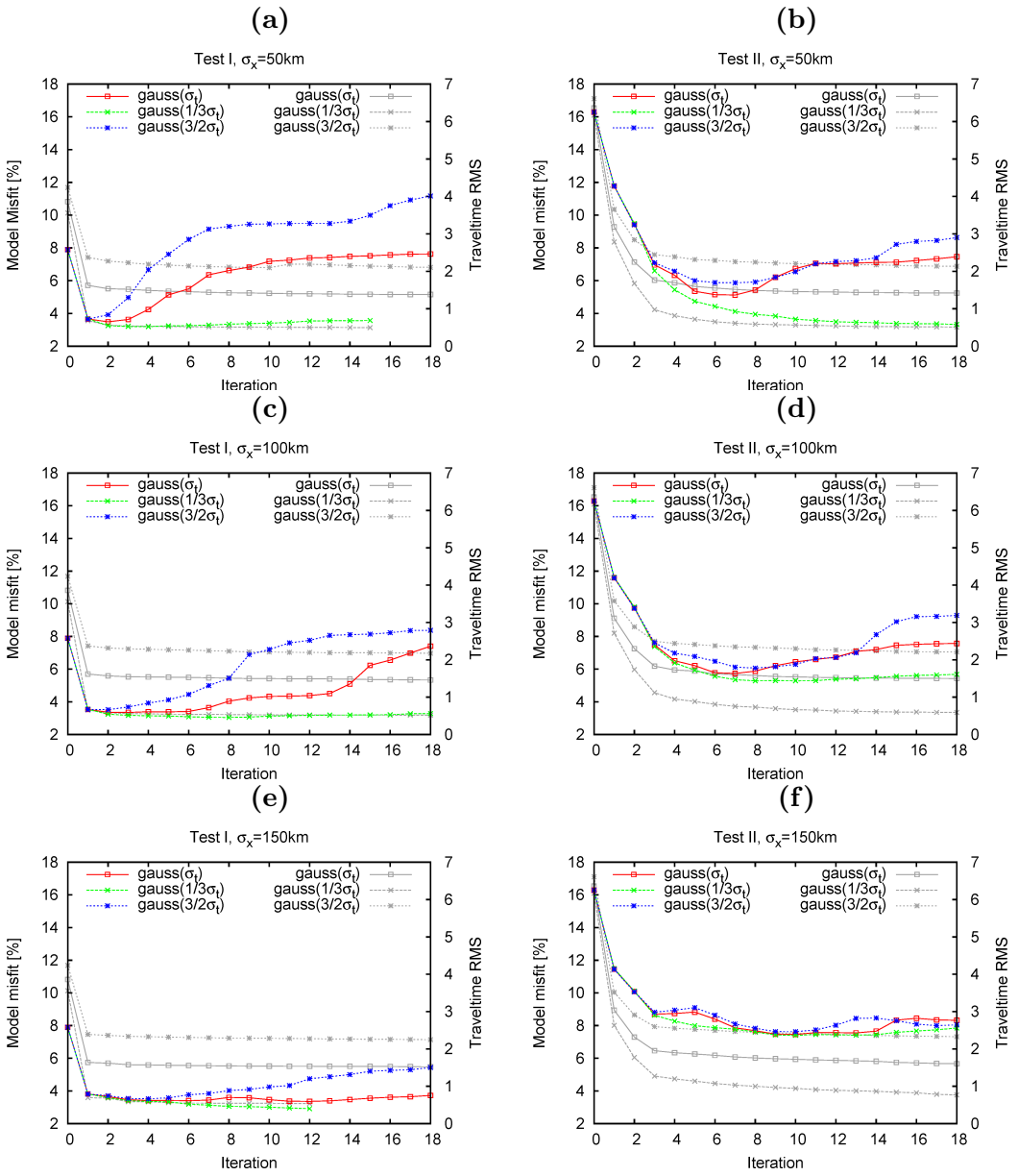

Figure 2.5: Results of Test I (left column) and Test II (right column). Colour curves and left vertical axis: convergence towards the target model represented by the model misfit. Grey curves and right vertical axis: data misfit in terms of traveltime residual RMS. Each row corresponds to a different level of the gradient smoothing: 50 km (top), 100 km (middle) and 150 km (bottom). The individual lines and symbols correspond to the different levels of noise applied to the synthetic data.

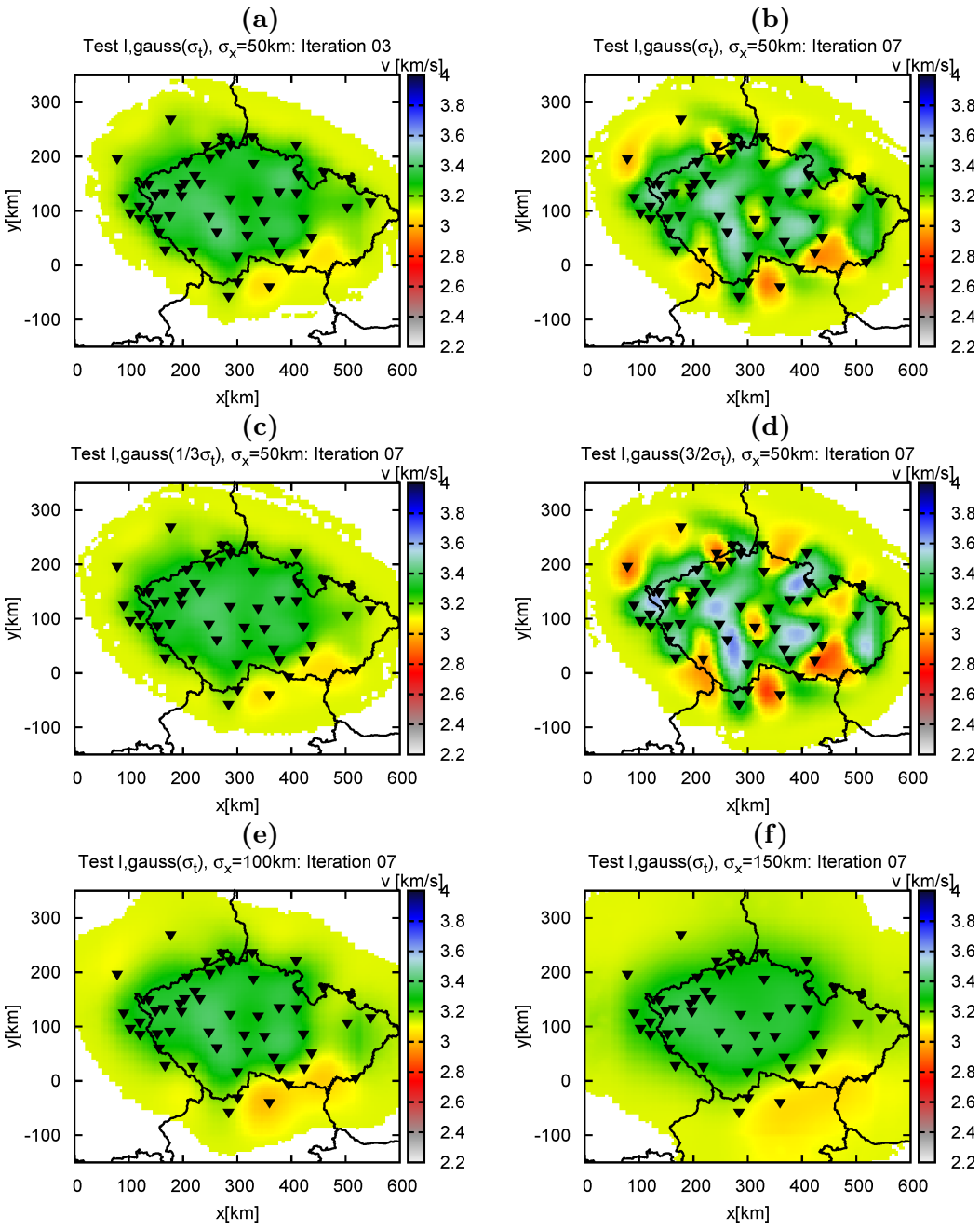

Figure 2.6: Models obtained in Test I for various noise levels (σ_t,$1/3\sigma_t$ and $3/2\sigma_t$; see Fig. 2.4), various smoothing levels σ_x and iterations.

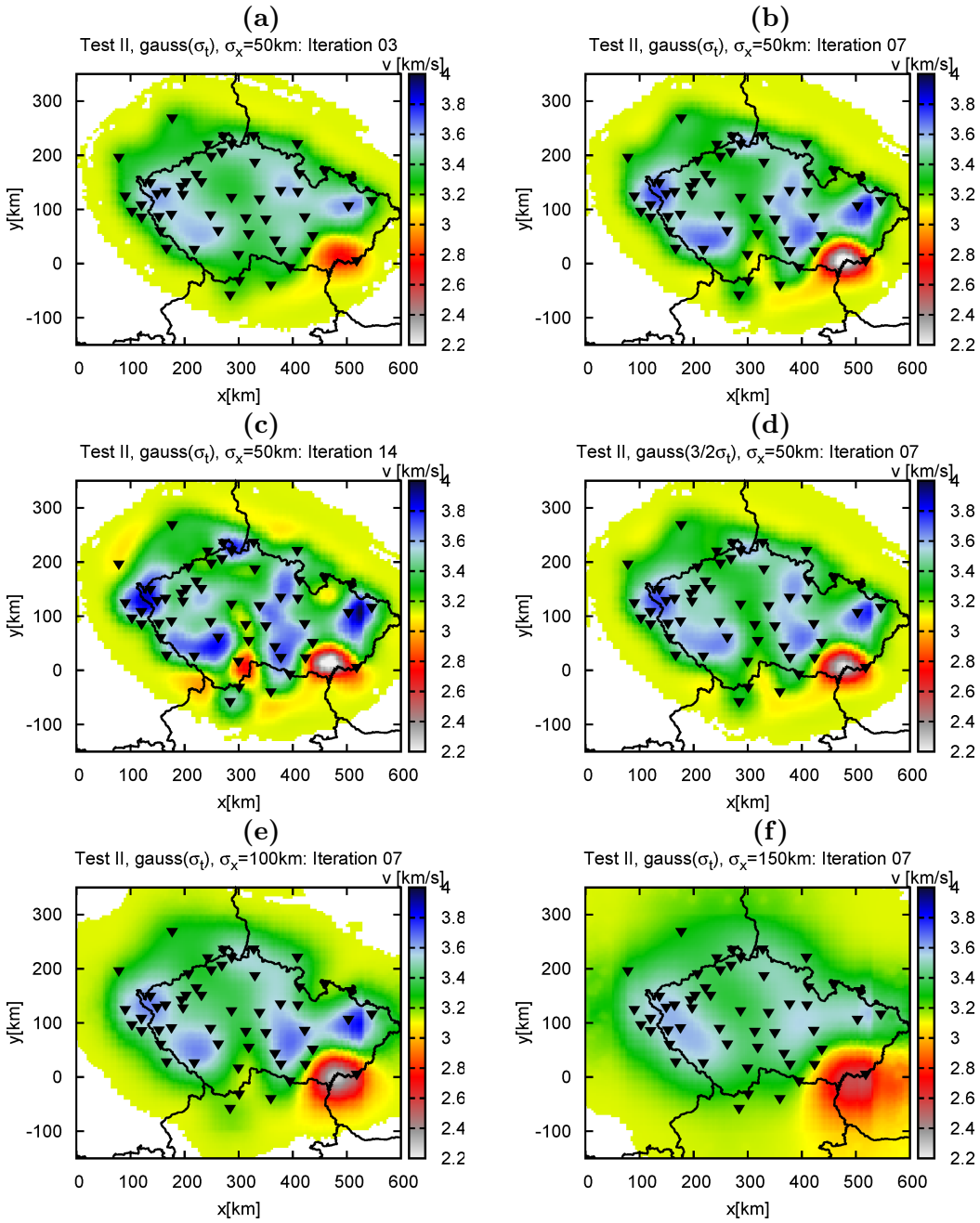

Figure 2.7: Same as Fig. 2.6, but for Test II.

To examine whether the conclusions of the previous tests are not biased by using the same station configuration in all cases, we perform a synthetic test with a different station configuration but with the same target model. Synthetic data are generated using Model II and noise level σ_t. The width of smoothing function σ_x applied to the misfit gradient is 50 km. The number of sources and receivers is unchanged, the locations of stations are simply swapped north to south and vice versa.

The results of the inversion are shown in Fig. 2.8 together with the results of the inversion using original stations' positions for comparison. In particular, Fig. 2.8a shows the model misfit and the decrease of the data misfit with iterations. The decrease of the data misfit is almost identical for both configurations. The behaviour of the model misfit with iterations also confirms that the optimal models are obtained at iterations 6 and 7 as found above. The model misfit in the case of the modified station configuration is generally higher than the one from Test II. Fig. 2.8b shows the model obtained at iteration 7 (i.e., the optimal model). The model does not contain false heterogeneities. However, the inversion is unable to recover all structures due to the insufficient station coverage (see, e.g., strong low-velocity anomaly in the southeast). This also explains the increase in the model misfit.

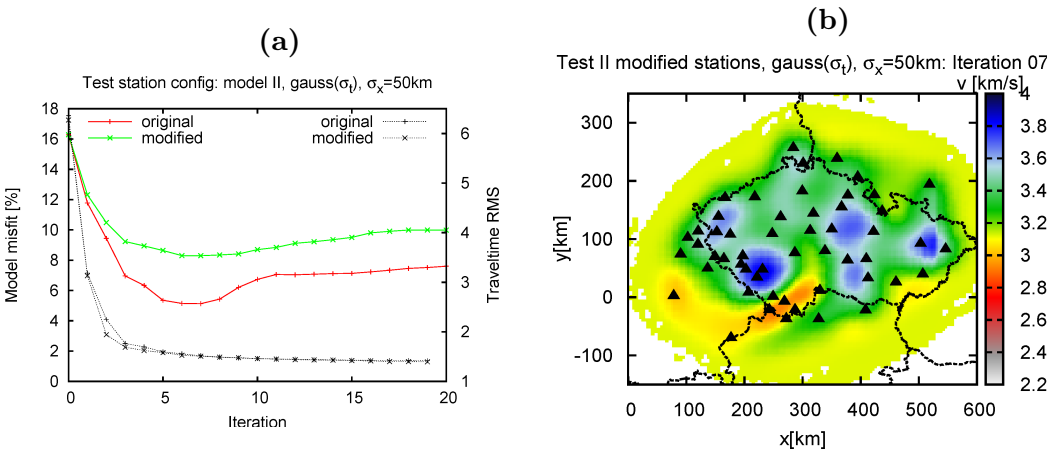

Figure 2.8: (a) Model and data misfit with iterations for the original and modified station configurations using Model II, noise level σ_t and gradient smoothing 50 km. (b) Model for modified station configuration in the minimum of the model misfit.

To conclude, Test II confirms several outcomes of Test I: in cases of the weaker smoothing, the false small-scale structures evolve during later iteration steps. However, extreme oversmoothing and/or too few iteration steps may lead to smoothed structures with underestimated amplitudes. The best results for Test II are achieved using gradient smoothing of maximally 100 km in 6–8 iteration steps. This conclusion is independent of the station configuration.

2.5 Real data example: inversion of the 20s and 16s Love-wave group traveltimes in the Czech Republic

We have applied the adjoint tomographic inversion to the 20 and 16 s Love-wave group traveltimes obtained by cross-correlation of the ambient noise on the territory of the Czech Republic.

We assume that the real data error corresponds to the variance σ_t, estimated from the differences between the 16 and 20 s traveltime data (Section 2.4.2). From the synthetic tests with the Gaussian errors (Section 2.4) it follows that the iteration process should be stopped after 6 iterations considering Gaussian smoothing width $\sigma_x = 100$km. If the noise level was underestimated and the real model contains small-scale structures, the result should be stable without any profound artefacts (Section 2.4.4). In case the real model does not contain small-scale structures and the data noise level is higher, the obtained model may suffer from false small-scale heterogeneities. However, the amplitudes of these anomalies should not be high (Section 2.4.3) and the artefacts should not be dominant features in the obtained models. When increasing the number of iterations above 6, the heterogeneities in the velocity image increase only in their amplitudes. Test II shows similar effects as a consequence of the data noise. Geological interpretations that consider only the shape of the heterogeneities would probably not be significantly affected. However, interpretations based on the local 1-D velocity profiles obtained from the inferred dispersion curves must carefully take this possibility into account.

The inferred models are shown in Figs. 2.9a and 2.9b for the 20 and 16 s waves, respectively. The recovered models are obviously similar which supports both the obtained results and the assumption that the differences between the two data sets originate in the measurement errors.

Figs. 2.9c and 2.9d show the decrease of the data misfit for the two periods considered. The distribution of the traveltime residuals is also shown (Figs. 2.9e and 2.9f). It is obvious that the greatest change in the data misfit is achieved already at the first iteration. However, some considerable improvement in the traveltime residuals is still evident up to iteration 3. There is only a small improvement during the next iteration steps. This behaviour suggests that the real structure is indeed more complicated than that of Test I.

Inversion using other frequency data and also the Rayleigh wave data together with geologic interpretations are the subject of our further study.

2.6 Discussion

We have analysed the choice of regularization parameters for the 2-D adjoint tomography using synthetic tests. We benefit from having two independently obtained data sets of close frequency content, namely the 20 and 16 s Love-wave group traveltimes from noise correlations. The differences between the data sets exhibit a Gaussian distribution. Assuming that these differences are mostly due

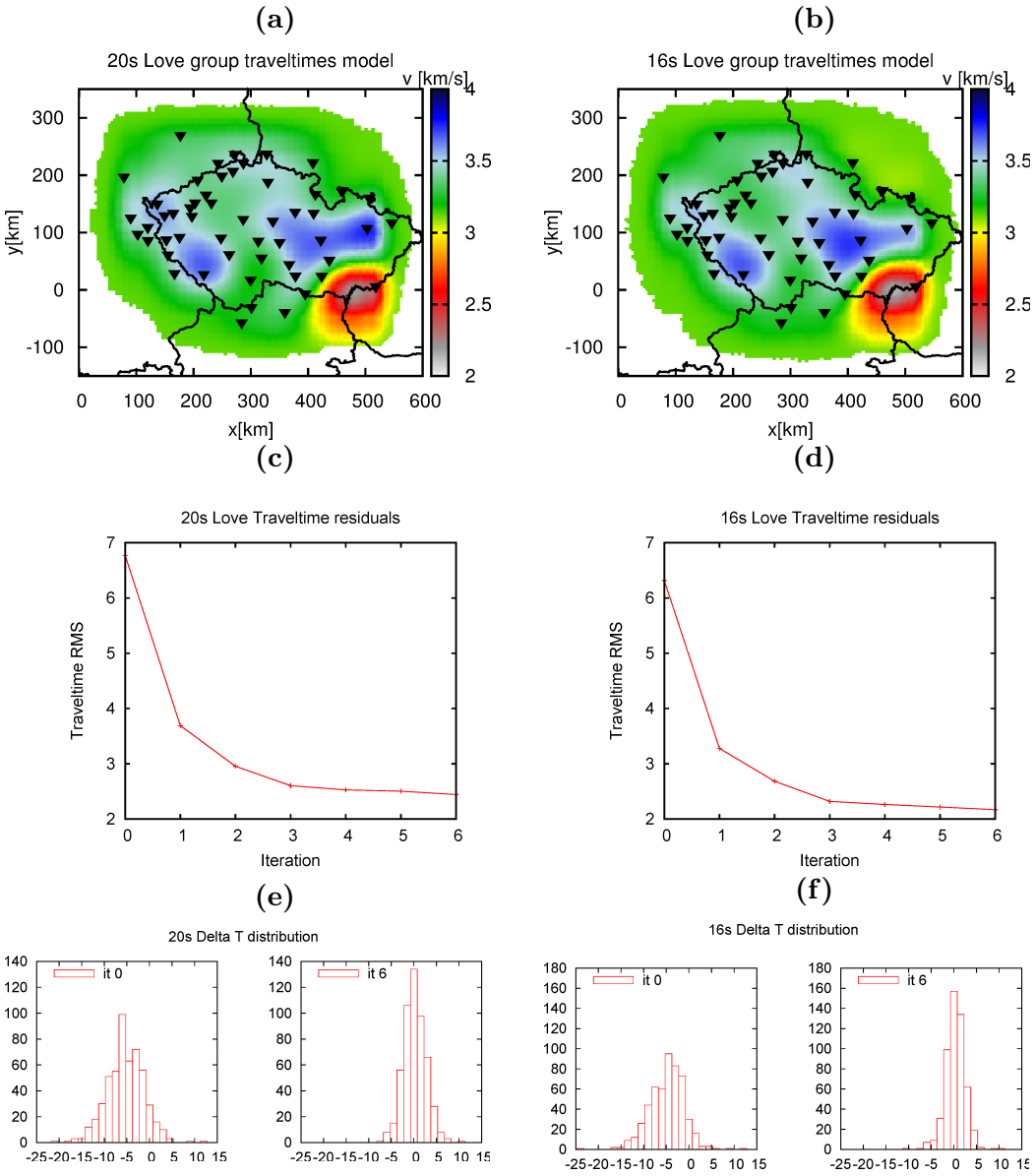

Figure 2.9: Results of real Love-wave data inversion: left 20 s; right 16 s. (a, b) Group velocity models; (c, d) traveltime residual RMS with iterations; (e, f) traveltime residual histograms for initial model and resulting model.

to the measurement error, their standard deviation is used as a reference data noise for the synthetic tests.

We have used the synthetic tests to identify the appropriate smoothing strength and the optimal number of iteration steps so that the inferred model contains not only large-scale but also small-scale heterogeneities without false structures. Instead of using the typical checkerboard test, we suggest performing synthetic tests with two very different structural characteristics – smooth (Test I) and complex (Test II) target models. We show that the optimal strength of regularization depends not only on the level of noise but also on the complexity of the target model. Furthermore, even the proper regularization does not necessarily stabilize the result in the optimum. Estimating where the obtained model is still reliable and where the false structures develop is the key purpose of our synthetic tests.

Generally, during the first iteration steps of the inversions the structural model is improved at the longest wavelengths and both the model and the data misfit decrease considerably. During further iterations, the shorter wavelength structures of the model are revealed. At this point the performance of the inversion depends on the noise and smoothing levels. For the lowest noise level the model does not evolve considerably. In case of strong noise and weak smoothing, the difference between the target and the inverted model grows with further iterations. This result may be surprising given the almost perfect station coverage. For the higher noise level the increase of the model misfit starts earlier and is steeper. Nevertheless, the data misfit keeps decreasing, showing that the inversion starts explaining the noise in the data by new artificial structures in the model.

A similar test where we changed the position of stations confirms that the conclusions are independent on the source–receiver coverage.

One important feature revealed by the synthetic tests is the behaviour of the data misfit with increasing number of iterations. In Test I, the data misfit improved significantly at the first iteration and remained almost unchanged for the rest of the inversion process. In Test II, the decrease of the data misfit exhibits more complex behaviour (resembling quadratic decrease) during the first few iterations and remains almost constant after iteration 4. The final data misfit value depends almost entirely on the level of noise applied to data. This means that the misfit behaviour with iterations gives indication not only of the noise level present in data but also of the complexity of the structural model with respect to the starting model.

We have also investigated application of the common L-curve method to identify the optimal iteration step for a given noise level and given smoothing (see Appendix A.2). Since it was difficult for us to draw conclusions based on the usual analysis of the L-curve, we prefer the above mentioned approach and rather combine both tests to obtain the optimal values of the gradient smoothing and number of iterations for our particular application.

The inferred optimal regularization parameter setting was applied to real data of the 20 and 16 s Love waves. In this way we obtained tomographic maps of the Czech Republic's shallow crust with the highest resolution and reliability possible, for the given data and method.

In the real data application, the decrease of the data misfit observed in Figs. 2.9c and 2.9d does not resemble the behaviour observed in Test I. This indicates that the real model is not extremely smooth, but contains rather smaller-scale

structures. The traveltime residual RMS value, to which the data misfit of the real-data inversion converged, is greater than 2 s which would correspond to the greatest noise level ($3/2\sigma_t$) in the synthetic tests. We note that when we underestimated the data noise, according to Test II the gradient smoothing $\sigma_x = 100$ km gave similar results for both noise levels up to iteration 13 (see Fig. 2.5d).

If the initial and target models are close, the first iteration of the nonlinear inversion may be considered as the result of the standard linearized inversion. Therefore, we are able to compare these methods reasonably only in Test I (the smooth model). In this case, almost the same model misfit was achieved at the first iteration for all noise levels and all regularizations. After this step (i.e., in the nonlinear part of the inversion), there is no considerable improvement with iterations; on the contrary, it may lead only to complex models formed by artefacts if improper smoothing is applied. However, when the initial model is not close to the target one (see our Test II), the model misfit does not reach the optimal value in the first iteration step. It is attained in further iteration steps (the nonlinear inversion). To sum up, the regularization, as applied in our problem, mostly affects the nonlinear part of the inversion.

We note that since the full 3-D adjoint inversions are computationally extremely expensive, the tests presented in this paper are feasible only in 2-D. We believe that they provide important insight into the method itself and reveal its main problems and limitations in general. From the presented numerical experiments, one may infer that the regularization (e.g., in the form of gradient smoothing) is recommended to be greater than the wavelength considered to prevent the bold structural artefacts. The proper number of iterations cannot be easily generalized because it strongly depends on the choice of the initial model. For applications similar to ours, the tests suggest that the number should be rather low ($\approx 5 - 10$). Otherwise, the model may be spoiled by the data noise artefacts even if the data coverage seems perfect.

2.7 Conclusion

The tomographic problem addressed in the present study is based on the iterative adjoint inversion of the Love-wave group traveltimes obtained from the ambient noise cross-correlations across the Czech Republic. In order to investigate the undesired regularization effects due to the Gaussian smoothing and the choice of the total number of iterations, we performed synthetic tests for two different target models – a simple smooth and a more complex heterogeneous model. We analysed effects of smoothing strengths and data noise levels. In particular, we used realistic noise levels derived from differences of the observed traveltimes at two adjacent periods (16 and 20 s). Tests with the simple model demonstrate the possibility of obtaining false small-scale structures even in areas with an ideal station coverage, when insufficient smoothing is applied or too many iteration steps are performed. Contrarily, the tests with the complex target model reveal the possible resolving power of the present data set. The tests made it possible to find the optimal regularization parameters for the investigated problem (100 km wide smoothing Gaussian and 6 iterations). The conclusions are relatively insensitive to the station distribution. Eventually, we applied the regularization parameter setting in the real data inversion of the Love-wave groups at the 16

and 20 s periods. The real data inversion results are very similar and show only minimal discrepancies.

Acknowledgments

We would like to thank Carl Tape and an anonymous reviewer for helpful and constructive comments. We gratefully acknowledge the funding by the European Union through the Initial Training Network QUEST (grant agreement No. 238007), a Marie Curie Action under the People Programme. This work was carried out under the HPC-EUROPA2 project (project number: 228398), with the support of the European Community - Research Infrastructure Action of the FP7. We would like to express our deepest gratitude to all that provided the data: CRSN network, all BOHEMA and PASSEQ experiments and other VEBSN stations. This work was partially supported also by MYGDONEMOTION APVV-0271-11 funded by the Slovak grant agency APVV. This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, as well as Czech Ministry of Education, Youth and Sports via the project Large Research, Development and Innovations Infrastructures (LM2011033). This research has been also supported by the Grant Agency of the Charles University under the project SVV-260218 and the Czech Science Foundation project 14-04372S.

Appendix

A.1 Ambient noise data processing

The ambient noise processing follows the procedures in Bensen et al. (2007). The first phase (i.e., the single-station data preparation) consists of data selection and basic preprocessing, such as demeaning and downsampling. In the next step, the temporal running-absolute-mean normalization using 150-s long window was applied to reduce the effect of the strong events in records.

During the next phase of the data processing, the records are cut into 1 hr segments and rotated to R-T-Z coordinates. The signals are then cross-correlated between the stations for every component to reveal the Greens' functions. All the obtained 1-hr long cross-correlations are summed to improve the signal-to-noise ratio (SNR). The shortest time interval of stacking the signal was 26 d, the longest was more than 8 yr, the mean interval is 2.66 yr. Only the Green functions with SNR$>$ 5 were used to estimate the traveltime data.

In the R-T-Z coordinate system, the transverse component corresponds to the Love wave Green's function. To obtain the dispersion measurements, narrow band-pass filters of given central frequencies were applied. The envelope of the filtered signal was calculated and the maximum of the envelope function was picked as the group traveltime corresponding to the filter frequency. Altogether 5525 Love-wave dispersion data were picked for all frequencies. The traveltime values corresponding to the dispersion curve at a given frequency of the station–station pair serve as input values in the inverse problem of this study.

A.2 L-curve criterion

One of the most common way to estimate the ideal value of necessary regularization is the so-called L-curve criterion. This criterion is based on the fact that the dependency of data misfit with respect to a model characteristic is supposed to have shape of letter L. The most suitable model is chosen as that corresponding to the corner of the L-curve, where the sufficient decrease of the data misfit is achieved by the model with the smallest complexity.

The gradient inversion method is known to change the model parameters on long wavelengths during the first iteration steps and the smaller-scale features of the model are obtained during the later steps. Indeed, the model complexity increases with iterations as shown in our Tests I and II. For several test examples, we demonstrate the performance of the L-curve and try to use this criterion to estimate which iteration gives the best result. The plot of data misfit with respect to the model complexity measured by the norm of the Laplace operator applied to the model, $||\nabla^2 m_n||$, for every iteration is used as the modification to the standard L-curve plot. The corner in the plot should reveal when to stop the iteration process.

The L-curves for the synthetic Tests I and II with the Gaussian noise of variance σ_t added to the synthetic data and different gradient smoothing levels σ_x are shown in Fig. 2.10. During the first few iterations, there is a sharp decrease of the data misfit with a small change in the model complexity. After that, the model complexity continues to increase while the decrease in the data misfit slows down. When the strongest smoothing is applied to the gradient (blue curves in Fig. 2.10), the curve tends to be more complicated containing one or several local edges. We assume that the model complexity is being artificially reduced during iterations while the data misfit is decreasing. This would also explain the absence of the edges in case of the weakest smoothing which corresponds to the original wavelength of the data set.

Let us remark that the optimal models for Tests I and II are achieved at iterations 2 and 6–8, respectively. After that, depending on the amount of smoothing, the model misfit starts to increase. Omitting the local edges in the curve, the optimal models determined according to the L-curve criterion are achieved for Test I at iteration 2 for 50 km smoothing, at iteration 3 for 100 km smoothing and at iteration 4 for 150 km smoothing. According to the results shown in Fig. 2.5 of the main text, these conclusions drawn according to the L-curve plot are reasonable. The optimal model according to the L-curve criterion for Test II would be at iteration 4 for 50 km smoothing, iteration 3 for 100 km smoothing and iteration 6 for 150 km smoothing. Comparing with Fig. 2.5, in case of a complex target model, the optimal iteration step suggested by the L-curve would be underestimated resulting in the oversmoothing the model. This analysis documents that the L-curve criterion is difficult to apply.

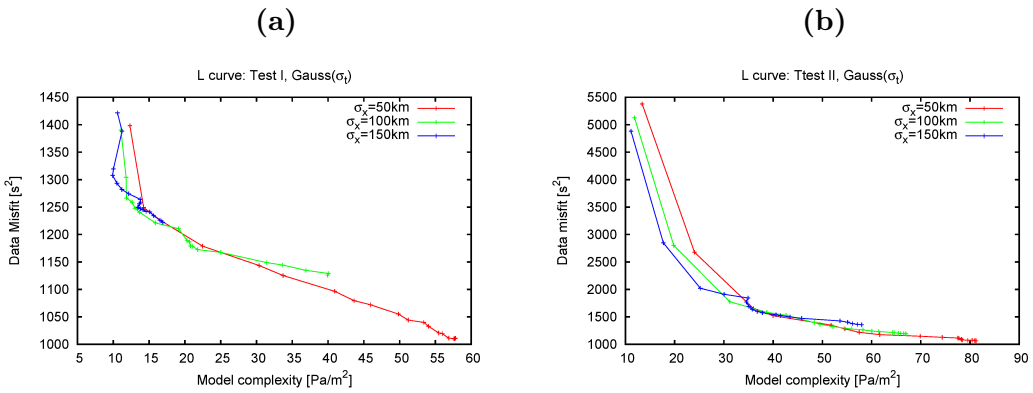

Figure 2.10: L-curve plot between model complexity and data misfit for Test I (a) and Test II (b); for Gaussian noise with standard deviation σ_t.

Chapter 3

Bayesian tomographic inversion of surface waves

The Bayesian solution of the inverse problem is defined as the posterior probability density function (PDF) on model parameter space. The solution of the problem, instead of one best fitting model, is usually given by a large number of model samples which may be used to estimate also the model parameter uncertainties, correlations and resolution. With increasing computational power, Monte Carlo (MC) methods for drawing samples on the model parameter space following the posterior PDF are increasingly used. The MC method may be considered as an alternative to grid search algorithm, which usually fails in high-dimensional model spaces. To increase the sampling efficiency, the algorithm is modified to improve the sampling in that parts of the model parameter space where the PDF attains significant value – so-called importance sampling.

Seismic tomography is an example of inverse problem in high-dimensional parameter space for which the MC sampling algorithms have been successfully applied. One of the more traditional representatives are the genetic algorithms, applied, for example, in crustal tomographic studies by Lomax and Snieder (1995). Seismic tomographic problems are often solved also by simulated annealing method (e.g., Pullammanappallil and Louie (1994)). Bodin et al. (2012) performed MC inversion to obtain a 1-D velocity model of South-East Australia. 3-D tomography combined with MC method was performed by Piana Agostinetti et al. (2015) for local earthquake data in Southern Italy or for ambient noise data in different regions by Young et al. (2013*b*); Galetti et al. (2017); Pilia et al. (2015).

In addition to the estimation of the uncertainty of the model parameters, the Bayesian approach may also give insight into the shortcomings of the employed methods. Galetti et al. (2015) used MC method to invert ambient noise surface wave data for a group velocity model employing ray method calculation of the surface wave traveltime. When analyzing the results, they found so-called 'uncertainty loops' surrounding the high velocity anomalies. These were interpreted as a consequence of employing ray method for the traveltime calculation.

Unlike the traditional approaches, there is no requirement on the linearity of the forward calculation in the MC algorithms. Therefore, the MC methods appear suitable for a wide variety of problems. However, their main drawback is the high number of forward simulations needed to determine the PDF value

for each tested model. Therefore, a very fast forward-problem solver is necessary, which presents the main limitation for the range of applications.

In this chapter, we introduce Bayesian approach to the inversion of surface wave dispersion curves into an S-wave velocity model. In the next section, we remind the formulae for Bayesian solution of inverse problem, define the model parameter space and briefly introduce the employed sampling algorithm (parallel tempering) and the method for forward calculation (matrix method). Then, we present the Bayesian solution of dispersion curve inversion by means of 1D synthetic tests. The tests enable us to explore the sensitivity of the employed data and effectivity of our sampling algorithm. The application to the 3D problem in the second step of the two-step surface wave inversion is shown in the next chapter for ambient noise tomography of the Bohemian Massif.

3.1 Methods

3.1.1 Bayesian solution of inverse problem

Tarantola (2005) defines the solution of the inverse problem as a posterior probability density function (PDF) on model parameter space $p(\boldsymbol{m})$ given as a conjunction of a priori information given by PDF $p_{\mathrm{prior}}(\boldsymbol{m})$, theoretical information usually expressed as conditional probability $\theta(\boldsymbol{d}|\boldsymbol{m})$ and data information from measurements with PDF $\rho(\boldsymbol{d})$ (i.e., independent on model prior and theoretical information). Further we assume that the model parameter space $\boldsymbol{m}$ is linear. Then the posterior PDF is

$$p(\boldsymbol{m}) = k_1 p_{\mathrm{prior}}(\boldsymbol{m}) \int \rho(\boldsymbol{d})\theta(\boldsymbol{d}|\boldsymbol{m})\,\mathrm{d}\boldsymbol{d}. \tag{3.1}$$

Although it is recommended to assume the theoretical information in this general form, in most cases the error due to the theoretical information (e.g., computational modelling) is considered much smaller than the measurement error and the theoretical information is approximated by Dirac distribution $\theta(\boldsymbol{d}|\boldsymbol{m}) \approx \delta(\boldsymbol{d} - \boldsymbol{g}(\boldsymbol{m}))$ where $\boldsymbol{d} = \boldsymbol{g}(\boldsymbol{m})$ is the forward problem,

$$\begin{aligned} p(\boldsymbol{m}) &= k_1 p_{\mathrm{prior}}(\boldsymbol{m}) \int \rho(\boldsymbol{d})\delta(\boldsymbol{d} - \boldsymbol{g}(\boldsymbol{m}))\mathrm{d}\boldsymbol{d} \\ &= k_1 p_{\mathrm{prior}}(\boldsymbol{m})\rho(\boldsymbol{g}(\boldsymbol{m})) \end{aligned} \tag{3.2}$$

This expression may be also easily obtained employing Bayes theorem.

Usually, the measurement uncertainties are considered in form of a Gaussian distribution with covariance matrix $\boldsymbol{C}_{\mathrm{d}}$ centered around the observed value $\boldsymbol{d}_{\mathrm{obs}}$

$$\rho(g(\boldsymbol{m})) \approx \exp\left(-\frac{1}{2}(\boldsymbol{d}_{\mathrm{obs}} - \boldsymbol{g}(\boldsymbol{m}))^T \boldsymbol{C}_{\mathrm{d}}^{-1}(\boldsymbol{d}_{\mathrm{obs}} - \boldsymbol{g}(\boldsymbol{m}))\right) \tag{3.3}$$

The model prior information p_{prior} is usually assumed either Gaussian centered around a prior value $\boldsymbol{m}_{\mathrm{prior}}$ with covariance matrix $\boldsymbol{C}_{\mathrm{m}}$, or as a special case with very large values on main diagonal of $\boldsymbol{C}_{\mathrm{m}}$ leading to constant PDF as no apriori information. The posterior PDF is then expressed

$$p(\boldsymbol{m}) = C \exp\left(-S(\boldsymbol{m})\right) \tag{3.4}$$

where $S(\boldsymbol{m})$ is called misfit:

$$S(\boldsymbol{m}) = \frac{1}{2}\left[(\boldsymbol{d}_{\text{obs}} - \boldsymbol{g}(\boldsymbol{m}))^T \boldsymbol{C}_{\text{d}}^{-1}(\boldsymbol{d}_{\text{obs}} - \boldsymbol{g}(\boldsymbol{m})) + (\boldsymbol{m} - \boldsymbol{m}_{\text{prior}})^T \boldsymbol{C}_{\text{m}}^{-1}(\boldsymbol{m} - \boldsymbol{m}_{\text{prior}})\right] \tag{3.5}$$

or in case of no apriori information ($\boldsymbol{C}_{\text{m}}^{-1} \to 0$), it is reduced to

$$S(\boldsymbol{m}) = \frac{1}{2}(\boldsymbol{d}_{\text{obs}} - \boldsymbol{g}(\boldsymbol{m}))^T \boldsymbol{C}_{\text{d}}^{-1}(\boldsymbol{d}_{\text{obs}} - \boldsymbol{g}(\boldsymbol{m})). \tag{3.6}$$

In the Bayesian framework, the inverse problem solution is represented by the full knowledge of the posterior PDF $p(\boldsymbol{m})$. When the forward problem is in addition linear, it can be shown, that the resulting posterior PDF is in form of a Gaussian function centered around the best/mean model $\boldsymbol{m}_{\text{best}} = \boldsymbol{m}_{\text{mean}}$ with posterior covariance matrix $\boldsymbol{C}$ (Tarantola (2005)). Therefore, if one estimates these two parameters of the Gaussian distribution, one obtains the complete solution of the inverse problem.

However, in case of nonlinear forward problem, the resulting posterior PDF may be very complex and obtaining the full posterior information may not be achievable. If the model parameter space is lower-dimensional, the grid search algorithm, i.e. calculation of PDF $p(\boldsymbol{m})$ for regularly distributed model parameters $\boldsymbol{m}$ over a priori chosen part of the model space, may be successfully applied to estimate the form of PDF. Nevertheless, for the most of the inverse problems (not excluding the tomographic problems) model parameter space is high-dimensional, therefore the evaluation of $p(\boldsymbol{m})$ is reduced to that parts of model space, where $p(\boldsymbol{m})$ attains more significant values. This is usually accomplished by employing the Monte Carlo method to generate the model samples.

3.1.2 Parallel Tempering algorithm

The parallel tempering method (PT, Earl and Deem (2005); Sambridge (2014)) belongs to Monte Carlo methods, that use random number generator to sample the model parameter space. The so-called Metropolis-Hastings algorithm (Metropolis et al. (1953)) ensures that the sampling follows the posterior PDF.

As in the most MC algorithms, the Markov chain random walker generates new model sample $\boldsymbol{m}_{i+1}$ depending only on a previous step $\boldsymbol{m}_i$. The new model sample is either accepted or rejected given the Metropolis-Hastings rule: if the new model has higher PDF value than the previous one ($p(\boldsymbol{m}_{i+1}) \geq p(\boldsymbol{m}_i)$), the new model is accepted and the next step of the chain starts from it. If the new model has lower PDF value, it is accepted with probability given by $\frac{p(\boldsymbol{m}_{i+1})}{p(\boldsymbol{m}_i)}$. If the model is not accepted, the chain is restarted again from the ith step. The Metropolis-Hastings rule is considered the most efficient rule for accepting the maximum of the proposed models. Still it has been shown that the sampling algorithm converges to $p(\boldsymbol{m})$ with infinite number of iteration steps (Tierney (1994)).

The success of the sampling algorithm depends on the acceptance rate of the proposed models. This can be affected mainly by the stepsize of the random walker. The smaller is the stepsize for the proposed model, the higher is the acceptance rate (since $p(\boldsymbol{m}_{i+1})$ is close to $p(\boldsymbol{m}_i)$). However, the walker then

moves in the model parameter space slowly and a large number of steps are performed, consuming large amount of computational time and being thus less efficient. When increasing the size of the perturbation, the sampler converges faster towards the PDF maxima. However, for a multimodal PDF the sampler may be unable to escape the local maxima, where the acceptance rate may be really low.

To avoid the sampler to be entrapped in local maxima, the PDF in the original Metropolis-Hastings algorithm is modified by adding a new parameter, usually called temperature T (the algorithm was motivated by the crystallization process during cooling which minimizes the energy of the crystal). The modified posterior PDF is assumed in form

$$p(\boldsymbol{m}, T) = p(\boldsymbol{m})^{\frac{1}{T}} \approx \exp\left(\frac{-S(\boldsymbol{m})}{T}\right). \tag{3.7}$$

The probability of accepting the new model in the tempered chain is

$$p(\boldsymbol{m}_{i\to i+1}, T) = \min\left[1, \frac{p(\boldsymbol{m}_{i+1}, T)}{p(\boldsymbol{m}_i, T)}\right]. \tag{3.8}$$

The effect of the temperature parameter T on the PDF is displayed in Fig. 3.1, where multimodal misfit $S(\boldsymbol{m})$ with two local minima (one of them also global) is assumed (shown by black curve). The normalized PDF corresponding to three different temperature levels $T = 1, 5, 100$ is displayed with different color curves. The PDF corresponding to the original temperature $T = 1$ (brown curve in Fig. 3.1), shows one distinct global maximum and one local maximum several times smaller. In case the Markov chain sampler reaches the global maximum, it might be difficult to explore other local maxima, as the PDF between them is very low. With increasing temperature to $T = 5$, the difference between the maxima is less pronounced and when assuming the highest temperature $T = 100$, the PDF is almost flat. The sampler would accept most of the proposed models and sample the model parameter space almost uniformly.

This convenient property of the tempered chain has been used in a well-known simulated annealing algorithm (Kirkpatrick et al. (1983)), which uses tempered Markov chains to sample the PDF starting from high temperatures and gradually decreasing the temperature to focalize in areas with high value of the original PDF.

In the PT algorithm, several chains, each with generally different temperature, sample the model space simultaneously. The chains at high temperatures sample the model parameter space more globally, whereas lower temperature chains concentrate more on areas with higher value of the PDF. To obtain the sampling of the original PDF, chains with $T = 1$ are employed. The efficiency of the algorithm is provided by temperature swaps between two Markov chains in-between the chainsteps. The probability $\alpha(i, j)$ of the temperature swap between chains $[\boldsymbol{m}_i, T_i]$ and $[\boldsymbol{m}_j, T_j]$ (leading to $[\boldsymbol{m}_i, T_j]$ and $[\boldsymbol{m}_j, T_i]$ if accepted) follows

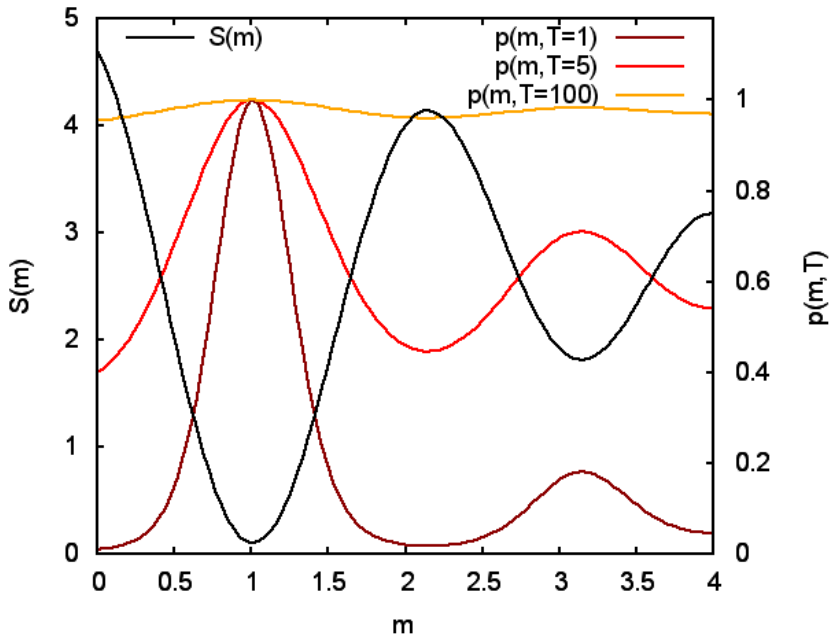

Figure 3.1: Effect of temperature for tempered Markov chain: multimodal misfit (black curve, left axis), PDF of the corresponding misfit modified by different $T = 1, 5, 100$ (color curves, right axis). Note that the PDFs are not properly normalized.

the Metropolis-Hastings rule:

$$\begin{aligned} \alpha(i,j) &= \min\left[1, \frac{p(\boldsymbol{m}_i, T_j)p(\boldsymbol{m}_j, T_i)}{p(\boldsymbol{m}_i, T_i)p(\boldsymbol{m}_j, T_j)}\right] \\ &= \min\left[1, \left(\frac{p(\boldsymbol{m}_i)}{p(\boldsymbol{m}_j)}\right)^{\frac{1}{T_j}} \left(\frac{p(\boldsymbol{m}_j)}{p(\boldsymbol{m}_i)}\right)^{\frac{1}{T_i}}\right] \\ &= \min\left[1, \left(\frac{p(\boldsymbol{m}_i)}{p(\boldsymbol{m}_j)}\right)^{\frac{1}{T_j}-\frac{1}{T_i}}\right]. \end{aligned} \tag{3.9}$$

The PT algorithm shows good convergence towards the posterior PDF and due to the mixing of the tempered chains, it is less likely to stuck in a local PDF maxima (see also Sambridge (2014)). Another great advantage is avoiding evaluation of the PDF normalizing constant C in (3.4).

Markov chain MC random walker usually generates new models by perturbing the current model. This may present several problems. Firstly, in the Markov chain MC sampler it is important from the theoretical point of view that the samples are independent on each other. However, for thus constructed sampler it is not necessarily fulfilled. For this purpose, several chainsteps in random walk are performed without being saved. Another problem occurs in case of no or poor information on the starting model. Therefore, the so-called burn-in phase is introduced: the starting model is generated (randomly or fixed) from which the random walk is launched. However, during the burn-in phase the walker transfers to the parts of the model space containing permissible models, but the visited samples are not saved for further processing.

3.1.3 Surface wave inversion

For the Bayesian inversion, we need to define model vector $\boldsymbol{m}$ and data vector $\boldsymbol{d}$, theoretical relation between the data and the model, $\boldsymbol{d} = \boldsymbol{g}(\boldsymbol{m})$, and their statistical distributions in form of PDFs. In the inversion based on the surface wave measurements, the dispersion curves obtained at a point on the Earth's surface serve as input data to estimate the vertical velocity structure beneath the given point.

The dispersion curves in our application were extracted from the ambient seismic noise cross-correlation (see Růžek et al. (2016)). We employ the phase velocity dispersion curves for all three components: transversal Love (T), and two Rayleigh – vertical (Z) and radial (R), i.e. $\boldsymbol{d} = \left[c^{\mathrm{T}}(\boldsymbol{x}, T), c^{\mathrm{R}}(\boldsymbol{x}, T) c^{\mathrm{Z}}(\boldsymbol{x}, T)\right]^T$. The input data are measured at N points on the Earth's surface (i.e., data points, see Fig. 3.2) and discretized in periods ranging between 4 and 20s. Here, only phase velocity dispersion curves are utilized although the group velocity data may be also assumed. The data vector in our inverse problem is thus given as

$$\boldsymbol{d} = \left[c_1^{\mathrm{T}}(4s), ..., c_1^{\mathrm{T}}(16s), c_1^{\mathrm{T}}(20s), c_1^{\mathrm{R}}(4s), ..., c_1^{\mathrm{R}}(20s), ...c_N^{\mathrm{Z}}(20s)\right]^T . \tag{3.10}$$

We assume the data measurement PDF, $\rho(\boldsymbol{d})$, in form of a Gaussian with diagonal covariance matrix with standard error σ_{d} estimated from the Rayleigh data differences between the two components (see further).

The model vector is composed of the S-wave velocity and alternatively also v_p/v_s ratio: $\boldsymbol{m} = [\beta(\boldsymbol{x}), (v_p/v_s)(\boldsymbol{x})]^T$. To calculate the dispersion curve, we employ the VDISP code which is based on the matrix method (Novotný (1999)). The matrix method assumes the seismic wave propagation in homogeneous vertical layers. Therefore, we parametrize our model space by a set of 1D layered S-wave velocity models (and possibly v_p/v_s assumed here as depth-independent) on a regular horizontal grid – model control points (see Fig. 3.2). The number of layers and their thicknesses are considered fixed during the inversion. The interfaces are located at 7 depths: $(2, 4, 8, 12, 18, 24, 32)$ km. The model vector is thus in form

$$\boldsymbol{m} = \left[\beta_{11}, \beta_{12}, ..., \beta_{17}, \beta_{18}, (v_p/v_s)_1, \beta_{21}, ..., \beta_{M8}, (v_p/v_s)_M\right]^T , \tag{3.11}$$

where M is the number of model control points. We set the distance between the model points larger than the distance of the data points (so that $N > M$). Since the model control points and data points are not collocated, the cubic spline interpolation of the model parameters from the model control points into the data points is performed when evaluating the misfit.

3.1.4 Matrix method for dispersion curve calculation in 1D layered media

As already mentioned, the forward problem si solved using a matrix method for the dispersion curve calculation. Here, the method is briefly reviewed assuming Love waves, following the derivation of Novotný (1999).

Assume 1D layered media with homogeneous layers over a homogeneous halfspace with interfaces located at depths $(z_2, z_3, ..., z_{n-1}, z_n)$. Each layer has thickness $d_i = z_{i+1} - z_i$ and material parameters P-wave velicity α_i, S wave velocity β_i and density ρ_i. The halfspace with $\alpha_n, \beta_n, \rho_n$ is located below depth z_n.

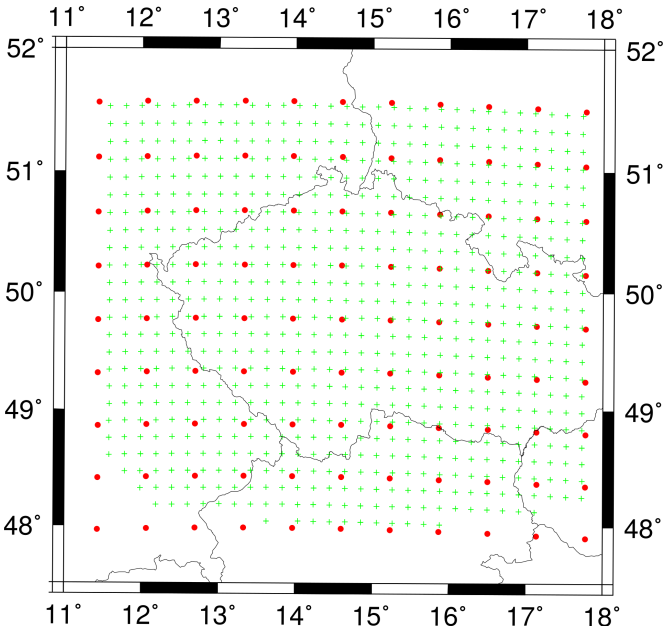

Figure 3.2: Horizontal distribution of data points (green pluses) and model control points (red dots) for Bayesian inversion of dispersion maps into 3D S-wave velocity model.

Let us assume Love waves displacement in form $\boldsymbol{u} = (0, v, 0)$ and the only nonzero component of the stress tensor $\tau_{yz} = \tau = \mu\frac{\partial v}{\partial z}$. The boundary conditions for each of the layer interfaces is the continuity of the displacement and stress. Free surface is assumed at z_1. In halfspace we require $v(x, z, t) \to 0$ for $z \to \infty$. In each layer the displacement fulfills the equation (Aki and Richards (2002))

$$\beta_i^2 \Delta v_i(x, z, t) = \frac{\partial^2 v_i}{\partial t^2}. \tag{3.12}$$

where Δ is the Laplace operator.

The ansatz to the displacement is assumed in form of harmonic plane wave propagating along x:

$$v_i(x, z, t) = f_i(z) e^{i\omega\left(t - \frac{x}{c}\right)}, \tag{3.13}$$

where $f_i(z)$ is the amplitude of the wave, ω is the angular frequency and c is the wavespeed.

Inserting ansatz (3.13) into (3.12) gives

$$\frac{\mathrm{d}^2 f_i(z)}{\mathrm{d}z^2} + \left(\frac{\omega^2}{\beta_i^2} - k^2\right) f_i(z) = 0, \tag{3.14}$$

where $k = \frac{\omega}{c}$ is the wave number. The solution for $f_i(z)$ to (3.14) is

$$f_i(z) = a_i \sin(s_i(z - z_i)) + b_i \cos(s_i(z - z_i)), \tag{3.15}$$

where we denote $s_i = \sqrt{\left(\frac{\omega}{\beta_i}\right)^2 - k^2}$.

The displacement and stress then reads

$$v_i(x,z,t) = [a_i \sin(s_i(z-z_i)) + b_i \cos(s_i(z-z_i))]\, e^{i\omega\left(t-\frac{x}{c}\right)} \tag{3.16}$$

$$\tau_i(x,z,t) = \mu_i s_i \,[a_i \cos(s_i(z-z_i)) - b_i \sin(s_i(z-z_i))]\, e^{i\omega\left(t-\frac{x}{c}\right)}. \tag{3.17}$$

Now, we compare the solution on top of each layer with the solution on the bottom. For $z = z_i$

$$v_i(x,z_i,t) = b_i e^{i\omega\left(t-\frac{x}{c}\right)} \tag{3.18}$$

$$\tau_i(x,z_i,t) = \mu_i s_i a_i e^{i\omega\left(t-\frac{x}{c}\right)} \tag{3.19}$$

and for $z = z_{i+1}$

$$v_i(x,z_{i+1},t) = [a_i \sin(s_i(d_i)) + b_i \cos(s_i(d_i))]\, e^{i\omega\left(t-\frac{x}{c}\right)} \tag{3.20}$$

$$\tau_i(x,z_{i+1},t) = \mu_i s_i \,[a_i \cos(s_i(d_i)) - b_i \sin(s_i(d_i))]\, e^{i\omega\left(t-\frac{x}{c}\right)}. \tag{3.21}$$

Thus one can express the displacement and stress on the bottom of the layer using the displacement and stress on top of the layer:

$$v_i(x,z_{i+1},t) = \frac{\sin(s_i d_i)}{\mu_i s_i}\tau_i(x,z_i,t) + \cos(s_i d_i) v_i(x,z_i,t) \tag{3.22}$$

$$\tau_i(x,z_{i+1},t) = \cos(s_i d_i)\tau_i(x,z_i,t) - \mu_i s_i \sin(s_i d_i) v_i(x,z_i,t), \tag{3.23}$$

rewritten using the matrix notation

$$\begin{pmatrix} v_i(x,z_{i+1},t) \\ \tau_i(x,z_{i+1},t) \end{pmatrix} = A_i \begin{pmatrix} v_i(x,z_i,t) \\ \tau_i(x,z_i,t) \end{pmatrix} \tag{3.24}$$

with A_i

$$A_i = \begin{pmatrix} \cos(s_i d_i) & \frac{\sin(s_i d_i)}{\mu_i s_i} \\ -\mu_i s_i \sin(s_i d_i) & \cos(s_i d_i) \end{pmatrix}. \tag{3.25}$$

Consider now the continuation of the solution between two layers on their common interface

$$\begin{pmatrix} v_i(x,z_i,t) \\ \tau_i(x,z_i,t) \end{pmatrix} = \begin{pmatrix} v_{i-1}(x,z_i,t) \\ \tau_{i-1}(x,z_i,t) \end{pmatrix}. \tag{3.26}$$

Starting from the bottommost interface and alternately using continuation condition (3.26) with matrix formula (3.24) one obtains

$$\begin{aligned} \begin{pmatrix} v_n(x,z_n,t) \\ \tau_n(x,z_n,t) \end{pmatrix} &= \begin{pmatrix} v_{n-1}(x,z_n,t) \\ \tau_{n-1}(x,z_n,t) \end{pmatrix} = A_{n-1} \begin{pmatrix} v_{n-1}(x,z_{n-1},t) \\ \tau_{n-1}(x,z_{n-1},t) \end{pmatrix} \\ &= \cdots = A_{n-1}\cdots A_2 A_1 \begin{pmatrix} v_1(x,z_1,t) \\ \tau_1(x,z_1,t) \end{pmatrix}, \end{aligned} \tag{3.27}$$

or denoting $A = A_{n-1}A_{n-2}\cdots A_2 A_1$

$$\begin{pmatrix} v_n(x,z_n,t) \\ \tau_n(x,z_n,t) \end{pmatrix} = A \begin{pmatrix} v_1(x,z_1,t) \\ \tau_1(x,z_1,t) \end{pmatrix}. \tag{3.28}$$

This gives the relation between the solution on the surface and in the halfspace.

Let us now consider the solution in the halfspace in form

$$v_n(x,z,t) = \left(e_n e^{-is_n(z-z_n)} + f_n e^{is_n(z-z_n)}\right) e^{i\omega\left(t-\frac{x}{c}\right)}, \tag{3.29}$$

where $s_n = \sqrt{\left(\frac{\omega}{\beta_n}\right)^2 - k^2}$. If $c > \beta_n$, then s_n is real and the solution in the halfspace oscilates. However, the physically reasonable solution requires amplitude decrease with increasing z. Therefore, we need $c < \beta_n$, then s_n is imaginary. Denote $s_n^* = is_n$, then

$$v_n(x,z,t) = \left(e_n e^{-s_n^*(z-z_n)} + f_n e^{s_n^*(z-z_n)}\right) e^{i\omega\left(t-\frac{x}{c}\right)} \tag{3.30}$$

and when requiring $v_n \to 0$ with $z \to \infty$, we must set $f_n = 0$.

For $z = z_n$, we thus have

$$v_n(x,z_n,t) = e_n e^{i\omega\left(t-\frac{x}{c}\right)} \tag{3.31}$$

$$\tau_n = -e_n s_n^* \mu_n e^{i\omega\left(t-\frac{x}{c}\right)} \tag{3.32}$$

Moreover, assuming free surface condition $\tau_1(x,z_1,t) = 0$, (3.28) simplifies to

$$\begin{pmatrix} e_n \\ -e_n s_n^* \mu_n \end{pmatrix} = A \begin{pmatrix} v_1(x,z_1,t) \\ 0 \end{pmatrix}, \tag{3.33}$$

which can be rewritten as

$$\begin{pmatrix} 1 & -A_{11} \\ -s_n^*\mu_n & -A_{21} \end{pmatrix} \begin{pmatrix} e_n \\ v_1(x,z_1,t) \end{pmatrix} = 0. \tag{3.34}$$

This system of equations has non-trivial solution when its determinant is zero which gives the dispersion equation of Love surface waves for $c(\omega)$

$$-A_{21} - \mu_n A_{11} \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{\omega}{\beta_n}\right)^2} = 0. \tag{3.35}$$

The equation has infinite solutions for $c(\omega)$, each representing different mode of Love wave.

Alternative notation for vector $(\frac{\dot{v}_i(x,z,t)}{c}, \tau_i(x,z,t))^T$ leads to the solution using so-called Thomson-Haskell matrices:

$$\begin{pmatrix} \frac{\dot{v}_i(x,z_{i+1},t)}{c} \\ \tau_i(x,z_{i+1},t) \end{pmatrix} = \tilde{A}_i \begin{pmatrix} \frac{\dot{v}_i(x,z_i,t)}{c} \\ \tau_i(x,z_i,t) \end{pmatrix} \tag{3.36}$$

with matrix

$$\tilde{A}_i = \begin{pmatrix} \cos(s_i d_i) & \frac{ik\sin(s_i d_i)}{\mu_i s_i} \\ \frac{i\mu_i s_i \sin(s_i d_i)}{k} & \cos(s_i d_i) \end{pmatrix} \tag{3.37}$$

and dispersion curve equation

$$\tilde{A}_{21} - i\mu_n \tilde{A}_{11} \sqrt{1 - \left(\frac{c}{\beta_n}\right)^2} = 0. \tag{3.38}$$

Similarly, one can derive the dispersion equation for Rayleigh waves.

The matrix method has been implemented in code VDISP by prof. Novotný, which we employ for the calculation of the forward problem.

3.2 Bayesian 1D synthetic tests

Here, the inverse problem (inversion of surface wave dispersion curves into 1D layered media) is solved in the Bayesian framework employing the PT method. The forward problem uses the matrix method for the calculation of dispersion curves.

We have performed several types of 1D synthetic tests to assess the capabilities (e.g., reliability, sensitivity, resolution) of the applied method (see Section 3.1).

The 1D reference (target) model for the tests is composed of 7 layers and half-space with fixed interfaces, although some of the layers share the same parameters (see Table 3.1).

Layer #	Depth [km]	S-wave velocity [km/s]
1	0-2	3.4
2	2-4	3.4
3	4-8	3.4
4	8-12	3.6
5	12-18	3.6
6	18-24	3.79
7	24-32	4.03
halfspace	> 32	4.13
v_p/v_s	> 0	1.5735

Table 3.1: Reference model for synthetic tests.

The synthetic phase dispersion curves are calculated in the reference model for periods corresponding to the real data inversion ($(4, 6, 8, 10, 12, 16, 20)s$) for all components (T, R, Z). The synthetic data were kept noise-less or were modified by Gaussian white noise with standard devation as assumed for the real dataset (0.14km/s, see Chapter 4).

The model parameters are S-wave velocity values for the same layers as the reference model and depth-independent v_p/v_s ratio. The number of layers and their thicknesses are kept fixed. As a prior information we assume homogenous PDF within fixed bounds. The data covariance matrix is assumed diagonal with variance given by real data ($\sigma_{\mathrm{d}} = 0.14$ km/s).

3.2.1 Single parameter tests

In this set of synthetic tests, we have run the MC inversion using dispersion curves at one data point (N=1) collocated with model control point (M=1), but only for each (single) model parameter – S-wave velocity model in each layer individually or eventually v_p/v_s ratio, i.e. model vector consists of only one value. All other model parameters were kept at the reference value. This basic test examines our data sensitivity to each model parameter independently. The tests were carried out using separate Love- or separate Rayleigh-wave dispersion curve and also using both. The synthetic data were not modified by random noise. We should keep in mind that even for perfect data the variance of the PDF is nonzero, as it is defined using, among other, PDF of data measurement for which nonzero standard deviation is assigned.

We run the inversion on 2 (in one case 4) CPUs, each with 12 Markov chains for 1k steps in burn-in phase and 10k steps in post-burn phase. The resulting number of models sampling the posterior PDF for each layer is ≈ 30k.

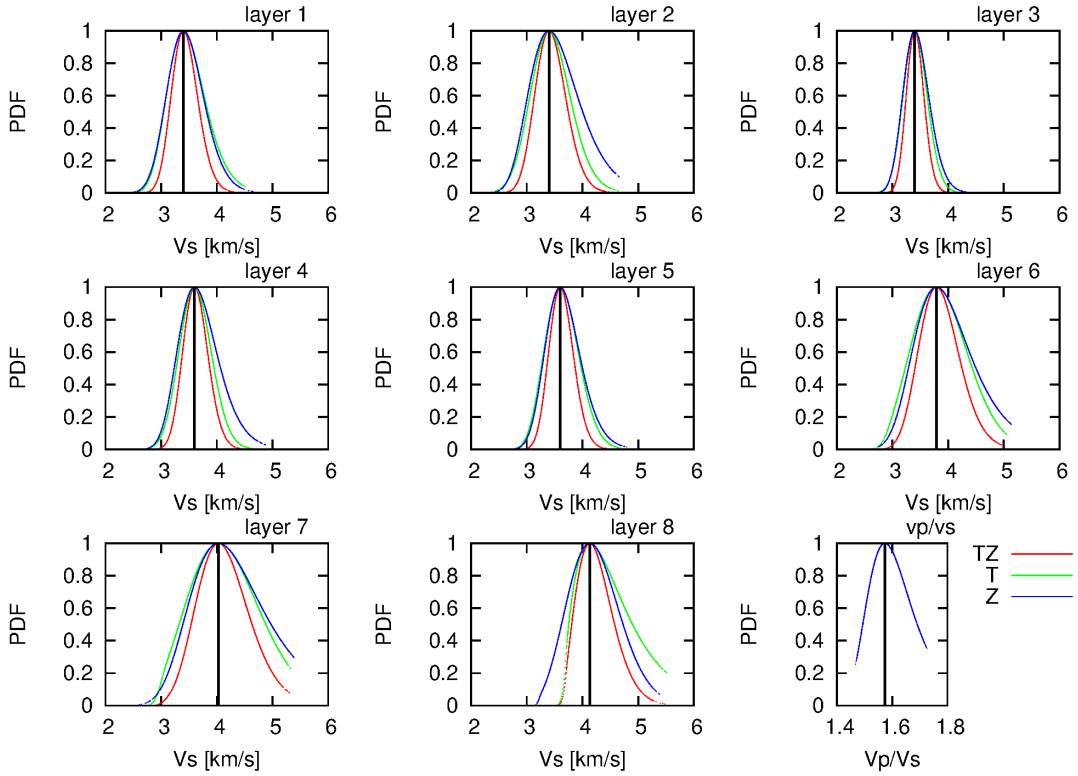

Figure 3.3: Normalized posterior PDF results of synthetic single parameter inversion for Love(T), Rayleigh(Z) or both (TZ) phase dispersion data. The reference value is shown by black line.

In Fig. 3.3 we present the results in form of PDF for each model parameter when Rayleigh, Love or both phase velocity dispersion curves are inverted. We have found no significant difference regarding the S-wave velocity model in any layer when inverting Love or Rayleigh wave data separately. The v_p/v_s ratio is determinable only when Rayleigh data are employed. Nevertheless, the resulting PDF is very broad and thus this parameter is very poorly resolved. The improvement in resolution of S-wave velocity parameters emerges when both Rayleigh and Love phase dispersion curves are employed.

We have also estimated the mean value and its standard error for each case. For all results, the mean value tends to overestimate the original value by ≈ 0.1 km/s, but lies within the margin of the estimated error. This is caused by the assymetry in the PDF. The location of the PDF maximum corresponds to the reference values.

From these tests, we may also draw some conclusions on the sensitivity of selected model parameters for the employed dataset (i.e. phase velocity dispersion curves in period range 4-20 s). The best resolution is obtained in S-wave velocity parameter in the third layer (depths 4-8 km). Then the variance of the PDF increases (i.e. resolution decreases) to similar value for the top layers and layers 4 and 5 (0-4 km, 8-18 km). Then the resolution deteriorates with increasing depth except for the halfspace.

It may seem surprising that the variance in the halfspace is better compared to the layer above. However, the bottommost parameter is not better resolved by the employed dataset than the parameters of the deep layers above. We carried out a set of synthetic tests, where we modified thickness of an additional artificial layer above the halfspace (see Fig. 3.4). There is anticorrelation between the variance of the model parameter and the corresponding layer thickness up to critical thickness $\approx$ 50 km. After this, the variance remained almost constant with increasing layer thickness, close to the value corresponding to that of purely halfspace.

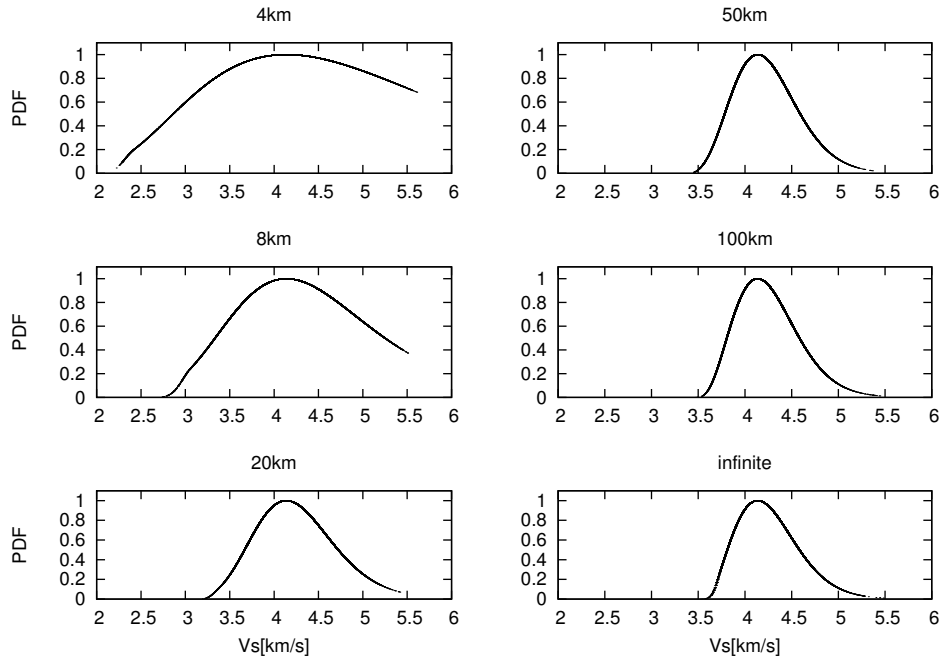

Figure 3.4: Normalized posterior PDF of 1-parameter inversion for velocity in the deepest additional layer or in halfspace. The model contains additional layer situated below 32km with varying thickness.

This test demonstrates the effects of the selected parametrization. When the model space is parameterized using layers with fixed thicknesses, the inversion estimates the average value of the parameter over the whole layer. The thicker is the layer, the greater is the impact of the model parameter on the data, resulting in the decrease of that model parameter variance. In addition, the model parameter is affected by the sensitivity of the dataset. In this case, the layer thickness recognizable by our dataset corresponds to the 50km.

The anticorrelation between the layer thickness and S-wave velocity variance in the particular layer may be viewed also as warning for models parametrized by a large number of thin layers, especially at greater depths.

3.2.2 1D inversion

The inversion is performed for one data point (N=1) synthetic dispersion curves (Love or Rayleigh or both) into a 1D layered model with fixed layers at the same point (M=1). The data are not modified by noise. In this test, the regularization by means of nonnegative velocity gradient with depth is applied: every new model

generated by perturbing the current model in the Markov chain is controlled whether it contains a layer with decreased S-wave velocity compared to the layer above. Until such layer is not present, new model is repeatedly generated from the current one.

The PT algorithm runs using 12 chains, for 1k burn-in and 10k post-burn steps. The total number of posterior PDF samples is $\approx$ 28k.

The results are shown in Fig. 3.5. The best and the mean model deviate more from the reference model than in the previous case, but the results lie within the margin of error. It may be observed that the best model reproduces the reference model better. The variance increases naturally with depth and the decrease in the halfspace is not present.

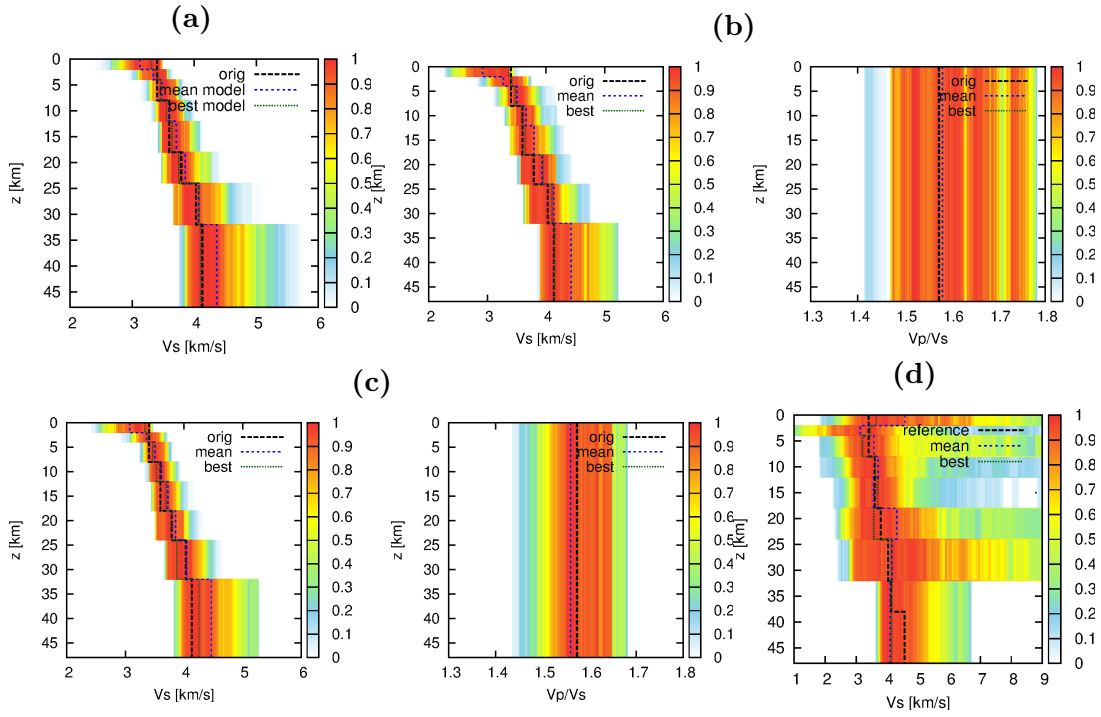

Figure 3.5: Resulting normalized posterior PDF of synthetic 1D inversion for one data point (color coded). The mean, the best and the reference model are also shown (see legend). a) Love wave dispersion curve inversion to S-wave velocity, b) Rayleigh wave dispersion curve inversion to S-wave velocity and v_p/v_s, c) both Love and Rayleigh wave dispersion curves inversion to S-wave velocity and v_p/v_s, d) Love and Rayleigh wave dispersion curve inversion to S-wave velocity without regularization constraint of non-negative velocity gradient.

The results of inversion of separate Love and Rayleigh components into S-wave velocity model are very similar as shown in Fig. 3.5a and 3.5b. When inverting Rayleigh component for both S-wave velocity and v_p/v_s, the S-wave velocity results are unchanged, but v_p/v_s has no clear single maximum of PDF and resolution of this parameter is very poor (Fig. 3.5b). This shows that the data-model relation for parameter v_p/v_s is much more complex than for S-wave velocity. The improvement in the resolution of the v_p/v_s arises when both Love and Rayleigh components are inverted at the same time (Fig. 3.5c). Although the variance in this parameter is still high, both the mean and the maximum of PDF are close to the reference value.

Another feature revealed by these tests is the trade-off between S-wave velocity parameters for different layers. For example, in layers 1 and 4, the posterior PDF is skewed towards lower values than the reference value. This is compensated in the next layers (in particular 3 and 5), where the model samples are drawn towards higher values than the reference ones. The calculated mean model therefore leads to the lower/higher values for the S-wave velocities in the corresponding layers. This observed assymetry in the PDF is a consequence of combined effect of the applied regularization (non-negative velocity gradient) and nonlinearity of the forward problem.

When we allow the model parameters to decrease with depth, the trade-off between the parameters in nearby layers intensifies and leads to the oscillatory solution. The variance in each model parameter becomes very high and neither the best nor the mean model is reasonable (Fig. 3.5d, mind the different v_s range). This example shows that our regularization by nonnegative velocity gradient adequately reduces the model parameter space and restricts the solution towards a sound result.

3.2.3 Inversion of dispersion maps into 1D model

In this case, we use the synthetic dispersion curves generated in the 1D reference model and "distribute" them into the data points of the real inversion ($N = N_{\mathrm{dat}} = 775$, see Fig. 3.2) – we obtain the dataset similar to the inversion of the real 3D all-component data. The data are kept noise-less or are modified by the white noise corresponding to the noise estimated from the real data. The inversion is performed using these distributed data into a 1D model (i.e. M=1).

The PT inversion is run for 10k steps in burn-in phase and 200k steps in output phase on 48 Markov chains. The resulting number of accepted samples exceeds 500k.

The results for both perfect (Fig. 3.6a) and noisy (Fig. 3.6b) data are very similar, the best model as well as the mean model lies close to the reference model within the standard error. Furhermore, the variances in the model parameters are also almost identical for both noise-free and noisy data. Since the variance of the model parameters for the noisy data inversion is similar to that of the perfect data inversion, it suggests that the data noise almost cancels out when high number of data (e.g., number of real data) was employed.

The general increase in the S-wave velocity variance with increasing depth can be traced for both synthetic tests (and real-data inversion as well, see Chapter 4). This might be interpreted as decrease of data sensitivity with increasing depth. Contrarily, PDF in the halfspace is remarkably narrow. We have performed several synthetic tests and found out, that when increasing the number of data, the model variance is reduced, but not evenly for all depths. The greatest impact of data addition is on the S-wave velocity in the halfspace and the v_p/v_s ratio. As already mentioned, for these two parameters the inversion is estimating the average value over their whole (infinite) depths. This average value is in general better determined.

We would like to note that this test showed the dependence of the posterior PDF on the number of data points. If the data (i.e. data errors) were not correlated, the resulting posterior covariance would be estimated properly. However,

in our case the data are correlated as a result of the applied Gaussian smoothing in the adjoint localization, mainly for the long periods. Theoretically, this should be accounted for in the data covariance matrix in the misfit calculation, otherwise the posterior model uncertainty is underestimated. To at least partially correct for this underestimation, we have increased the data variance by a factor of 9, which was estimated as follows: assuming 10 s surface wave data, the width of the Gaussian smoothing function is aproximately 50 km covering 3×3 data points in our inversion of dispersion maps. For simplicity, this factor was used for all periods.

The results of these last synthetic tests may be compared with the results from the real data 1D inversion shown in Fig. 4.4c (see also Section 4.5.2). In the real data case, greater variance in the S-wave velocity parameters is observed, predominantly in the deeper layers and in the subsurface parts. The tests presented above have shown that this is not to be attributed to the data noise. Therefore, we explain the increased variance as a consequence of the variability in the real data due to the true 3D velocity structure.

3.3 Notes on implementation of MC(PT) algorithm

3.3.1 MC setting for real data application

The performance of the MC (PT) sampling algorithm depends on several parameter settings, which may have impact on the efficiency of the resulting sampling. In this part, we would like to specify setting for the real data inversion.

As already mentioned in Section 3.1, one of the important parameters is the stepsize used to generate the new/proposed model from the current one. If the stepsize is small enough, the new model is accepted and the Markov chain samples the domain successfully but in too many steps. If the stepsize is chosen too large, the sampler is prone to stuck in the local maxima of the PDF. Since the PT algorithm enables the walker to escape the local maxima, we may be able to use larger stepsize for the model perturbation. We use Gaussian perturbation in each model parameter with deviation 0.01 km/s and 0.001 for the S-wave velocity and v_p/v_s, respectively. In the later part (i.e., after the sampler does not accept the new models indicating convergence to a local minimum of the misfit) the size of this Gaussian is reduced to 0.002 km/s and 0.0001. To ensure that the samples are indendepent, we have chosen to save every 100th chainstep. Only after the chains converge for the smaller perturbations, we stop the random walk. In the real-data inversion, this leads to more than 100k steps.

The distribution of the temperature parameter T affects the ability of the Markov chains to leave the local extrema of the PDF. We have decided to use randomly distributed temperature values between 1 and 50. It was not necessary to assume higher values for temperature, as the sampled PDF does not generally show such pronounced local extrema. Furthermore, to ensure the sufficient sampling of the original PDF, several chains (in particular every 5th) are forced to have $T = 1$.

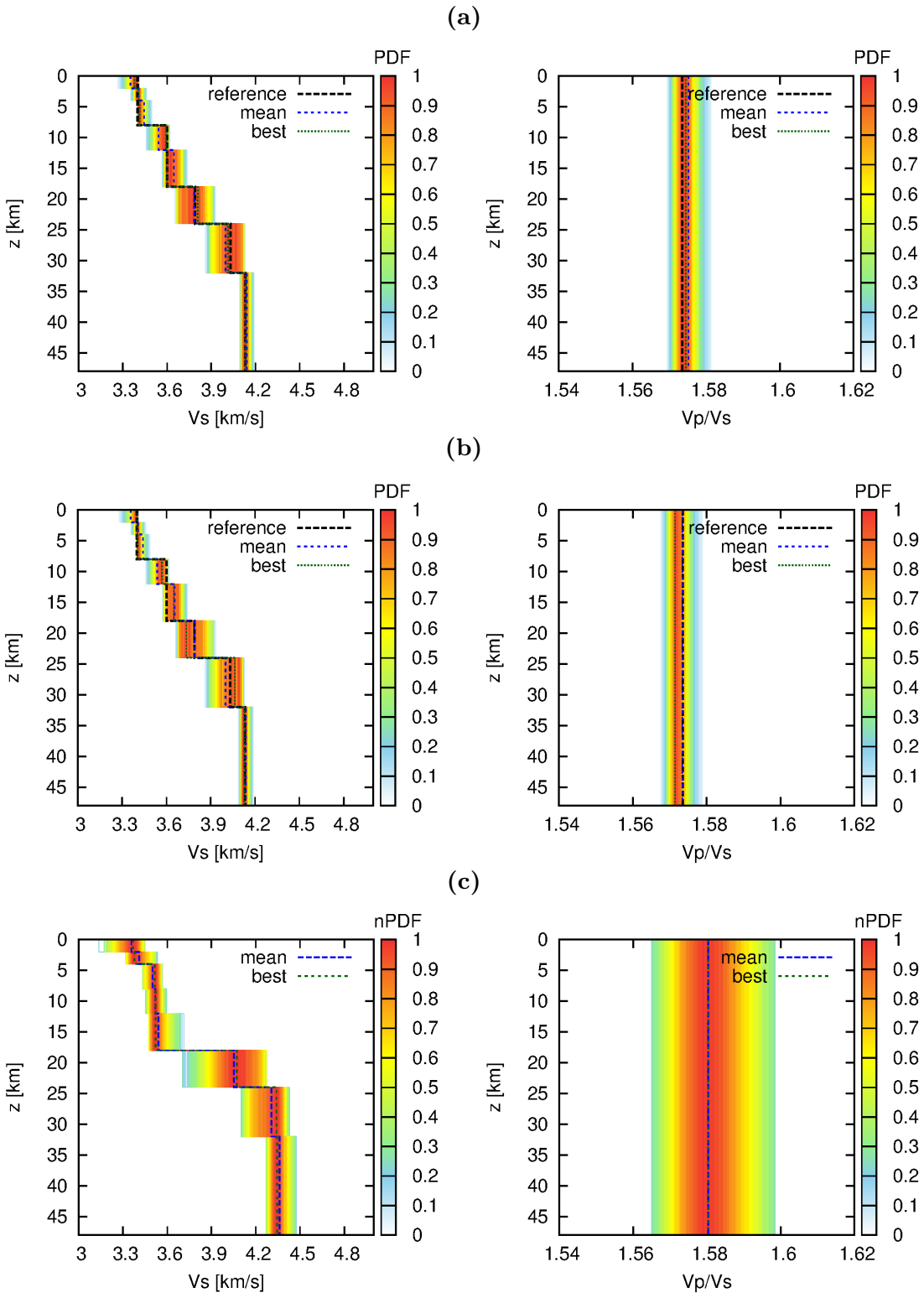

Figure 3.6: Posterior PDF of the synthetic inversion of dispersion maps into a 1D model. PDF is normalized to the best model shown. The mean and reference models are also shown. Inversion of: a) error-free data, b) data modified by noise, c) real dispersion maps using the same model parameters as synthetic tests.

Since we have poor prior information on the model parameters (in particular constant on wide interval), in the real-data application we use burn-in phase consisting of 1000 chainsteps. This was found to be sufficient mainly thanks to the larger stepsize.

In the next section, the effect of several parameters (e.g. number of chains distributed between the CPUs and number of chainsteps) on the performance of the PT sampling algorithm is examined with the help of 1D synthetic tests. Therefore, here we only specify that the overall number of Markov chains in the real application is > 500.

3.3.2 Effect of some of the PT parameters

Here we use 1D synthetic tests to examine the performance of the PT algorithm for different technical parameters, in particular number of CPUs (nCPU), number of chains and chainsteps. The synthetic data as well as model parameters for the 1D inversion are the same as in Section 3.2.2. We assess the efficiency of the sampling algorithm by evaluating the number of saved model samples as well as the best found misfit but we should keep in mind the randomness in the sampling process. Furthermore, we also investigate the computatational time which then plays important part for 3D problem. The results are shown in Table 3.2. Also keep in mind that the efficiency of the sampling algorithm depends also on the other setting parameters. In these tests they were kept fixed.

According to the minimal obtained misfit (see the last column), all results show very good convergence towards the maximum PDF except for one – the MC inversion without the temperature swaps between the chains (last line). When comparing the performance of the PT algorithm, we rather concentrate on the number of successful samples (i.e., samples accepted every 100th chain step with temperature $T = 1$ after the burn-in phase, see also Section 3.1.2). When increasing the number of chain steps, the resulting number of samples scales accordingly. The same holds when increasing the number of CPUs, but keeping the number of chains per CPU. However, this is not valid when scaling the number of chains: increasing the number of chains twice while keeping the nCPU does not generate twice more samples. It appears that the PT sampling algorithm performs better for lower number of chains per CPU. To compensate for this effect, either the higher number of CPUs or the chain steps should be performed. The calculation time for PT with temperature swaps differs only slightly from the one where the swaps are not allowed showing the efficiency of the PT algorithm. The calculation time scales with number of chains and also number of chainsteps (assuming the small difference to be caused by the chainsteps in the burn-in phase). When employing more CPUs, the calculation time scales linearly if the number of chains per CPU is kept.

To conclude, the efficiency of the sampling is the same when increasing the number of chains (when sufficient number of CPUs is employed) as when performing larger number of chainsteps. This shows good scalability of the PT algorithm making it suitable for parallel implementation.

nCPU	# chains	# steps	# samples	comp. time	min. misfit
2	24	5000	28.5k	3m6s	3.86e-03
2	24	10000	57.3k	5m43s	3.86e-03
2	12	5000	18.9k	1m33s	3.76e-03
2	12	10000	38.3k	3m46s	1.72e-03
4	48	5000	55.8k	5m59s	7.36e-04
4	24	10000	76.7k	4m32s	4.81e-03
2	24	5000	28.3k	3m	1.47e-02

Table 3.2: Performance of the PT algorithm for simple 1D synthetic inversion. The number of CPUs (nCPU) and PT parameters: number of chains and chain-steps were varied. The effect on computational time, number of saved samples and best misfit is examined. The last line shows the results when the temperature swaps between the chains are not allowed.

3.4 Conclusion

In this chapter, we have presented the inversion of surface wave dispersion curves into a velocity model in the Bayesian framework. Contrarily to the traditional approach when the solution is represented by a model minimizing misfit (i.e., maximizing posterior PDF) and its covariance matrix, our Bayesian solution is represented by a large number of model samples. The selection of the eligible models employs Monte Carlo method, which draws model samples according to the posterior PDF. This method is also suitable for problems with rather complex data-model relationship which have complex or multimodal PDFs.

In our application, the parallel tempering method is employed and tested on 1D synthetic inversion of the dispersion curves. The tests show that the mean and best model are close to the reference model. Moreover, the variance of the model parameters is easily examined. The algorithm shows good performance in parallel implementation which makes it suitable for large-scale problems. Thus, in the next chapter this method is applied to invert the dispersion maps (i.e., regularly distributed dispersion curves) into the 3D S-wave velocity model of the Bohemian Massif (Chapter 4).

Chapter 4

Three-dimensional S-wave velocity model of the Bohemian Massif from Bayesian ambient noise tomography

Published in Tectonophysics,
Volume 717, 16 October 2017, 484–498, doi:10.1016/j.tecto.2017.08.033

Lubica Valentová[1],František Gallovič[1],Petra Maierová[2]

Highlights

- First ambient noise surface-wave tomography of the Bohemian Massif was performed.
- Dispersion curves were inverted into phase velocity maps by adjoint inversion.
- Phase velocity maps were inverted to 3D S-wave velocity model by Bayesian approach.
- Mean S-wave velocity model correlates well with main geologic structures.
- Two distinct velocity anomalies with unknown geologic interpretation were recovered.

Abstract We perform two-step surface wave tomography of phase-velocity dispersion curves obtained by ambient noise cross-correlations in the Bohemian Massif. In the first step, the inter-station dispersion curves were inverted for each period (ranging between 4 and 20 s) separately into phase-velocity maps using 2D adjoint method. In the second step, we perform Bayesian inversion of the set of the phase-velocity maps into an S-wave velocity model. To sample the posterior probability density function, the parallel tempering algorithm is employed providing over 1 million models. From the model samples, not only mean model but

[1]Department of Geophysics, Faculty of Mathematics and Physics, Charles University in Prague, V Holešovičkách 2, 18000 Prague, Czech Republic
[2]Center for Lithospheric Research, Czech Geological Survey, Klárov 3, 11821 Prague 1, Czech Republic

also its uncertainty is determined to appraise the reliable features. The model is correlated with known main geologic structures of the Bohemian Massif. The uppermost low-velocity anomalies are in agreement with thick sedimentary basins. In deeper parts (4-20km), the S-wave velocity anomalies correspond, in general, to main tectonic domains of the Bohemian Massif. The exception is a stable low-velocity body in the middle of the high-velocity Moldanubian domain and high-velocity body resembling a promontory of the Moldanubian into the Teplá-Barrandian domain. The most pronounced (high-velocity) anomaly is located beneath the Eger Rift that is a part of a Tertiary rift system across Europe.

Keywords: ambient noise tomography; Bayesian inversion; Bohemian Massif; geologic domains

4.1 Introduction

Earth's ambient seismic noise, generated mainly by processes in oceans and atmosphere, has been present in seismic records to distress and vex seismologists for many years. It was only recently that the recordings of the ambient noise were recognized to be useful: by cross-correlating long series of noise recordings between two stations the Green's function between them may be obtained (Campillo and Paul (2003); Shapiro and Campillo (2004)). The Greens' functions represent response in a given location to an impulsive source in another point and thus contain purely information about the seismic properties of the media between the two points. Therefore, it is natural to use them in tomography.

In the first ambient noise tomography applications, the results were presented by a set of 2D group (and later phase) velocity dispersion maps obtained by the inversion of traveltimes between station pairs for each period separately. The dispersion maps for various regions were estimated: US – Shapiro et al. (2005) and Lin et al. (2008), New Zealand – Lin et al. (2007), Korean peninsula – Cho et al. (2007), Europe – Yang et al. (2007), etc. They show good correlations with known geologic structures and may be used for preliminary intepretations.

The phase/group velocity dispersion maps at a set of periods can be also translated into a 3D velocity model. This has been done on various scales, from global (Nishida et al. (2009); Haned et al. (2015)) to regional (Li et al. (2010); Badal et al. (2013); Pang et al. (2016)), up to local networks, for example, around volcanos (Matos et al. (2015); Spica et al. (2015); Ryberg et al. (2016); Obermann et al. (2016)). The ambient noise traveltime dataset may be also combined with another dataset, for instance, with teleseismic traveltimes to increase the resolution at greater depths (e.g., Yang et al. (2008); Ouyang et al. (2014); Guo et al. (2016); Rawlinson et al. (2016)), or with receiver functions beneath the stations to improve sensitivity to interfaces in the velocity structure (e.g., Bodin et al. (2012); Guo et al. (2015); Shen et al. (2012); Růžek et al. (2012)). The ambient noise tomography has proved useful also for broader applications, such as tomography of ocean ridges using a network of ocean-bottom seismometers (Mordret et al. (2014); Zha et al. (2014)) or assessment of seismic anisotropy (Guo et al. (2012); Shirzad and Shomali (2014)).

Inversion of surface wave dispersion data into a 3D velocity model traditionally employs a two-step approach. In the first step, the period-dependent traveltimes based on dispersion curves measured between two points are inverted into phase/group velocity dispersion maps. In the second step, the dispersion curves on a regular grid are extracted and inverted for each grid point independently into a set of 1D velocity models. These 1D models are then assembled into a final 3D velocity model. In this second step, various techniques are employed. Standard deterministic approach is based on iterative linearized least square inversion (e.g. Li et al. (2010); Luo et al. (2012); Matos et al. (2015); Porritt et al. (2016)). However, stochastic approaches based on Monte Carlo (MC) methods are becoming more popular. The great advantage of the MC methods is that the result is not represented by a single model, but by a rather large set of models sampling the posterior probability density function. This is essential in the case of multimodal probability density functions, when the solution is nonunique and should not be represented by one particular model. The MC methods are able to explore several areas of model parameter space where the probability density function attains significant values. One may either extract coherent properties of the models (e.g., by model averaging), or assess uncertainties and correlations between the retrieved model parameters. Moreover, the MC solutions are numerically more stable – one does not have to deal with matrix inversion as in the case of the linearized inversion. Various MC methods have already been employed in ambient noise tomographic applications, for example, simulated annealing (Spica et al. (2014, 2015)), neighborhood algorithm (Mordret et al. (2014); Gao et al. (2011)) or other MC search algorithms (Jiang et al. (2014, 2016); Guo et al. (2015)). In our application, we employ another MC technique called parallel tempering, which was introduced into geophysical problems only recently by Sambridge (2014). The method is well balanced between fast convergence and avoiding entrapment in local maxima of the probability density function.

In the traditional two-step inversion, a 3D model is compiled from separately inverted 1D vertical models defined on a regular horizontal grid. The 1D models may be defined by a large number of velocity layers (Li et al. (2010); Pang et al. (2016); Porritt et al. (2016)) but also in a transdimensional way, where the number of model parameters is regarded as a hyperparameter (e.g., Young et al. (2013*b*,*a*); Pilia et al. (2015); Galetti et al. (2017)). However, we take a different approach: instead of performing 1D inversion for each point of the dispersion map separately, the 3D model is parametrized using fixed layers and the inversion is carried out for all the model parameters simultaneously.

We apply the methodology to the Bohemian Massif – a remnant of the Variscan orogen with a complex structure and history (see Fig. 4.1a) summarized in Section 4.2. The input data consist of phase velocity dispersion curves between the station pairs in periods 4–20s obtained from ambient-noise cross-correlations – see Section 4.3. In Section 4.4, we outline the methods applied in our problem. Then, we present the results in a form of dispersion maps and S-wave velocity model in Section 4.5. As a result of the limited period range of the dispersion data, the resulting model is bound to recover only the top 25 km of the Bohemian Massif crustal structure. In Section 4.6, the models are interpreted in terms of known geologic structures and compared with models obtained by other authors. The work represents the first 3D ambient noise tomography of the Bohemian

Massif. Moreover, since the inversion is performed in the Bayesian framework, the model uncertainty is estimated as well.

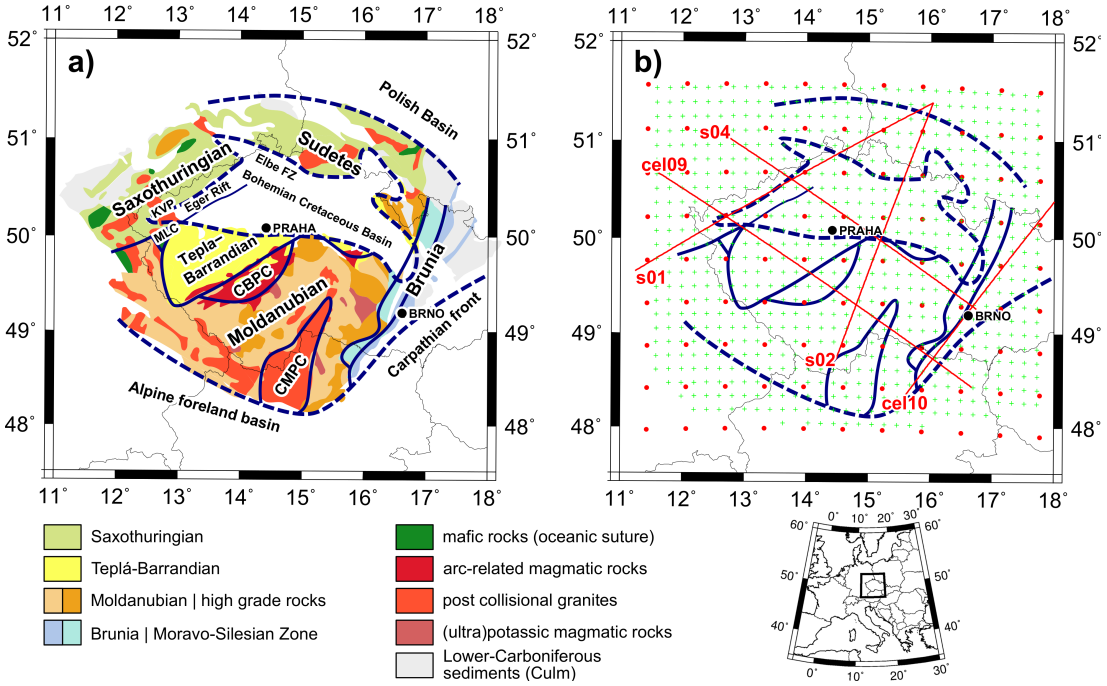

Figure 4.1: Map of the studied region: a) Main geologic structures of the Bohemian Massif modified after Franke (2000) and Schulmann et al. (2014). MLC – Mariánské Lázně Complex, CBPC – Central Bohemian Plutonic Complex, CMPC – Central Moldanubian Plutonic Complex, KVP – Karlovy Vary Pluton, FZ – fault zone. b) Distribution of points where the models are defined: Green pluses – data points where phase velocity dispersion maps are determined by 2D adjoint localization (first step). Red dots – model control points from which 3D velocity model is interpolated in the 3D inversion (second step). Selected seismic profiles measured during seismic experiments CELEBRATION 2000 (CEL09, CEL10) and SUDETES 2003 (S01, S02 and S04) across the studied region, are shown by red lines. Blue outlines show main geologic structures (solid – tectonic domains, dashed – post-Variscan sedimentary cover). Black thin outlines show state borders.

4.2 Bohemian Massif

4.2.1 Tectonic setting

The Bohemian Massif (Fig. 4.1a) is a relic of the European Variscan orogenic belt that formed ∼400–300 Myr ago as a result of convergence between Gondwana and Laurussia (Franke (2000); Matte (2001)). It consists of several major tectonic domains with different history and dominant rock types (for overview see Schulmann et al. (2009)). Most of them originally formed a part of the Gondwanan active margin, and separated as continental micro-plates during the Late-Cambrian–Ordovician times. When the motion of the plates changed and the region between Gondwana and Laurussia was closing, the Saxothuringian

domain was a part of the subducting plate (e.g., Franke (2000)). It recorded significant deformation and metamorphism with intensity increasing towards its south-eastern boundary where the oceanic suture was located. The most prominent relic of the suture is the Mariánské Lázně Complex that contains mafic rocks buried along a cold geotherm in a sequence typical for closure of an oceanic domain (Beard et al. (1995)).

The Teplá-Barrandian domain was a part of the upper plate during the orogeny and preserved the pre-Variscan upper crust including sedimentary sequences affected only by low-grade deformation and metamorphism (Drost et al. (2004)). The south-eastern margin of the Teplá-Barrandian domain was intruded by magmatic rocks together forming the Central Bohemian Plutonic Complex. The composition of these rocks corresponds to melting of mantle variably enriched by crustal component, which is typical for a magmatic arc above a subduction zone (Janoušek et al. (2000)).

The Moldanubian domain was a continuation of the Teplá-Barrandian domain during the orogeny, but it was affected by medium-to-high grade metamorphic conditions (Schulmann et al. (2009)). The contrasting character of the two adjacent domains results from their respective vertical displacement at a shear zone that developed in the weakened magmatic arc region (the Central Bohemian Plutonic Complex, Dörr and Zulauf (2010)). Erosion of the elevated Moldanubian surface then led to exposure of its middle and lower crust. The central part of the Moldanubian domain also contains a large accumulation of plutonic bodies that formed by crustal melting at the late stage of the orogeny – the Central Moldanubian Plutonic Complex (Finger et al. (2009)).

The easternmost part of the Massif is the Brunia (or Brunovistulian) domain. Along the margin with the Moldanubian domain, the Brunia-derived rocks were strongly deformed and metamorphosed within the so-called Moravo-Silesian Zone as a result of thrusting of Brunia underneath the Moldanubian domain (see references in Schulmann et al. (2009)). The interpretation of the Bouguer gravity anomaly suggests that the Brunia basement continues 50–70 km west of its surface margin and is covered only by a thin layer of Moldanubian rocks (Guy et al. (2011)). In addition, westward continuation of Brunia on the level of the mantle lithosphere was inferred from the seismic anisotropy (Babuška and Plomerová (2013)). The Brunia basement is covered by Variscan as well as younger sedimentary sequences (Kalvoda et al. (2008)).

During the collapse of the Variscan orogen, its elevated topography was subject to extension and gradual erosion which resulted in formation of several Permo–Carboniferous sedimentary basins. Later on in the Permian, subsidence of the Polish basin started eventually leading to formation of a several kilometers thick sedimentary cover north of the Bohemian Massif.

At ∼70 Myr ago, the onset of the Alpine orogeny led to activation of the Elbe Fault Zone, subsidence of the surrounding area and formation of the Bohemian Cretaceous Basin (Uličný et al. (2009)). Its today's thickness reaches 1 km and it covers a significant part of the Teplá-Barrandian domain. North-east of the Elbe Fault Zone, the Sudetes domain is located showing similar characteristics as the Saxothuringian domain.

The Alpine orogeny further reactivated existing fractured zones across Europe by rifting and related volcanism and sedimentation (Dèzes et al. (2004)). In the

Bohemian Massif the Eger Rift was activated along the Saxothuringian–Teplá-Barrandian boundary showing recent regular earthquake swarms, CO_2 emanations and elevated surface heat flow (Čermák (1994); Cloetingh et al. (2010)). Various seismic studies pointed to anomalous character of the crust and to asthenospheric updoming beneath the Eger Rift (e.g., Heuer et al. (2006); Geissler et al. (2005); Plomerová et al. (2007); Hrubcová and Środa (2015)). The Alpine orogeny also shaped the southern margin of the Bohemain Massif that is now covered by Alpine-Carpathian foreland basins and overthrust by the Western Carpathians in the east.

4.2.2 Previous tomographic studies

Due to its complex history and structure, the region of the Bohemian Massif has been subject to many seismologic studies starting from the 80s (see Novotný and Urban (1988)). Most of the recent crustal studies use active experiments along profiles, such as CELEBRATION 2000 and SUDETES 2003. The applied tomographic methods result in 2D vertical mainly P-wave velocity models. The crustal models along the active experiment profiles (see also Fig. 4.1b) were determined: CEL09 crossing most of the tectonic domains (Hrubcová et al. (2005); Novotný (2012)), CEL10 along the Moravo-Silesian Zone (Hrubcová et al. (2008)), S01 along the Eger Rift (Grad et al. (2008); Novotný et al. (2009)), S04 almost parallel with CEL09 (Hrubcová et al. (2010)). Models along several profiles concentrating on the Bohemian Massif were acquired by Růžek et al. (2007) and in the Sudetes by Majdański et al. (2006). The models show some common basic characteristics of the Bohemian Massif: firstly, there is very thin or no sedimentary coverage on the Bohemian Massif in the seismic data. Secondly, one can divide the crust into upper and lower one with P-wave velocity contrast of $\sim$0.5km/s in almost all tectonic domains. Usually, both upper and lower crust is strongly homogenized with low vertical gradient. In the subsurface parts, the high-velocity bodies are correlated with mafic intrusions whereas the low-velocity anomalies with granitic structures. In the upper crust, the distinct anomalies were found in the area of the Eger Rift. The Moldanubian domain shows higher P-wave velocities compared to the other domains representing the high-grade rocks. The Moho beneath the Bohemian Massif reaches to depths 30–40 km.

Thanks to numerous experimental data, tomographic inversion into 3D models was performed as well, for example, focused on the Moravo-Silesian region by Růžek et al. (2011), or over larger regions: P-wave velocity model for the Sudetes region by Majdański et al. (2007) or S-wave velocity model of the Bohemian Massif and Alps by Behm (2009). These models show good correlation of the subsurface structures with known geology. 3D P- and S-wave local velocity model of West Bohemia using not only experimental but also local earthquake data was acquired by Růžek and Horálek (2013). In this model, the areas with low Poisson ratio correlate with focal zones of the West-Bohemia earthquakes. Several crustal models were compiled together into a 3D model using different interpolation techniques by Karousová et al. (2012). The map of the Moho depth extracted from this model shows significant crustal thickening in the southern part of the Moldanubian domain and in the Brunia domain.

The crustal domains of the Bohemian Massif were found to continue also in the mantle lithosphere where they are characterized by different P-wave anisotropy patterns (Plomerová et al. (2005); Babuška and Plomerová (2013)).

1D S-wave velocity models beneath the stations in the Bohemian Massif are provided by receiver function studies (e.g., Wilde-Piórko et al. (2005); Geissler et al. (2012); Heuer et al. (2006)). Receiver function studies combined with ambient noise cross-correlation data focused on the crust of the Bohemian Massif were performed by Růžek et al. (2012). Beneath the stations located in different tectonic domains, the coherent properties in the receiver functions may be found. Except for the Moho deepening under the Moldanubian, the Moho updoming was found not only beneath the Eger Rift, but also in the area of the Bohemian Cretaceous Basin where it was connected to the tectonic activity along the Elbe Fault Zone.

The stepping stone for our work are the ambient noise cross-correlation studies by Růžek et al. (2016) who investigated the S-wave velocity properties for different domains of the Bohemian Massif. The domains show slightly different characteristics, but the variability within each domain is comparable with the differences among them.

In our work, we utilize these ambient noise cross-correlation data to obtain 3D S-wave crustal velocity model of the whole Bohemian Massif.

4.3 Data

We adopt inter-station dispersion curves estimated independently for all components (transverse T, radial R and vertical Z), which were extracted from ambient noise cross-correlations by Růžek et al. (2016). The stations used in the processing are a) permanent stations belonging either to the Czech Regional Seismological Network (CRSN) or the Virtual European Broadband Seismological Network (VEBSN); b) temporary stations operating within experiments BOHEMA I–III or PASSEQ (Plomerová et al. (2003); Babuška et al. (2005); Wilde-Piórko et al. (2008)); c) stations belonging to adjacent regional networks (Saxonian and Bavarian). The total number of stations is 72. The selected stations are equipped with broadband sensors and the noise measurements span over a time period of 12 years. To extract dispersion curves between the stations, the cross-correlations of the preprocessed noise between pairs of stations were calculated. More details on the method and resulting inter-station dispersion curves can be found in Růžek et al. (2016). The resulting inter-station dispersion curves range between periods 4 and 20 s. In the inversion, we employ only phase velocity dispersion curves with high signal-to-noise ratio in the cross-correlation function.

Fig. 4.2 shows the station and data coverage for the R component. The number of measured traveltimes is almost 700. Note, that for the other two components, the data coverage is even better due to generally higher-quality cross-correlation functions.

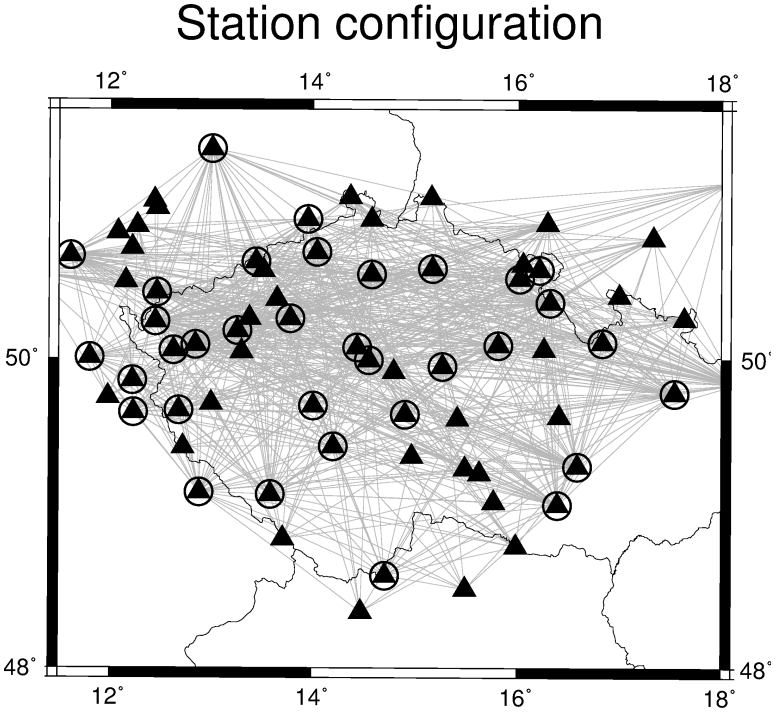

Figure 4.2: Station configuration of the investigated area. The station-pairs with measured phase-velocity dispersion curve for the R component are connected by lines. The selected stations (marked by circle) serve as source for the adjoint inversion.

4.4 Methods

The inversion method is derived from the traditional two-step approach described in Section 4.1. In the first step, we invert the inter-station dispersion curves separately for the selected periods into a 2D regular grid (so-called phase velocity dispersion maps) using 2D adjoint method. For the second step, we perform the inversion of the dispersion maps into a 3D S-wave velocity model in a Bayesian framework.

4.4.1 Adjoint localization

To obtain phase velocity dispersion maps, we apply 2D adjoint inversion (Tromp et al. (2005); Fichtner et al. (2006); Tape et al. (2007)) to inter-station dispersion curves for selected periods. The station coverage as well as the number of dispersion traveltimes in the studied region may be considered sufficient for the tomographic problem (see Fig. 4.2).

The adjoint method belongs to iterative gradient methods of the misfit minimization. The misfit considered in our problem is the L2 norm of the cross-correlation traveltime differences. The misfit gradient (i.e., the derivative of the misfit with respect to the model parameters) is calculated using one forward and one so-called adjoint calculation. In our problem, the adjoint calculation is represented by backpropagation of so-called adjoint wavefield from the receivers back to the sources (Peter et al. (2007); Tape et al. (2007)). The forward and adjoint

wavefields are then combined into finite-frequency or sensitivity kernels, which emphasize areas with increased values of the misfit gradient.

The greatest advantage of the adjoint method is that it accounts for the finite-frequency effects of the wave propagation. Nevertheless, the method requires regularization, for example, by means of Gaussian smoothing that is applied to the calculated sensitivity kernels (Tape et al. (2007, 2010); Peter et al. (2011)). The width of the Gaussian smoothing function as well as the optimal number of iteration steps was determined using preliminary synthetic tests with actual data noise (Valentová et al. (2015)). More details are to be found in Section 4.4.3.

4.4.2 Bayesian inversion

We apply a Bayesian approach to solve the second stage of the inverse problem. The result of the inversion is represented by a set of model samples obtained by a random walk according to the posterior PDF. The method is appropriate for non-linear problems, such as inversion of dispersion curves. Moreover, from the model samples one may estimate not only the best or the mean model, but also its uncertainty.

The posterior PDF is defined as conditional PDF on model parameter space after measurement of data $\boldsymbol{d}$ is acquired (e.g., Tarantola and Valette (1982); Mosegaard and Tarantola (1995); Tarantola (2005)). This is usually expressed using the Bayes theorem:

$$p(\boldsymbol{m}|\boldsymbol{d}) = \frac{p(\boldsymbol{m})p(\boldsymbol{d}|\boldsymbol{m})}{p(\boldsymbol{d})}, \tag{4.1}$$

where $p(\boldsymbol{m})$ is the model parameter PDF which is independent on the data measurement (i.e., prior PDF, denoted p_{prior}). Conditional probability of observing data given model $\boldsymbol{m}$, $p(\boldsymbol{d}|\boldsymbol{m})$, is so-called likelihood function. The likelihood function for measured data $\boldsymbol{d} = \boldsymbol{d}_{\text{obs}}$ contains statistical information on the data measurement error, but may also include modeling error. However, the information on the modeling error is usually difficult to estimate or negligible compared to the measurement error, and is therefore not assumed. The posterior PDF as a function of model parameters $\boldsymbol{m}$ for measured data $\boldsymbol{d} = \boldsymbol{d}_{\text{obs}}$ can be rewritten as

$$p(\boldsymbol{m}) = k\ p_{\text{prior}}(\boldsymbol{m})p(\boldsymbol{d}_{\text{obs}}|\boldsymbol{m}) \tag{4.2}$$

where k is a PDF normalization constant.

Usually, the data PDF is considered in a form of Gaussian distribution, then

$$p(\boldsymbol{d}_{obs}|\boldsymbol{m}) \propto exp(-S(\boldsymbol{m})), \tag{4.3}$$

where $S(\boldsymbol{m})$ defines misfit between measured data $\boldsymbol{d}_{\text{obs}}$ and synthetics calculated using a theoretical relation $\boldsymbol{g}(\boldsymbol{m})$ with Gaussian covariance matrix $\mathbf{C}_{\text{d}}$:

$$S(\boldsymbol{m}) = (\boldsymbol{d}_{\text{obs}} - \boldsymbol{g}(\boldsymbol{m}))^T \mathbf{C}_{\text{d}}^{-1} (\boldsymbol{d}_{\text{obs}} - \boldsymbol{g}(\boldsymbol{m})) \tag{4.4}$$

To draw samples from the model space according to the posterior PDF, the Markov chain MC random walker is employed. In each step of the chain, the model parameters are randomly perturbed considering Gaussian probability function. The new (proposed) model is accepted or rejected based on the Metropolis

algorithm: if the posterior PDF of the proposed model is higher, the model is accepted; if the PDF is lower, the model may be still accepted with some probability. The acceptance rate depends on the width of the Gaussian generating the perturbations, i.e., the perturbation of the misfit. To increase the efficiency of the sampler we apply a method called parallel tempering (PT, Sambridge (2014)). The PT algorithm is similar to the better-known simulated annealing, as it introduces modification of the PDF by an additional parameter called temperature T. The modified PDF $p(\boldsymbol{m}, T)$ is given by

$$p(\boldsymbol{m}, T) = k\ p_{\text{prior}}(\boldsymbol{m}) \exp\left(\frac{-S(\boldsymbol{m})}{T}\right). \tag{4.5}$$

The samples are drawn following this modified PDF assuming multiple values of the temperature T.

For high temperatures, the PDF becomes smooth (note that for $T \to \infty$, $p \to p_{prior}$). For low temperatures, the PDF maxima are more pronounced and as $T \to 0$, PDF converges to δ function located in the global maximum (i.e., global minimum of $S(\boldsymbol{m})$). For $T = 1$, the modified PDF equals to the original PDF. During the simulated annealing, temperature T gradually decreases from high values to lower values so that the sampler gradually concentrates into areas with higher PDF values. In the PT method, multiple Markov chains each with different temperature sample the model parameter space simultaneously. The chains with lower temperature values sample locally areas of PDF maxima, whereas chains with higher temperatures are able to escape the local maxima of PDF. Moreover, two chains can exchange their temperature values between the chain advances. The probability that two chains (denoted $[\boldsymbol{m}_i, T_i]$ and $[\boldsymbol{m}_j, T_j]$) swap their temperatures (resulting in $[\boldsymbol{m}_i, T_j]$ and $[\boldsymbol{m}_j, T_i]$) is given by the Metropolis-Hastings criterion. This condition ensures, that the subset of samples for temperature $T = 1$ represents sampling of the original untempered PDF $p(\boldsymbol{m})$. The PT algorithm is well balanced between efficiency and stability for complex multimodal PDFs (Sambridge (2014)). Another advantage is straightforward parallelization of the sampling computer code.

4.4.3 Implementation details

In the 2D adjoint localization, the regularization is applied primarily by means of Gaussian smoothing of the sensitivity kernels and by the number of iterations. The regularization parameters are estimated by synthetic tests using a smooth and a complex model to examine reliability of both long-scale and short-scale features. For the 20 s Love data inversion, the width of the Gaussian smoothing function is set to 100km with iterations stopping at 6th step (for more information and related synthetic tests see Valentová et al. (2015)). For other periods (both Love and Rayleigh datasets), the size of the Gaussian function scales with the corresponding wavelength.

The phase velocity dispersion maps are calculated for periods (4, 6, 8, 10, 12, 16, 20) s in the first stage and serve as input data in the second step of the inversion in which the Bayesian approach is applied. We extract dispersion curves in a selected 2D regular grid of data points (green pluses in Fig. 4.1b). The 3D velocity model is represented by a set of vertical 1D layered models on a regular horizontal grid of model control points (red dots in Fig. 4.1b). The spacing of the data

and model control points is 16 km and 50 km, respectively. In each model control point, we assume 7 layers above a halfspace with interfaces at depths of 2, 4, 8, 12, 18, 24, and 32 km. Although the model is parametrized in 3D, the synthetic dispersion curves are calculated for computational reasons assuming 1D layered model at each data point. In the MC methods, the forward problem is calculated numerous times and thus a fast solver (such as 1D) is necessary. We employ the code VDISP which is based on a matrix method using Thomson-Haskell and Watson's matrices for Love and Rayleigh waves, respectively (Novotný (1999)). To obtain a 1D layered model for the synthetic dispersion curve calculation, the model is interpolated in each layer from the model control points into the data grid points by cubic spline interpolation.

The Bayesian inversion is regularized by the adopted parametrization. Firstly, the number of layers is rather low and with fixed thicknesses; the number of model control points is kept lower than the number of data points. Furthermore, the model is interpolated into the data points using smooth functions (cubic splines).

The Bayesian inversion employs the PT algorithm to sample the posterior PDF. As the model prior information $p_{\text{prior}}(\mathbf{m})$ we use homogeneous PDF on a selected interval for all model parameters, in particular between 1 and 15 km/s. Another constraint on the model parameters is the requirement of a nonnegative velocity gradient with depth for each model control point. For simplicity, the data covariance matrix $\mathbf{C}_{\text{d}}$ is assumed to be diagonal with standard deviation corresponding to error of dispersion maps (see Section 4.5.1).

We run the computations in parallel on 12–24 CPUs, the temperature T ranges between 1 and 50. To remove the dependency of Markov chains on the starting model, we introduce long burn-in phase (10k steps) when the accepted models are not saved. The main (post-burn) phase is stopped when the sampler appears to have converged on the model space (i.e., no improvement in model space sampling occurs with further thousands of chain steps). Only the samples from chains at temperature $T = 1$ accepted every 100th step are saved and further processed. The resulting number of samples from the posterior PDF exceeds 1 million.

4.5 Results

In the first step, the inter-station dispersion curves were inverted for each period separately by the 2D adjoint method. The results are represented by a set of phase velocity maps in Section 4.5.1.

In Section 4.5.2, we present the results of the Bayesian inversion of the dispersion maps into a 1D S-wave velocity model (and v_p/v_s) using different datasets and assumptions. The 1D Bayesian model of the Bohemian Massif reflects its average structure. Moreover, with the help of synthetic tests we discuss to what extent the variance in the 1D velocity model parameters contains also variability due to the 3D structures.

In Section 4.5.3, the results of the Bayesian inversion of the dispersion maps into a 3D model is presented as 2D depth-slices for both absolute and relative S-wave velocities, as well as horizontal variations of the v_p/v_s ratio. The correlations of the results with geology are discussed in Section 4.6.

4.5.1 Phase velocity maps

Fig. 4.3 shows the inverted phase velocity dispersion maps for each period as obtained from the three components individually: Love (T) and both Rayleigh (R and Z). The perturbations in phase velocities are shown with respect to average values specified on the right side of each map.

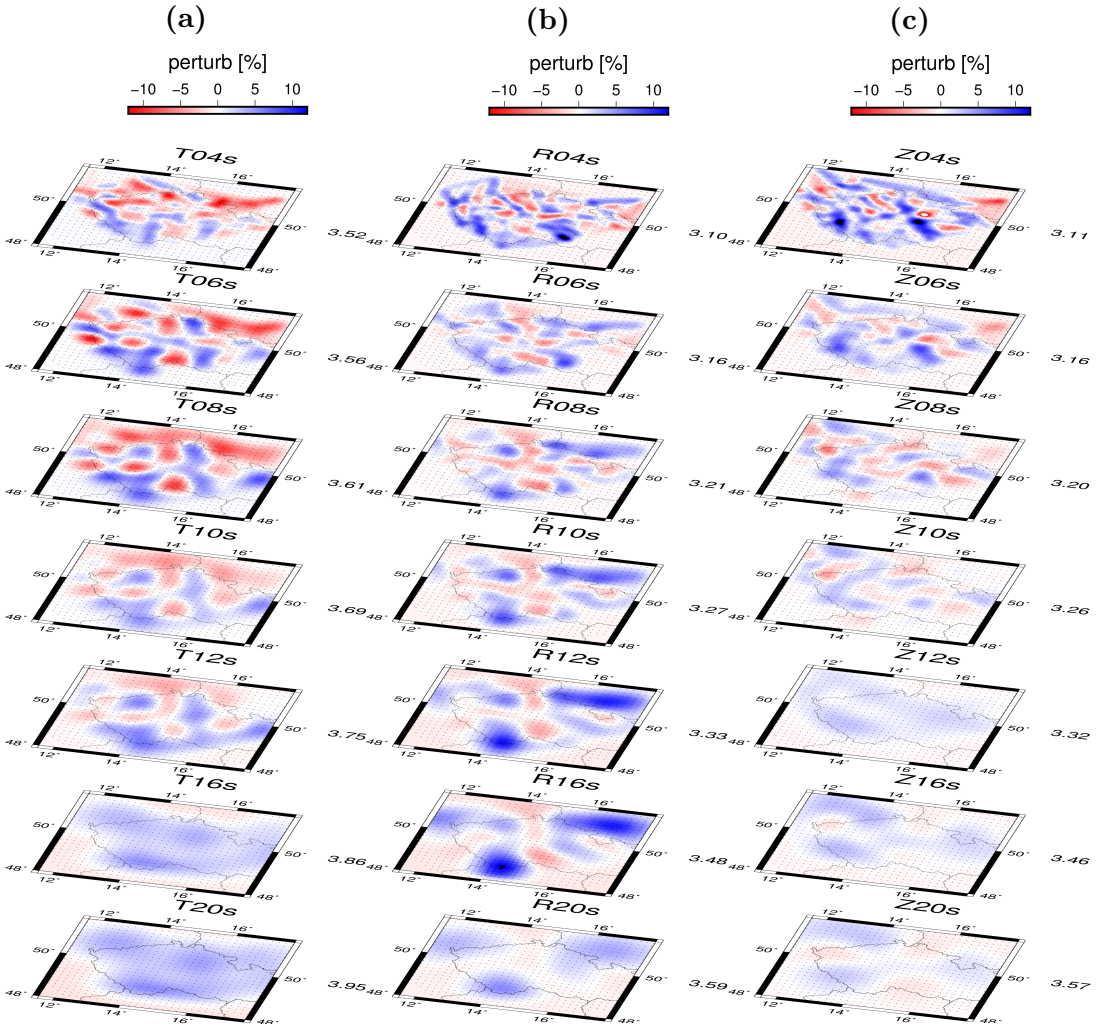

Figure 4.3: Phase velocity dispersion maps obtained by the 2D adjoint inversion for each period from all components: a) transverse T (i.e., Love wave), b) radial R c) vertical Z (both Rayleigh waves). The period increases from top to bottom (see legend). The model perturbations (color scale) are presented relative to mean phase velocity model, denoted for each map to the right in km/s.

We treat the dispersion curves for R and Z component as two independent datasets, so that by comparing the resulting models we get insight into the accuracy of the first part of the inversion. Regarding the average models, both components give very similar results. Contrarily, the phase-velocity perturbations of the R and Z component, although similar in some features, differ in some details as well.

We use differences between both Rayleigh dispersion maps to estimate the error of the dispersion maps. The RMS of the differences ranges between $0.10 - 0.15$ s for different periods, which agrees well with the value of 0.14 s estimated

from Rayleigh phase velocity differences between inter-station pairs by Růžek et al. (2016).

4.5.2 1D velocity model

Here, we apply our stochastic MC inversion to obtain a representative 1D layered velocity model for the real dataset. We assemble data from all phase velocity maps, excluding points where the initial velocity remains unchanged (i.e., there is no sensitivity of the data on the velocity structure).

1D inversion is computationally cheap ($\approx$ 4 h on 8 CPUs) and thus may be used with the help of synthetic 1D tests to investigate properties of the inverse problem. The 1D inversions were performed employing different datasets: only Love phase velocities or all-component phase velocities. The 1D model is composed of 2 km thick layers down to 32 km and a halfspace. Furthermore, the addition of the v_p/v_s model parameter to 1D S-wave velocity parameters was examined.

Fig. 4.4a shows results of Love phase velocity inversion into a 1D S-wave velocity model. The model appears to have a relatively wide but unique maximum. This shows that although the result is stable, either the data sensitivity to the S-wave velocity model is low or the 1D model is inappropriate to represent the dataset due to the actual higher spatial variability of the true velocity model, or both.

In Fig. 4.4b, we show the results of inversion of Love and Rayleigh phase velocity dispersion data into a 1D layered S-wave velocity model with fixed $v_p/v_s = 1.57$. By increasing the number of data, the variance of the model parameters decreases. This suggests that the data from all components are well represented by a common 1D S-wave velocity model.

Fig. 4.4c displays models obtained from the inversion of all components into a 1D S-wave velocity model and an additional parameter – depth-independent v_p/v_s. The resulting S-wave velocity model is similar to the previous one (Fig. 4.4b), both in terms of the best model and the variance. This shows that the inversion of the v_p/v_s parameter does not affect the inversion of the S-wave velocity model much, except for slightly amplifying the velocity increase at $\sim$20 km. The obtained v_p/v_s ratio has low variance and appears to be well resolved. However, by assuming depth-independent v_p/v_s, the inversion estimates v_p/v_s corresponding to the average over all depths. In general, the averaged parameter is better constrained than v_p/v_s at a given depth. Therefore, it must not be interpreted as a low variability of the v_p/v_s of the real 3D structure.

From the 1D velocity models, it is difficult to distinguish between the effects of data resolution and inherent variability due to the 3D structure. To separate these two effects, we performed several synthetic tests. The synthetic tests with both noiseless and noisy data were performed with similar setting as the real data inversion. The resulting variance of the real data inversion is distinctly larger than that of the synthetic tests. This suggests that the most of the 1D model variability is due to the lateral inhomogeneity of the real crust.

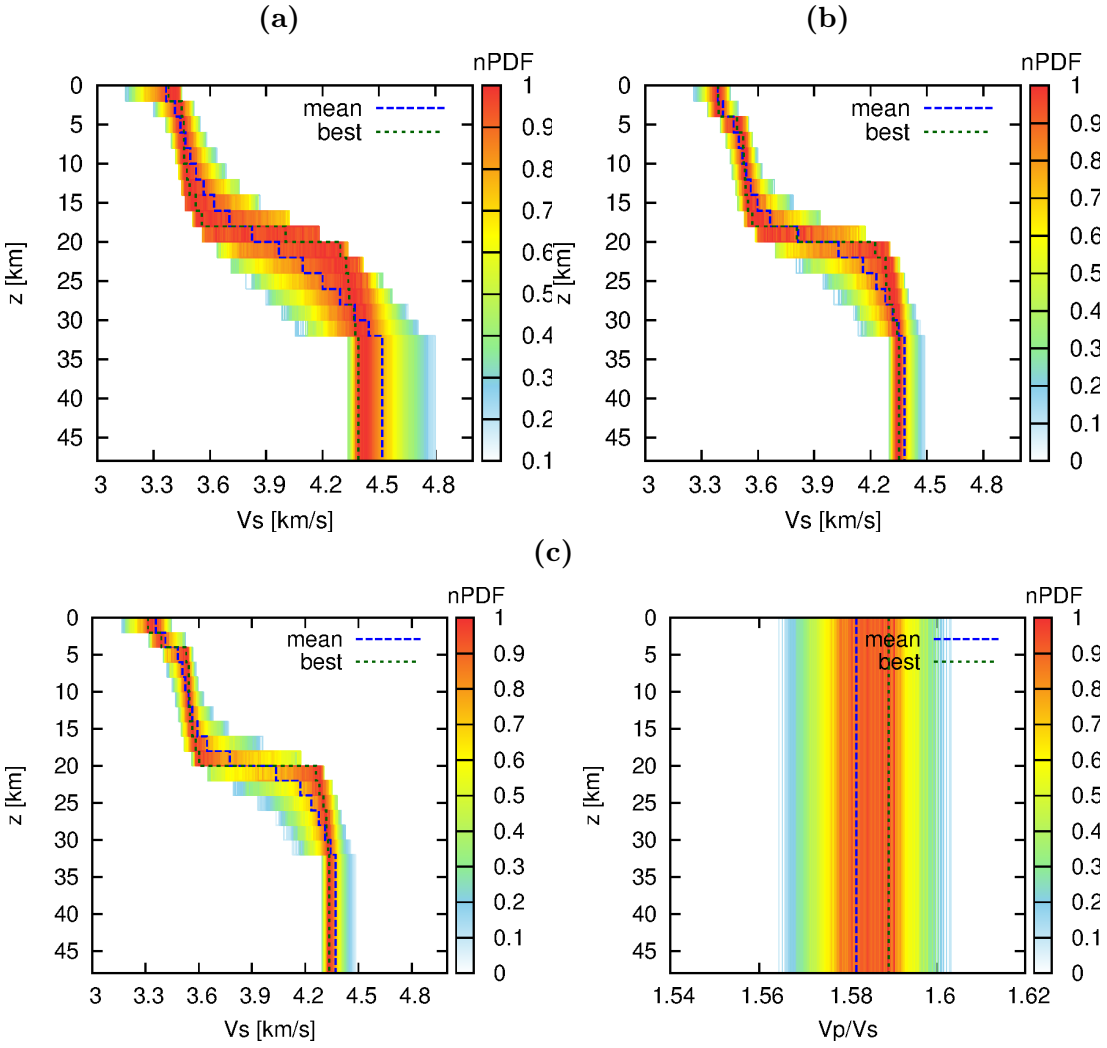

Figure 4.4: Results of the inversions of the phase velocity dispersion maps into 1D layered models by the PT algorithm. The color scale corresponds to the normalized PDF (nPDF) scaled by the PDF value of the best model (black dashed line). a) Inversion of Love dispersion maps into a 1D S-wave velocity model. b) Inversion of all Love and Rayleigh dispersion maps into a 1D S-wave velocity model, v_p/v_s being fixed. c) Inversion of all Love and Rayleigh dispersion maps into a 1D S-wave velocity model (left) and depth-independent v_p/v_s ratio (right).

4.5.3 3D velocity models

The result of our Bayesian inversion consists of > 1 million PDF samples (i.e., 3D models). To present the results, we display the mean model calculated from all models and the best model. The advantage of the mean model is that it presents only stable features. Therefore, the mean model is typically smoother than any single model drawn by the MC sampler. The best model is shown as a representative model to examine differences in properties between a single model and the averaged one.

Fig. 4.5 presents the depth-slices of the mean and the best 3D S-wave velocity model. To better visualize its lateral variations, Fig. 4.6 shows the same models but in terms of perturbations relative to the horizontally averaged model.

All models show increase of S-wave velocity with depth. Regarding the lateral variations, both mean and best model share the most significant structures. The differences between them appear on smaller scales.

Simultaneously with S-wave velocities, we have performed the PT exploration also for the depth-independent v_p/v_s ratio. The reasons for the v_p/v_s depth-independence are: a) poor resolution of v_p/v_s and b) to reduce the parameter space. The resulting mean model and its standard deviation are shown in Fig. 4.7. Although, the range of the v_p/v_s ratio is wide (1.5–1.8), the standard deviation in recovered areas lies mostly below 0.05, suggesting strong horizontal structural variability of this parameter.

4.5.4 Uncertainty of the 3D model

The greatest advantage of employing MC methods to solve the inverse problem lies in the plurality of models representing the solution. As an example, Fig. 4.8a shows vertical 1D models in a model control point located in the middle of our domain. From this example, we see that the best resolved part in the inversion lies at depths of 2–18 km. At greater depths, the variance in the S-wave velocities is very high. It also appears that the PDF of the S-wave velocities in deeper parts as well as v_p/v_s ratio have 2 local maxima. We ascribe this to the undersampling of the PDF in the particular parameter domain. Also note that for model control points located at the boundaries of our domain, the overall uncertainty increases.

Uncertainty of the model along a profile can be estimated by standard deviation of the mean model (Fig. 4.8b right). In general, the lowest uncertainty (as indicated also by the 1D models in Fig. 4.8a) is achieved down to ∼20 km. However, the uncertainty changes also laterally along the profile (between 1–2% for the well resolved part). Alternatively, we visualize these changes via standard deviation of selected S-wave velocity isolines (Fig. 4.8b left).

We emphasize that we should be extremely cautious when interpreting the imaged structures below 25 km, where the model variance increases rapidly. In particular, our models are not suited for the search of the Moho. The following geologic interpretation should be confined to large-scale structures only. This limitation is a consequence of the method applied, namely: a) employment of surface wave data, which is inherently sensitive to the averaged (smoothed) structures both horizontally and vertically, making it impossible to obtain velocity interfaces; b) model parametrization – horizontal grid with relatively large (50 km)

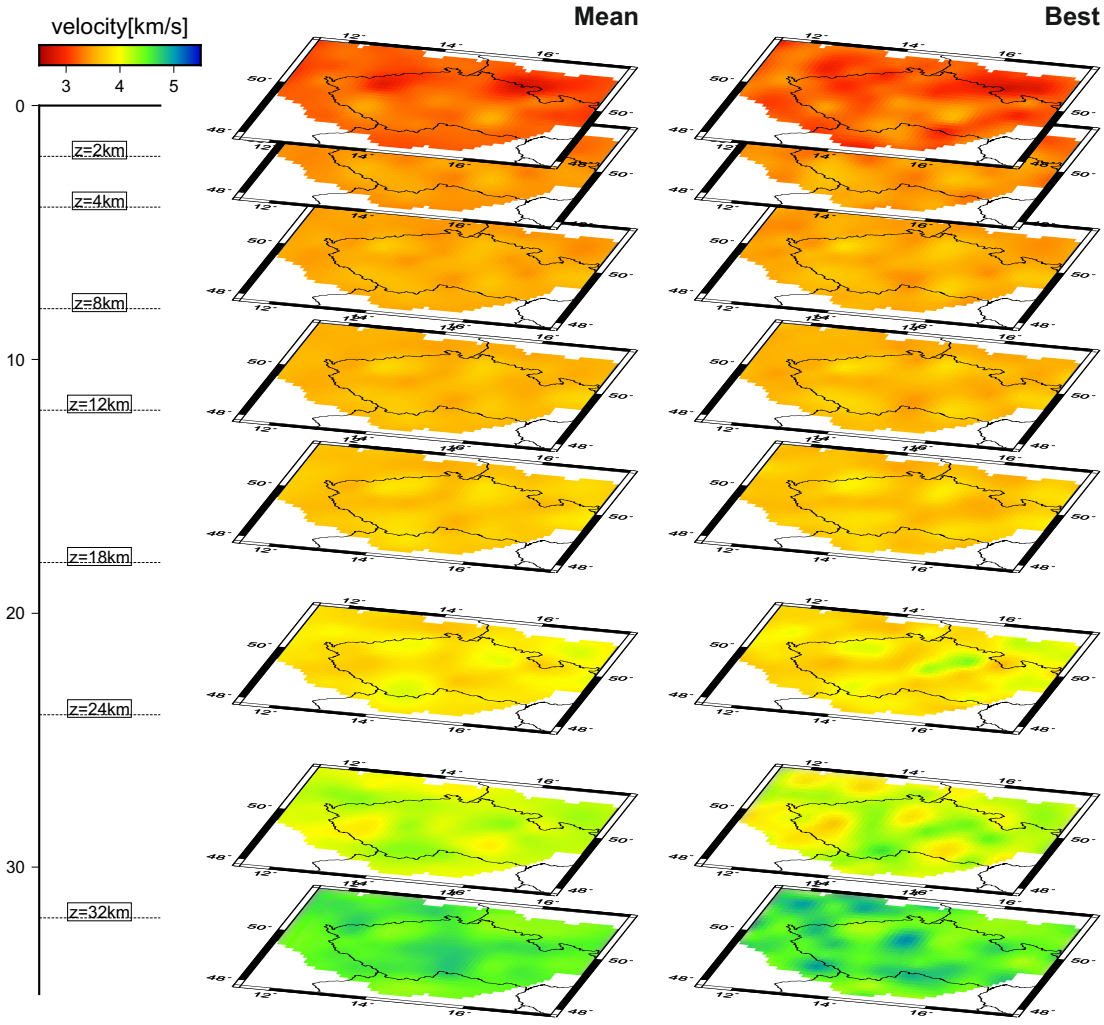

Figure 4.5: Depth slices through the mean (left) and best (right) S-wave velocity models obtained by our 3D inversion using the PT algorithm. The layers are shown in the left. The areas with no station coverage are masked.

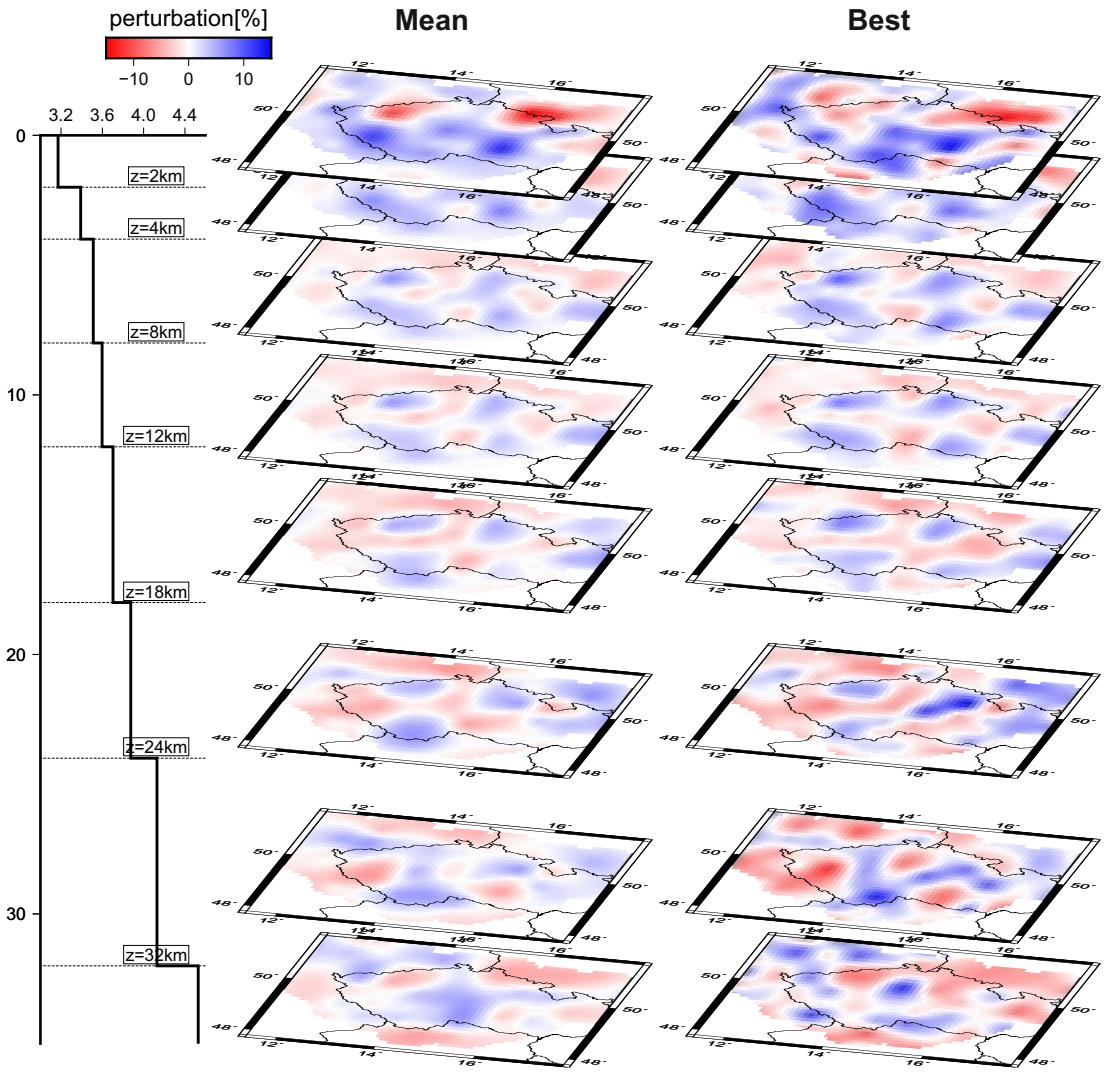

Figure 4.6: The same as Fig. 4.5 but for S-wave velocity perturbations calculated as relative differences from horizontally averaged velocities (left panel).

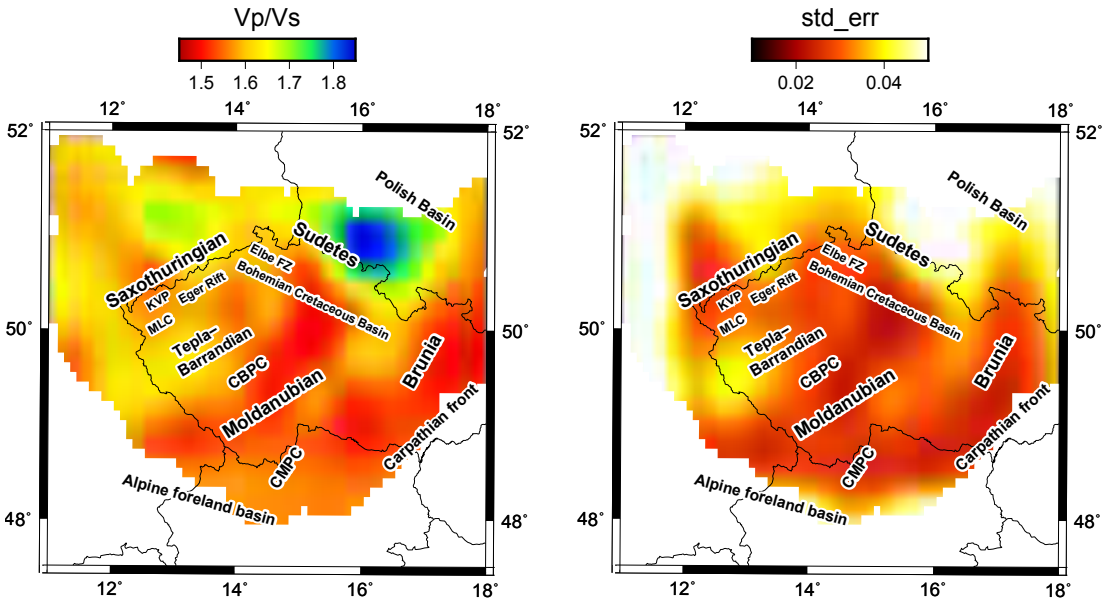

Figure 4.7: Mean model and standard deviation of v_p/v_s value obtained by our 3D inversion. The v_p/v_s ratio is assumed to be depth-independent.

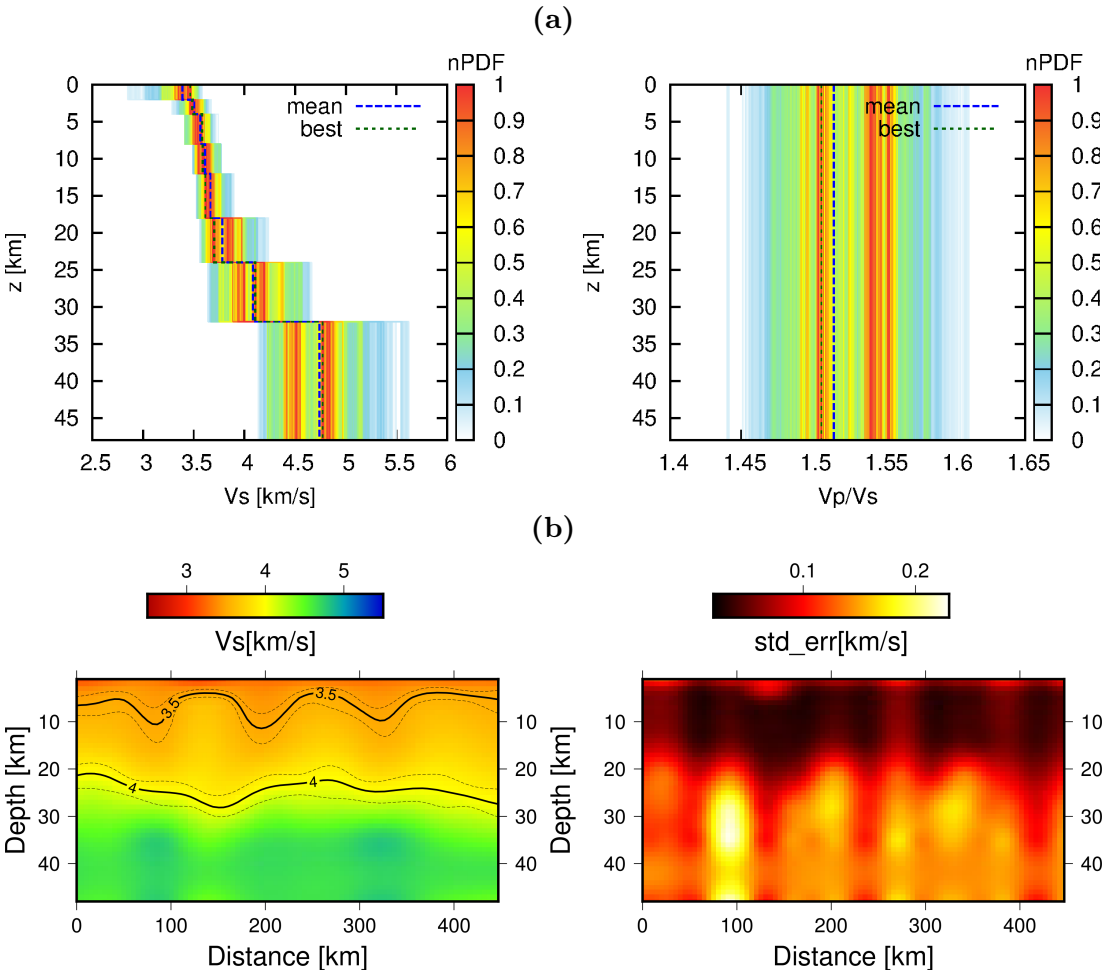

Figure 4.8: a) 1D vertical models in a selected model control point in the middle of the domain. Color palette shows nPDF – PDF normalized to its maximum (best model, see legend). b) 2D mean model and its variance interpolated along the CEL09 profile. Two velocity isolines are shown with their standard deviation.

spacing; and c) averaging of great amount of single models generated by the MC inversion which produces stable but smooth structures.

4.6 Discussion

4.6.1 Geological interpretation of dispersion maps and 1D S-wave velocity profile

The result of the first part of the inversion – the phase-velocity dispersion maps – usually correlate well with known geology and are used for preliminary interpretation (e.g., Saygin and Kennett (2010); Nicolson et al. (2012)). In our dispersion maps (see Section 4.5.1, Fig. 4.3), there is a high velocity structure in the southern part of the domain present in almost all maps, which may be related to the Moldanubian domain. Another stable high velocity anomaly, located in the center of the north-west border of the Czech Republic is found easily on maps for components T and R for periods 8–16 s, where it is surrounded by low velocities. Moreover, this anomaly can be also tracked for the Z component. This anomaly is situated beneath the Eger Rift zone, where a high velocity body is usually found in the tomography (Grad et al. (2008); Růžek et al. (2007); Alexandrakis et al. (2014); Mousavi et al. (2015)). For the shortest periods, the structures are much more complex and stable features present for all components are more difficult to determine. In the Elbe Fault Zone, one may observe a narrow low velocity anomaly for all maps at periods 4–6 s.

Before the inversion into 3D model, we performed Bayesian inversion of the dispersion maps into 1D S-wave velocity model and v_p/v_s ratio (see Section 4.5.2, Fig. 4.4). This 1D model may be considered as a representative model of the Bohemian massif. In the near-surface part (depth less than 4 km), there is a moderate velocity gradient and relatively large variance. This may point out to uneven sedimentary cover of the Bohemian Massif. The rest of the upper crust (from 4 to $\sim$ 12 km) shows very low velocity gradient and a very low variance. This suggests strongly homogenized upper crust across the whole domain.

Around 20 km depth, there is a strong S-wave velocity increase indicating significant structural difference between the upper and lower crust. The increased variance in these depths may be associated with the high velocity gradient being laterally heterogeneous in the real structure.

Below 25 km, there is no significant velocity jump corresponding to the Moho. Moreover, the variance in these depths decreases, namely in the halfspace. This is due to the fact that the variance estimated for the S-wave velocity at these depths is only formal as it represents variance of the S-wave velocity averaged over all depths below 32 km. Therefore, the obtained variance reflects neither vertical nor lateral variability of the real 3D structure, and thus the Moho is not resolved in our model.

The Bohemian Massif shows low v_p/v_s value $\sim$1.6 indicating rather rigid, consolidated material. The interpretation of the v_p/v_s horizontal perturbations in Fig. 4.7 is rather ambiguous as the uncertainty of this parameter may be poorly estimated. In similar sense as the S-wave velocity of the halfspace velocity mentioned above, the estimated v_p/v_s corresponds to the average over all depths.

The increase in the northern part (Sudetes), may be attributed to the presence of the thick sedimentary coverage.

4.6.2 Geological interpretation of 3D S-wave velocity model

Here, we discuss the correlation of our 3D mean S-wave velocity model with known geology with the help of vertical cross-sections (Fig. 4.9, 4.10), corresponding to selected profile measurements of active seismic experiments CELEBRATION 2000 (CEL09, CEL10) and SUDETES 2003 (S01, S02, S04). We also compare the cross-sections with 2D models obtained along these profiles by other authors, but with great caution. Indeed, the latter models were interpreted in terms of P-wave velocities and were based on other kind of data than in our case (body waves, frequency ranges).

Fig. 4.9 shows the vertical cross-sections through the mean and the best 3D models interpolated along the profiles. The mean isovelocity lines of 3.5 and 4.0km/s are delineated together with standard deviation representing their uncertainty. Furthermore, Fig. 4.10 shows the velocity model perturbations relative to their horizontal averages to better distinguish different structures.

Although the Bohemian Massif is composed of several main tectonic domains, its overall structure is much more complicated. As our results also imply, the correlation with geologic domains is not that obvious. This was already suggested by Růžek et al. (2016) who obtained rather heterogeneous models for each domain.

For the following discussion, to better distinguish the horizontal extent of the velocity structures, Fig. 4.11 displays maps of two isovelocity topographies – for values 3.25 and 3.6km/s. The values were chosen to show the near-surface structures (0–4 km) and the deeper parts (4–16 km), respectively.

Sedimentary basins

For the shallowest layers (0–2 km in Figs. 4.5 and 4.6, and Fig. 4.11a), we find good correlation of low velocity perturbations with sedimentary basins. The extensive Bohemian Cretaceous Basin is visible in our model mainly where the sedimentary cover is presumably thicker. Such area is located at the northern rim of the Bohemian Cretaceous Basin in the Sudetes (including the Intra-Sudetic Basin) continuing further to the Polish Basin, where the pronounced low-velocity anomaly is present. Thicker lower-velocity layer in this area is also visible on profile S02 at distances > 250km (see Figs. 4.9, 4.10), and can be also found in S02 velocity profiles by Růžek et al. (2007) and Majdański et al. (2006). Other such area corresponds to the Eger Rift where the low near-surface velocities are retrieved in a $\sim$50 km wide zone (see also profiles S01 and S04 in Figs. 4.9 and 4.10).

The anomaly corresponding to the sedimentary cover of Brunia is rather variable with its largest amplitude in the north-western part, where the thickness of the sedimentary layer increases. Similar anomalies were identified on CEL10 profile by Hrubcová et al. (2008) and Růžek et al. (2007).

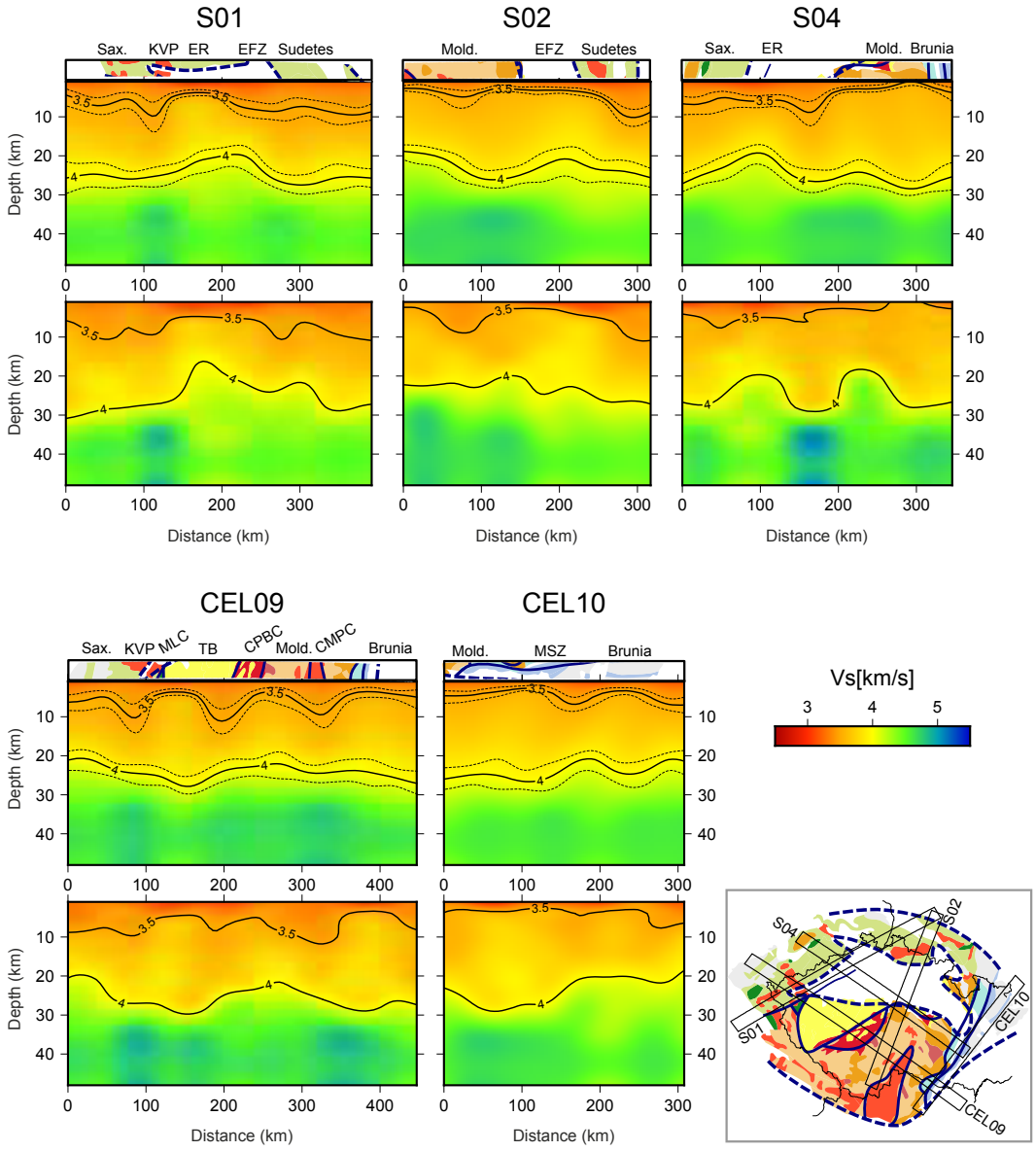

Figure 4.9: The mean (top) and best (bottom) S-wave velocity model in selected profiles across the studied domain. The isoline depths of S-wave velocity values 3.5 and 4.0 km/s are shown by black curves. In the mean model, the standard deviation around these values is shown by dashed curves. The main geologic structures are denoted on the top of each profile as indicated in the map in the inset panel. MLC – Mariánské Lázně Complex, CBPC – Central Bohemian Plutonic Complex, CMPC – Central Moldanubian Plutonic Complex, ER – Eger Rift, KVP – Karlovy Vary Pluton, EFZ – Elbe Fault Zone, Mold – Moldanubian, Sax – Saxothuringian, TB – Teplá-Barrandian, MSZ – Moravo-Silesian Zone.

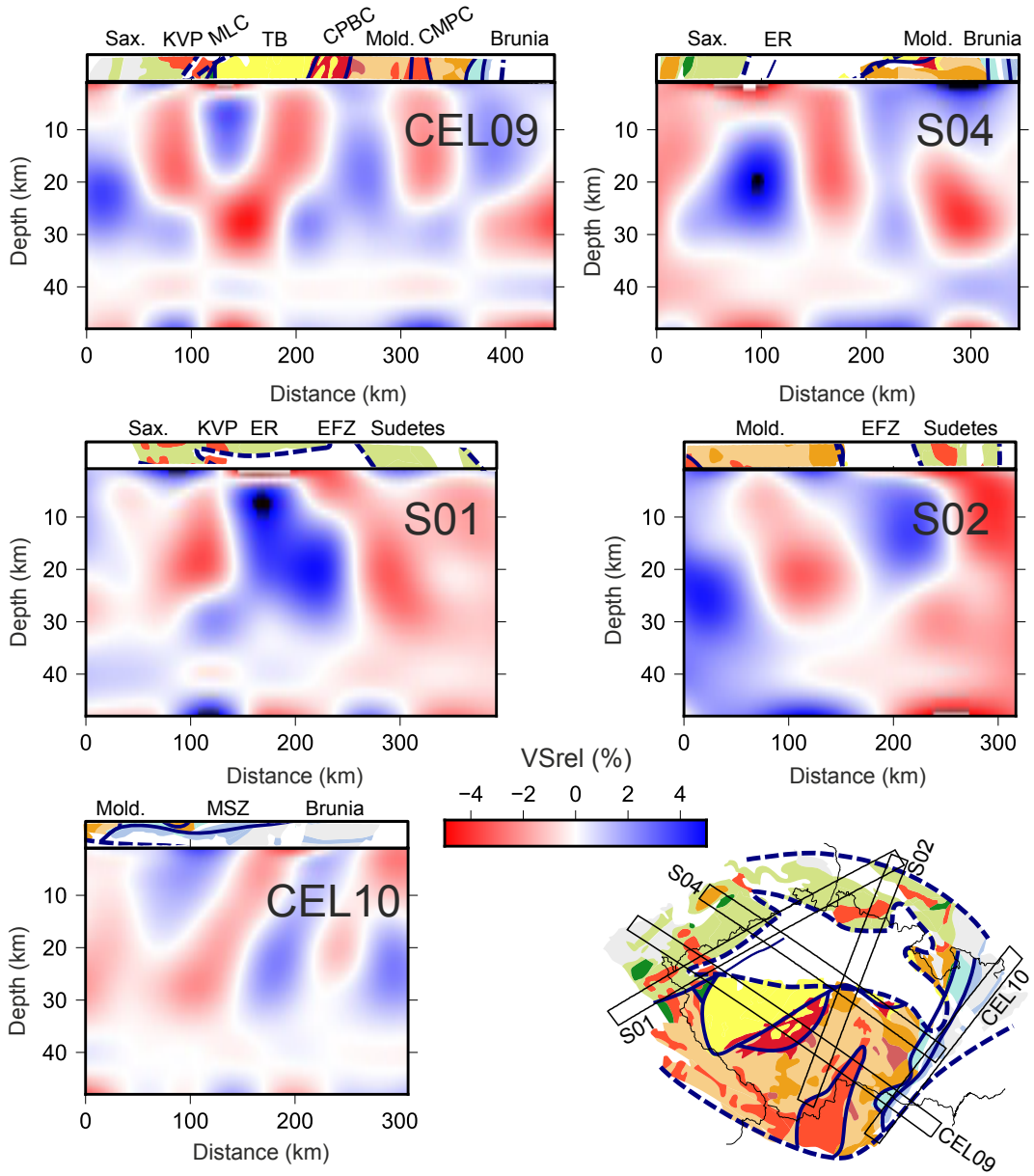

Figure 4.10: The S-wave velocity perturbations with respect to average model for each depth calculated from the mean model along the studied profiles. The main geologic structures are denoted on the top of each profile as indicated in the map in the inset panel. MLC – Mariánské Lázně Complex, CBPC – Central Bohemian Plutonic Complex, CMPC – Central Moldanubian Plutonic Complex, ER – Eger Rift, KVP – Karlovy Vary Pluton, EFZ – Elbe Fault Zone, Mold – Moldanubian, Sax – Saxothuringian, TB – Teplá-Barrandian, MSZ – Moravo-Silesian Zone.

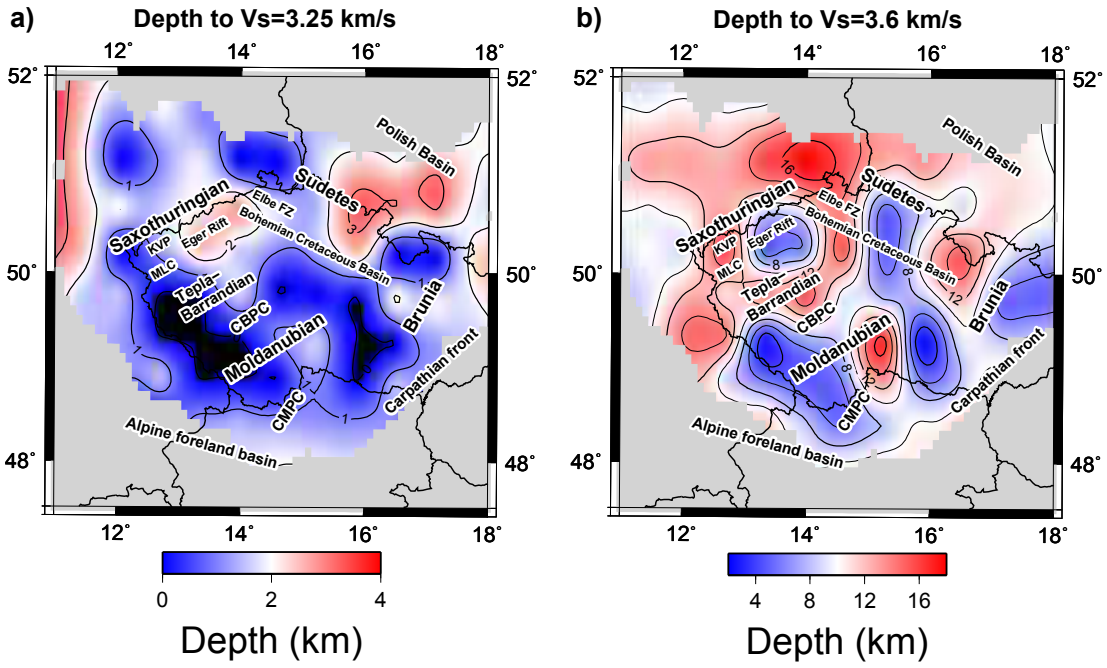

Figure 4.11: Topography of the iso-S-wave velocity surface for value 3.25km/s and 3.6km/s of the mean model. The main geologic structures are shown (MLC – Mariánské Lázně Complex, CBPC – Central Bohemian Plutonic Complex, CMPC – Central Moldanubian Plutonic Complex, KVP – Karlovy Vary Pluton, FZ – Fault Zone).

Tectonic domains

Deeper parts of our model (depths 4–24 km in Figs. 4.5 and 4.6, Fig. 4.11b) reflect the main geologic domains.

- The Moldanubian domain is characterized by a high-velocity anomaly, as it represents middle/lower crust denuded by erosion. The low-velocity uppermost layers are usually not present there (see also models of Růžek et al. (2007); Hrubcová et al. (2005); Majdański et al. (2006); Novotný (2012); Růžek et al. (2016)). Conversely, exceptionally high-velocity structures near the surface (see Fig. 4.11a and profile S04 in Figs. 4.9, 4.10) were detected in the same area also by Růžek et al. (2007) and Hrubcová and Środa (2015). Some of these anomalies spatially coincide with exposed high-grade rocks (see Fig. 4.1a). In the middle of the high velocity Moldanubian domain, a low-velocity body is present in our model (see also profiles CEL09 – distance 300 km, S04 – distance 300 km and S02 – distance 100 km in Figs. 4.9 and 4.10). Similar low-velocity zone down to 15 km was found on CEL09 profile by Novotný (2012), who interpreted it as an accretionary wedge on the contact of the Moldanubian and Moravian (i.e., Brunia) domains.

- The Teplá-Barrandian domain (partially hidden under the Bohemian Cretaceous Basin) is on the contrary to the Moldanubian characterized by a low-velocity anomaly corresponding to its composition of less consolidated upper crustal rocks. An exception is a high velocity anomaly extending northward from the Moldanubian domain in the middle of the Bohemian Cretaceous Basin. The structure is also displayed on profile S02 at the distance of 200km (Figs. 4.9 and 4.10) and evokes a promontory of the Moldanubian into the Teplá-Barrandian under the Bohemian Cretaceous Basin (see Fig. 4.11b). However, this is not confirmed by known geology (Uličný et al. (2009)).

- The Saxothuringian domain and the Sudetes show mainly low-velocity anomalies (see also profiles S01, S02 and S04 in Figs. 4.9 and 4.10) which reflect their similar composition and origin. Lower velocities on S02 were also obtained by Majdański et al. (2006).

- The Brunia domain in the north-eastern part shows a high velocity anomaly representing a well consolidated crystalline basement. Higher S-wave velocity values in this domain were also found by Růžek et al. (2016).

Plutonic bodies

In the Bohemian Massif, numerous granitic intrusions are present. Our inversion is not able to distinguish small-scale structures and thus we focus on large plutons and plutonic complexes only. Among the large plutons, the low-velocity anomaly corresponding to the Karlovy Vary Pluton may be found in our model in Fig. 4.11b and on profiles CEL09 and S01 in Fig. 4.9. The low-velocity anomaly corresponding to the Karlovy Vary Pluton is also present in models of Hrubcová et al. (2005) and Novotný et al. (2009). Moreover, Málek et al. (2004) derived 1D models for different units of the West-Bohemia region obtaining lower P-wave

velocities for the plutons compared to the crystalline units. Such results are also in agreement with laboratory measurements for granitic rocks in this area (see also Pros et al. (1998)).

Another large plutonic bodies that might be inferred by our inversion are the Central Bohemian Plutonic Complex and the Central Moldanubian Plutonic Complex. In the vicinity of the Central Bohemian Pluton a low-velocity anomaly is observed. This anomaly may, however, belong to the adjacent Teplá-Barrandian domain, where the lower velocities are typical. In the case of the Central Moldanubian Pluton, a low-velocity anomaly is present just below its northern tip. However, it does not continue below the southern part of the Pluton, which makes interpretation of this anomaly debatable.

All these three anomalies are clearly visible as low-velocity perturbations in Fig. 4.10 – profiles CEL09 and S01, where they extend down to ∼20 km depth. For plutonic bodies, such a large depth extent is unlikely as confirmed by other studies: i) gravity modeling by Guy et al. (2011) indicated that the Central Bohemian Plutonic Complex and Central Moldanubian Plutonic Complex reach to 5 and 10 km, respectively, and ii) the Karlovy Vary Pluton was identified as a low velocity anomaly extending down to 10 km on the CEL09 profile by Hrubcová et al. (2005) and Novotný (2012). The supposed overestimation of the depth extent of these anomalies in our model may be ascribed to the vertical smoothing effect of the surface waves.

Eger Rift

The most distinct structure in our model is in the area of the Eger Rift: the Saxothuringian – Teplá-Barrandian boundary fault active even nowadays. Under the sediments, the Eger Rift is characterized by a very strong high-velocity anomaly extending deep in the lower crust (see Fig. 4.11b; profiles S01 and S04 and the southern rim of the Eger Rift anomaly is also seen at distances 100–150 km of CEL09, Figs. 4.9 and 4.10). In this area, higher S-wave velocities down to lower crust were already found by Kolínský et al. (2011). High P-wave velocities were found by Alexandrakis et al. (2014); Mousavi et al. (2015). The S01 profile extends along the Eger Rift where the high velocity bodies were already found by Růžek et al. (2007) and by Grad et al. (2008). Grad et al. (2008) interpreted their two separate high-velocity bodies originating in the lower crust and continuing to the shallower depths as a result of the Saxothuringian subduction. The Moho updoming in this area was found by Heuer et al. (2006) and in the S04 model of Hrubcová and Środa (2015). However, our model, despite apparent reversal of the anomaly at greater depths (see bottomost map in Figs. 4.5 and 4.6), is less sensitive to the deep structures and is unable to contradict or confirm the continuation of the high-velocity anomaly to greater depths.

4.7 Conclusions

We have inverted ambient noise inter-station dispersion curves into the 3D S-wave velocity model of the Bohemian Massif. The traditional two-step approach is modified by employing the finite-frequency adjoint method in the first step, and

in the second step by performing Bayesian inversion by the parallel tempering algorithm for all model parameters simultaneously.

For the interpretation, we have considered the mean velocity model and its uncertainty which were calculated from all resulting 3D models sampling the posterior PDF (more than 1 million models). The mean model reveals the structures that are stable in the inversion. The estimated uncertainty helps us to assess reliability of the individual model features. As the main drawback of our model one may consider its lack of the more detailed structures for the geological interpretations. In addition, the surface geology is not perfectly related to the deep crustal structure that is imaged by surface waves inherently smoothing both laterally and vertically the true velocity structure.

The inferred 3D S-wave velocity model shows good correlation with main geologic domains of the Bohemian Massif. The Moldanubian domain is characterized by high S-wave velocities representing exposed middle/lower crustal material, except for a distinct anomaly located in its central part. Similarly to the Moldanubian, Brunia shows higher S-wave velocity anomalies. In contrast, the Teplá-Barrandian, Saxothuringian and Sudetes domains show lower S-wave velocities. In the Teplá-Barrandian domain, we have recovered a high velocity region with unclear geological interpretation resembling a promontory of the Moldanubian domain.

The most prominent high-velocity anomaly in the model is found beneath the Eger Rift. Moreover, some of the low S-wave velocity anomalies present in our model may correspond to large plutonic bodies, and the topmost low velocity structures correlate well with the sedimentary cover of the Bohemian Massif.

Acknowledgments

We would like to thank two anonymous reviewers for their helpful comments to improve the manuscript. This research has been supported by Charles University grant SVV 260447/2017. We acknowledge financial support through the GACR (Grantova Agentura Ceske Republiky) project no. 17-22207S. This work was supported by The Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project "IT4Innovations National Supercomputing Center – LM2015070". We would like to express our deepest gratitude to all that provided the data: CRSN network, all BOHEMA and PASSEQ experiments and other VEBSN stations. We thank Malcolm Sambridge for providing parallel tempering software (available from http://www.iearth.org.au) and Oldřich Novotný for VDISP code.

Chapter 5

Discussion

In this chapter we discuss some aspects of our results. Firstly, we compare the input data with the synthetics corresponding to the inferred models. Next we compare the inferred 1D model with other published S-wave velocity models, in particular the mean 1D model of the Bohemian Massif and a local model of Western Bohemia.

5.1 Dispersion curves and maps

We have calculated the synthetic phase velocity dispersion curves for the inverted 1D mean model of the Bohemian Massif (see Section 4.5.2). We display these synthetic phase velocity dispersion curves together with all the original inter-station dispersion curves obtained by processing the ambient noise (after selection for high SNR), which served as the input data for the adjoint inversion (see Fig. 5.1). Here we omit displaying synthetic dispersion curves corresponding to all model samples as they are very similar (see, e.g. Fig. 5.6). From Fig. 5.1 one can see that the inverted 1D mean model of the Bohemian Massif explains sufficiently the inter-station dispersion curves.

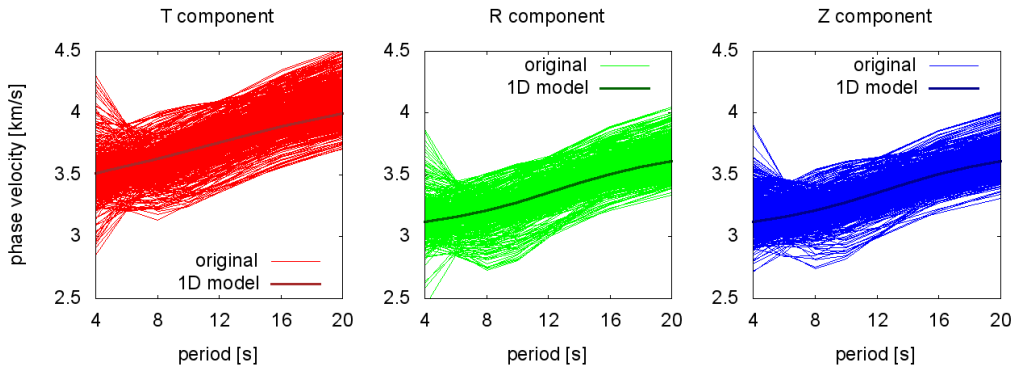

Figure 5.1: Inter-station dispersion curves serving as input data for the 2D adjoint localization for all components and corresponding synthetic dispersion curves calculated in the resulting mean 1D model of the Bohemian Massif (darker color).

Next, we compare the synthetic dispersion curves of the 1D model with dispersion curves obtained by the adjoint localization in all data points (Fig. 5.2a).

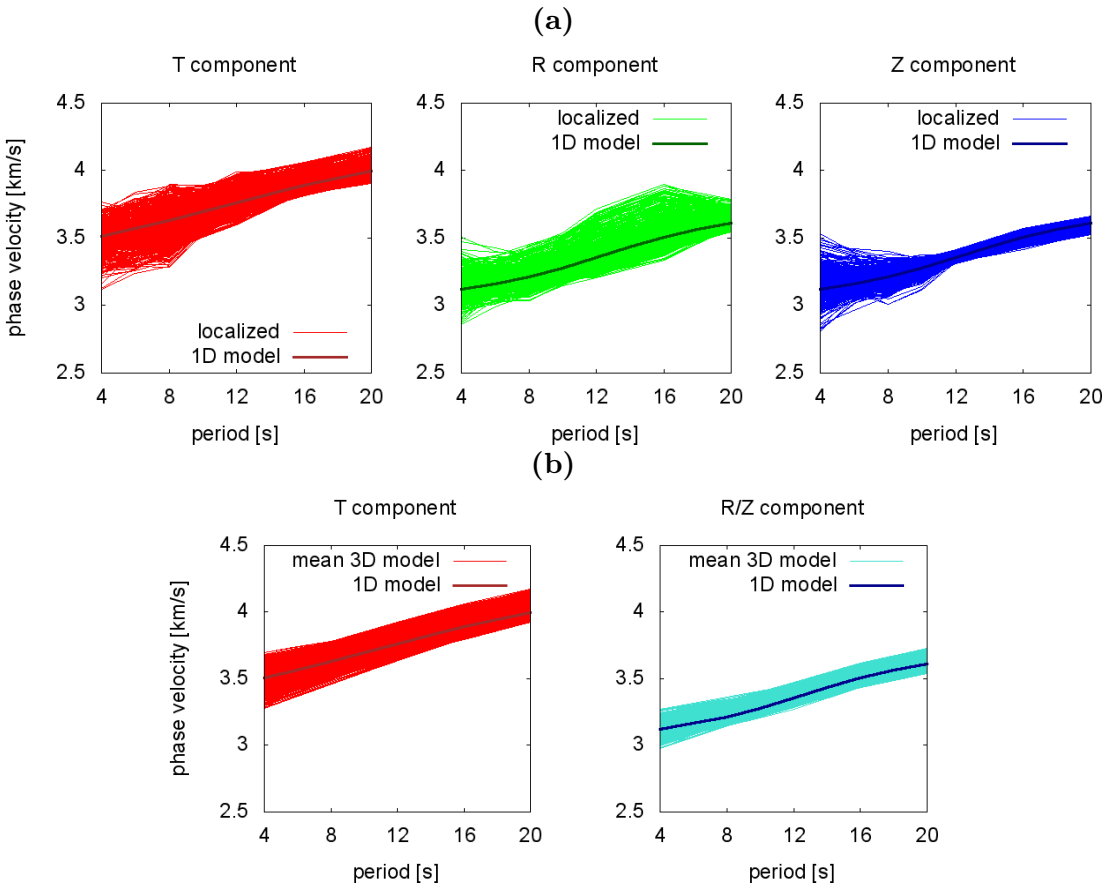

Figure 5.2: Synthetic dispersion curves for all components corresponding to the inverted mean 1D model of the Bohemian Massif (darker color) compared with: a) dispersion curves extracted from the phase velocity dispersion maps for all data points, b) synthetic dispersion curves calculated in all data points using the mean 3D model of the Bohemian Massif.

The variability of the localized data decreased mainly for the longer periods – this may be partially explained by the applied Gaussian smoothing which is frequency-dependendent. However, for the Z component the localized data are close even for period 12 s, which may be explained rather by homogeneous structure of the deeper parts. We also display the synthetic dispersion curves calculated in all data points from the mean 3D model (see Fig. 5.2b). They show smaller variability for both shorter and longer periods than the localized data suggesting that the smaller-scale shallow structures were not inverted in the 3D model.

To review the results of the second step of the 3D inversion, we calculate synthetic dispersion maps for the mean model and compare them with the real dispersion maps of the adjoint inversion. The results are displayed relatively with respect to the mean phase velocity (denoted on the right of each map) for each period for Love waves in Fig. 5.3 and for Rayleigh waves in Fig. 5.4. We also remind that the error assumed for the phase velocities of the dispersion map was 0.14km/s.

Similarly as seen for the 1D model, the mean values of the phase velocity dispersion are fitted very well for all components and all periods. Regarding the 2D perturbations in the phase velocities, most of the long-wavelength structures

present in the data are also present in the synthetic dispersion maps. One may also observe that the amplitudes of the perturbations in the synthetic maps do not always reflect the original (localized) amplitudes. However, as already mentioned in Chapter 2, the amplitudes of the perturbations are strongly dependent on the employed regularization in the adjoint localization (i.e., size of the smoothing Gaussian function and the number of iterations). Therefore, it should not be considered as drawback if the Bayesian inversion fails to invert the amplitudes of these phase velocity perturbations correctly.

The greatest drawback of our inversion may be considered its inability to fit the short-wavelength perturbations which are present in the short-period phase velocity data. Besides random errors, these are affected by the small-scale subsurface structures which are not present in our resulting model. To be on the safe side, we have chosen the model space parameterization, in particular the distance between the model control points, relatively large ($\approx$ 50km), thus not allowing to invert for such small-scale structures. However, when comparing the small-scale phase velocity perturbations for the two components of the Rayleigh waves, they were found to be less reliable than the large scale perturbations. Therefore, the inversion is regularized to prevent the artificial small-scale structures.

5.2 Comparison of 1D models

In the previous Chapter 4, 2D cross-sections through our model along selected profiles were compared with the corresponding 2D models of other authors. Here in addition, we compare the 1D S-wave velocity models, both local and that of the whole Bohemian Massif, with models of other authors. Although numerous tomographic studies were performed for the Bohemian Massif, they provide mainly P-wave velocity models while S-wave velocity models are rather sparse.

5.2.1 1D models of the Bohemian Massif

Firstly, we compare our 1D model of the Bohemian Massif obtained by the Bayesian inversion of the dispersion maps with the following models (see Fig. 5.5):

- DSS1988: an older model corresponding to the Bohemian Massif from the review paper on the deep seismic sounding (DSS) studies by Novotný and Urban (1988), obtained by Beránek and Tobyáš (1971). The original P-wave velocity model was recalculated into S-wave velocity model using traditional formula $v_s = v_p/\sqrt{3}$.

- Kolinsky2011: Kolínský et al. (2011) performed inversion of two-station phase velocity Love wave dispersion curves into 1D S-wave velocity model. The model represents average model along the profile between the stations and for the longest profile (200 km long) may be considered as a representation of the 1D model of the Bohemian Massif.

- Ruzek2012: Růžek et al. (2012) performed joint inversion of the teleseismic receiver functions with group velocity dispersion maps from the ambient noise cross-correlation into a 1D layered S-wave velocity models beneath

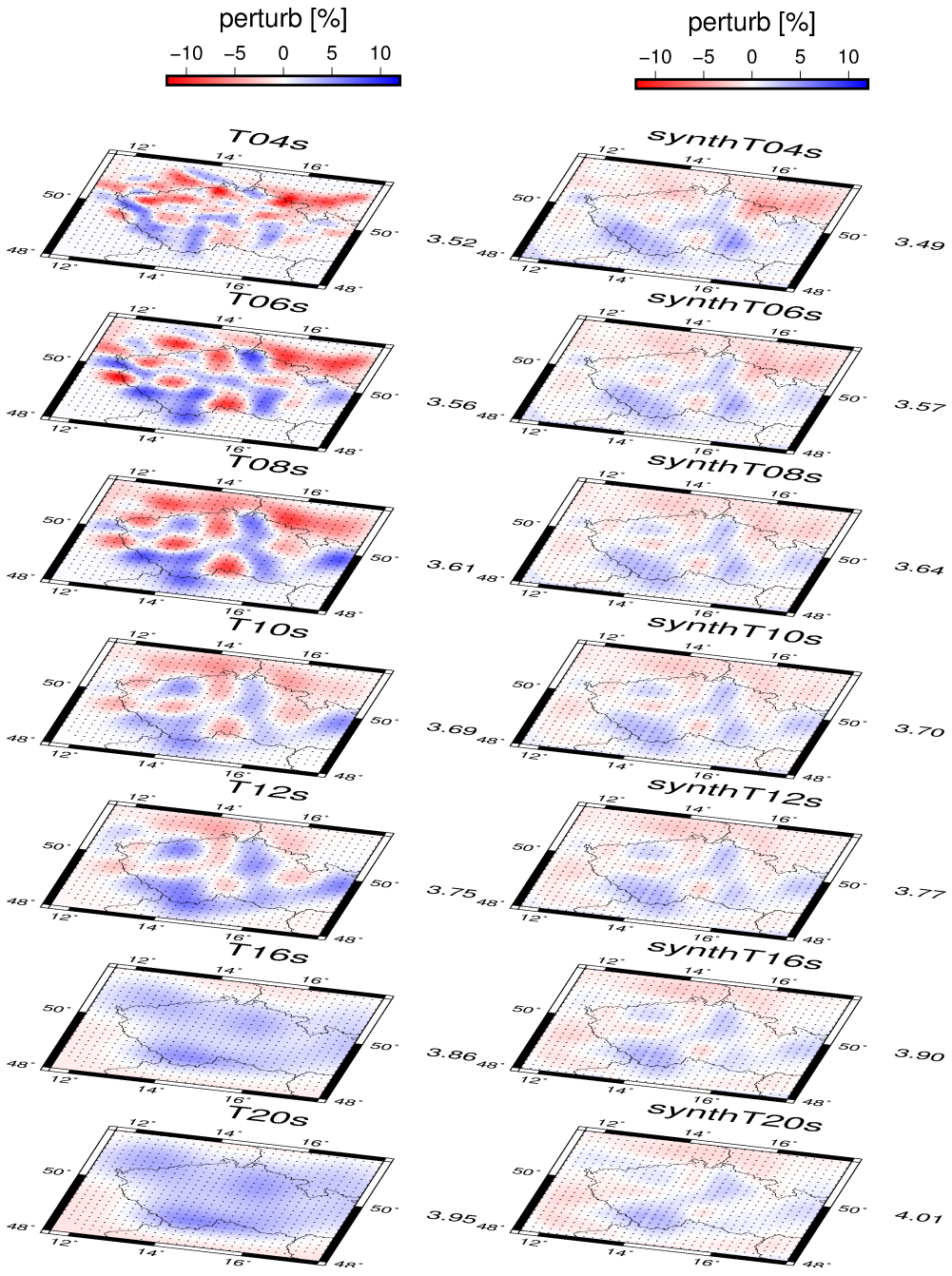

Figure 5.3: Phase velocity perturbations for each period for Love wave component, the mean value (in km/s) is shown on the right for each map. Left: phase velocity dispersion maps obtained by the 2D adjoint localization. Right: synthetic phase velocity dispersion maps calculated for the mean S-wave velocity model obtained by our 3D Bayesian inversion.

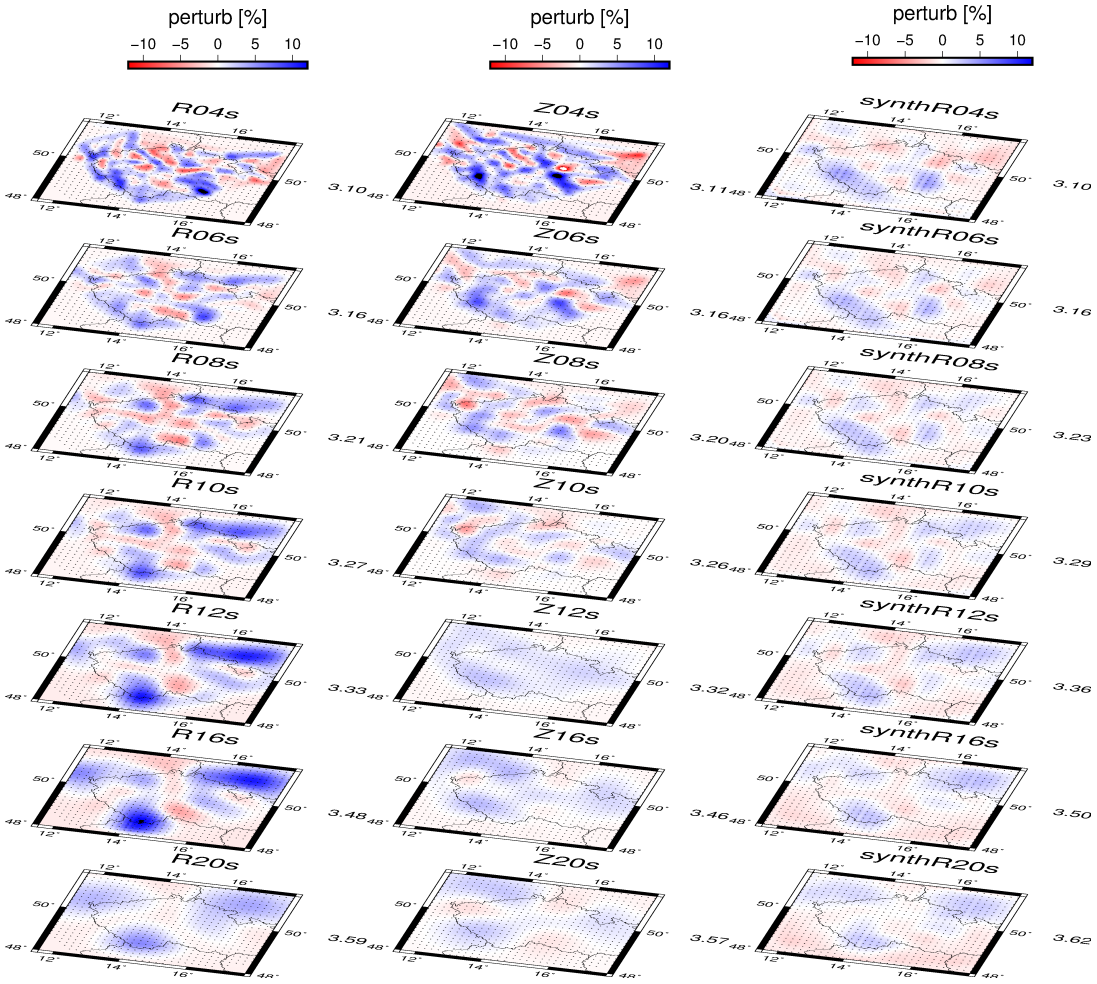

Figure 5.4: Phase velocity perturbations for each period for Rayleigh wave components, the mean value (in km/s) is shown on the right for each map. Left and middle: phase velocity dispersion maps obtained by the 2D adjoint localization for R and Z components, respectively. Right: synthetic phase velocity dispersion maps calculated for the mean S-wave velocity model obtained by our 3D Bayesian inversion.

the stations and v_p/v_s ratio. These models were averaged over all stations into 1D S-wave velocity model and v_p/v_s ratio.

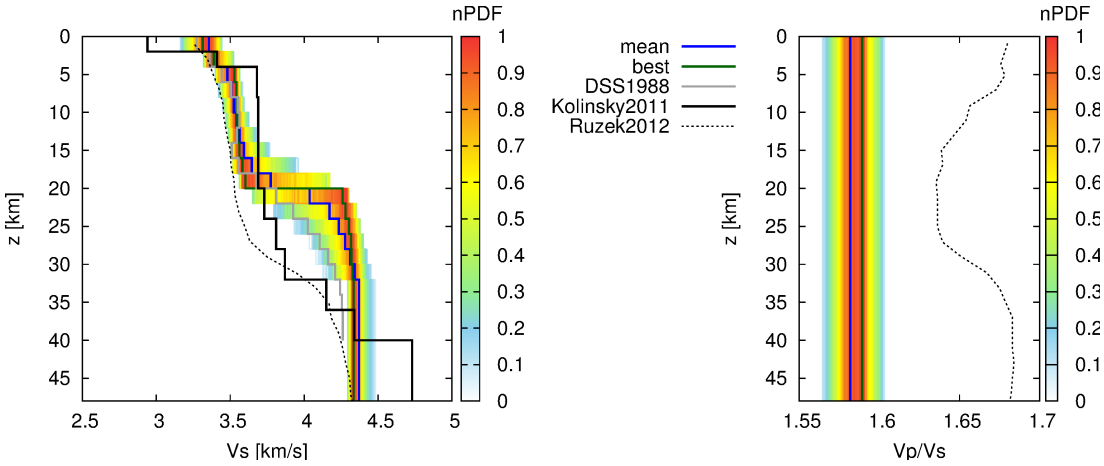

Figure 5.5: Result of the 1D Bayesian inversion (color coded): left – S-wave velocity model, right – v_p/v_s ratio. The resulting model samples are colored according to the normalized PDF value (nPDF), and the mean and best models are shown (see legend). The models are compared with 1D S-wave velocity models of the Bohemian Massif by other authors (for more details see text).

For the top 20km, all the models show rather homogeneous upper crust with low velocity gradient. The best correlation of our model is obtained for the DSS1988 model despite its advanced age and simplified relation between the P- and S-wave velocity. Model by Růžek et al. (2012) is systematically biased towards the lower values of our model, whereas model of Kolínský et al. (2011) shows very low subsurface velocities, which appear to be compensated by the higher velocities below 5 km leading to discrepancies between the models for the top 15 km. Our model lies between these two models. Also note, that Kolínský et al. (2011) used phase velocity dispersion curves in period range 10-40 s in their inversion.

Below the 20 km, our model prefers a sharp increase in the S-wave velocity (red-color models in Fig. 5.5) which is not present in other published models. All of the published models show increased gradient in these depths but keep lower values in the S-wave velocity. The gradient increase can be found also in part of our resulting models. It is also appropriate to account for the estimated errors of the models by other authors. Kolínský et al. (2011) indicate rms error of their model 0.136 km/s, Růžek et al. (2012) obtained standard deviation increasing with depth from ≈0.2 km/s for the top 20 km to ≈0.5 km/s below the 30 km. When considering the estimated errors of all the models, the disagreements between them may be explained.

As already mentioned, Růžek et al. (2012) inverted not only for S-wave velocity model but also for v_p/v_s ratio displayed for the comparison with our model in right panel of Fig. 5.5. Although they found a low value of v_p/v_s than usually assumed, we have obtained even smaller v_p/v_s even when including the estimated variance.

For all the models we have also calculated corresponding phase velocity dispersion curves (see Fig. 5.6). First of all, despite variability of our resulting 1D

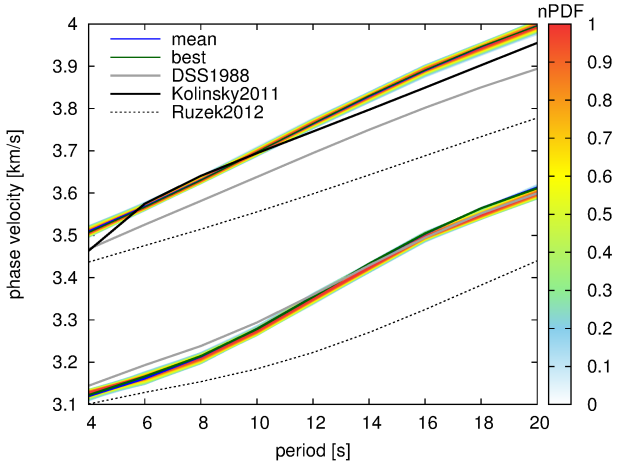

Figure 5.6: Synthetic dispersion curves of Love waves (upper) and Rayleigh waves (lower) corresponding to models of Fig. 5.5.

models (Fig. 5.5), the corresponding dispersion curves vary only slightly. This may explain why the inversion is unable to absolutely distinguish between the model with sharp velocity increase from the model with gradual velocity increase around 20 km depth. The dispersion curves of the DSS1988 model fit the Rayleigh dispersion curve almost perfectly but diverge in the Love dispersion curves, probably due to different v_p/v_s ratio. Note that DSS1988 model was derived using P-wave reflection and refraction traveltimes. The Love wave dispersion curve of model by Kolínský et al. (2011) fits the synthetic dispersion curve corresponding to our model well for periods 6-10 s and differs only slightly for other periods.

We would like to note that the disagreement of our 1D model of the Bohemian Massif with other published model is probably only apparent as it is related rather to how and which parts of the region were averaged. Therefore, we focus on models representing smaller area where better agreement between the models is expected.

5.2.2 Local 1D models of Western Bohemia

We have extracted 1D model in a specific point (namely in station Nový Kostel, NKC) from our 3D model using cubic spline interpolation. This station is located in Western Bohemia region with periodic occurrence of earthquake swarms. Therefore, this area was subject to many tomographic invesigations, results of which are compared with ours (see Fig. 5.7):

- Novotny1988: older DSS studies layered model of the Ore Mountains from Novotný and Urban (1988). We have calculated S-wave velocity model from the P-wave velocity using the formula $v_s = v_p/\sqrt{3}$.
- Novotny1996: the top 5 km is assembled from local models of the region, for the greater depths the model Novotny1988 is used (Novotný (1996)). The S-wave velocity was also recalculated from the P-wave velocity.
- Malek2000: Málek et al. (2000) inverted local earthquake data into 1D layered P- and S-wave velocity model individually for different subregions of the Western Bohemia as well as the whole Western Bohemian region.

- Ruzek2013: local earthquake and experimental data were used by Růžek and Horálek (2013) to obtain a 3D P- and S-wave velocity model and calculated their average along horizontal planes. We should keep in mind, that our 3D model is rather averaged (and thus smoothed) model also when considering a single point. Therefore, we compare our extracted 1D model with the models of Málek et al. (2000) and Růžek and Horálek (2013) corresponding to whole Western Bohemia.

- Malek2005iso: Málek et al. (2005) used both local earthquakes as well as controlled shot data to infer 1D isotropic and anisotropic model of the Western Bohemia. Here, we compare our model with its isotropic part.

- WildePiorko2005: the S-wave velocity model was derived using teleseismic receiver function beneath the station NKC (Wilde-Piórko et al. (2005)).

- Ruzek2012: the model was obtained from joint inversion of teleseismic reciever function and group dispersion data beneath the station NKC (Růžek et al. (2012)).

- Kolinsky2011: model obtained by Kolínský et al. (2011) for short profile NEC-NKC crossing the Eger Rift.

In general, all models are very close to our extracted 1D model mainly at depths 2-25 km. A larger discrepancy is in the subsurface part between our model and the shallow models (models reaching only to 10-15 km, i.e. Malek2000, Malek2005iso and RuzekHoralek2013, see also zoomed inset in the middle of Fig. 5.7), which is probably caused by low sensitivity and resolution of our data to these depths. Model by Wilde-Piórko et al. (2005) deviates from all other models in top 15 km, but its lower part (20-40 km) agrees with them.

Where available, we also compared the value of v_p/v_s (see Fig. 5.7 right). We have obtained generally smaller value of the v_p/v_s ratio when compared with other models, although values of Málek et al. (2000) and Růžek et al. (2012) lie at the upper bound of the permissible values of our model.

To sum up, we have compared our inverted 1D S-wave velocity model with other models of the Bohemian Massif. Moreover, we have compared the 1D S-wave velocity model extracted from the 3D model in Western Bohemia with other available S-wave velocity models of the particular region. Except for the shallowest subsurface structure (<1 km), we have found good agreement between our model and most of the models which were derived using different methods and input data. The comparison between the local models supports our results much better than the 1D model of the Bohemian Massif, averaged over different geologic domains. Therefore, we claim that our model sufficiently represents the S-wave velocity structure of the Bohemian Massif as well as its smaller parts.

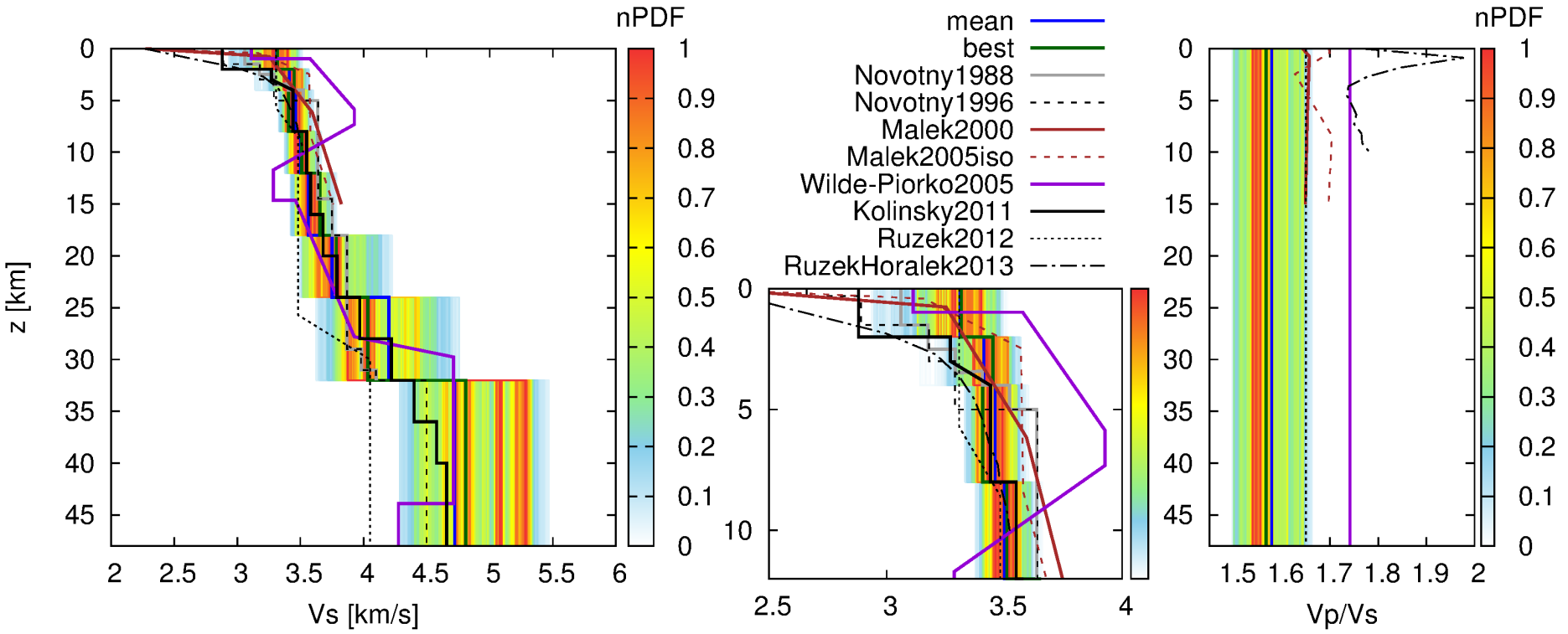

Figure 5.7: 1D model extracted from our 3D Bayesian model at station NKC (color coded): left – S-wave velocity model, right – v_p/v_s ratio. The resulting model samples are colored according to the normalized PDF value (nPDF), and the mean and best models are shown (see legend). The model is compared with 1D S-wave velocity models of Western Bohemia/NKC station of other authors (for more details see text). The inset in the middle is a zoom of the top 12km.

Conclusions

We have performed the first 3D ambient noise tomography of the Bohemian Massif. The input data comprise of inter-station phase-velocity dispersion curves in range 4–20s prepared from ambient noise processing by Růžek et al. (2016). The inversion was carried out using two-step approach: in the first step, the dispersion curves were localized for each period into phase velocity dispersion maps and in the second step, these maps were inverted into a 3D S-wave velocity model.

To account for the finite-frequency effect, 2D adjoint inversion was utilized in the first step. Particular attention has been given to estimate the regularization parameters, such as width of the gradient smoothing function and number of iterations – subject not properly covered by literature despite having a crucial effect on the resulting model. To this aim, synthetic tests for the longest periods with two distinct target models were used, the first model was simple smooth model to demonstrate the emergence of model artifacts when insufficient smoothing is applied or too many iterations are performed. The second target model contains pronounced smaller-scale heterogeneities revealing the possibility to recover more-detailed structures in the resulting model. The synthetic data were modified by noise estimated from the differences between the 16s and 20s Love group data. According to the tests, relatively large smoothing width (almost twice the wavelength) and rather small number of iterations (≈ 5) were found to be optimal. These values may appear to give too simple or over-regularized models. On the other hand, the resulting models contain weak but apparent smaller-scale heterogeneities which are, moreover, considered reliable assuming the estimated noise level.

The phase velocity dispersion maps were determined with the regularization parameters derived from the synthetic tests with the smoothing width scaled according to the respective wavelength. These maps serve as input data in the Bayesian inversion to obtain the 3D S-wave velocity model of the Bohemian Massif (and additionally depth-independent v_p/v_s ratio). The inversion was carried out employing Monte-Carlo sampler (in particular the so-called parallel tempering algorithm), combined with calculation of the dispersion curves in the layered model using matrix method. The main advantage of the Bayesian inversion is that the solution is represented not only by one 'best' model but by a large set of models, which are used to determine a mean model together with the uncertainty of the model parameters. As a result, more than one million model samples (i.e., 3D velocity models) following the posterior probability density function were obtained. From the estimated standard deviation we find that the best resolved part lies at depths 2-18 km ruling out determination of the Moho depth.

The mean model was correlated with known geologic structures of the Bohemian Massif. We must keep in mind that our method is able to recover only large-scale (smoothed) structures. The most prominent structure in the resulting model is a high velocity anomaly located beneath the Eger Rift, presently the most active area of the Bohemian Massif. The individual domains of the Bohemian Massif show higher velocity anomalies where high-grade metamorphism is expected (Moldanubian, Brunia), whereas other domains are characterized by rather low-velocity anomalies (Teplá-Barrandian, Saxothuringian and Sudetes). The uppermost part of the 3D velocity model shows good correlation between the low-velocity anomalies and sedimentary cover of the Bohemian Massif. Moreover, two significant anomalies with unknown geological interpretation were recovered: a low velocity anomaly in the Moldanubian domain and a high-velocity anomaly resembling a promontory of the Moldanubian into the Teplá-Barrandian domain under the Bohemian Cretaceous Basin.

Finally, our 1D S-wave velocity model of the Western Bohemia was compared with local models found by other authors, showing in general very good agreement in depths 2-18 km (i.e., where low variance of our model is obtained). This supports the results of our ambient noise tomography of the Bohemian Massif's crust as whole, which may give additional insight into its structure and development.

Bibliography

Czech regional seismological network, https://www.ig.cas.cz/en/structure/observatories/czech-regional-seismological-network.

Aki, K., and P. Richards, *Quantitative Seismology*. University Science Books, 2002.

Aki, K., A. Christoffersson, and E. S. Husebye, Determination of the three-dimensional seismic structure of the lithosphere, *Journal of Geophysical Research*, *82* **(2)**, 277–296, 1977.

Aki, K., and W. H. K. Lee, Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model, *Journal of Geophysical Research*, *81* **(23)**, 4381–4399, 1976.

Alexandrakis, C., M. Calò, F. Bouchaala, and V. Vavryčuk, Velocity structure and the role of fluids in the West Bohemia Seismic Zone, *Solid Earth*, *5* **(2)**, 863, 2014.

Babuška, V., J. Plomerová, L. Vecsey, P. Jedlička, and B. Růžek, Ongoing passive seismic experiments unravel deep lithosphere structure of the Bohemian Massif, *Studia Geophysica et Geodaetica*, *49* **(3)**, 423–430, 2005.

Babuška, V., and J. Plomerová, Boundaries of mantle–lithosphere domains in the Bohemian Massif as extinct exhumation channels for high-pressure rocks, *Gondwana Research*, *23* **(3)**, 973 – 987, 2013.

Badal, J., Y. Chen, M. Chourak, and J. Stankiewicz, S-wave velocity images of the Dead Sea Basin provided by ambient seismic noise, *Journal of Asian Earth Sciences*, *75*, 26–35, 2013.

Barmin, M. P., M. H. Ritzwoller, and A. L. Levshin, A Fast and Reliable Method for Surface Wave Tomography, *Pure and Applied Geophysics*, *158* **(8)**, 1351–1375, 2001.

Beard, B. L., L. G. Medaris, C. M. Johnson, E. Jelínek, J. Tonika, and L. R. Riciputi, Geochronology and geochemistry of eclogites from the Mariánské Lázně Complex, Czech Republic: Implications for Variscan orogenesis, *Geologische Rundschau*, *84* **(3)**, 552–567, 1995.

Behm, M., 3-D modelling of the crustal S-wave velocity structure from active source data: application to the Eastern Alps and the Bohemian Massif, *Geophysical Journal International*, *179* **(1)**, 265, 2009.

Bensen, G. D., M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Yang, Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International*, *169* **(3)**, 1239–1260, 2007.

Bensen, G. D., M. H. Ritzwoller, and N. M. Shapiro, Broadband ambient noise surface wave tomography across the United States, *Journal of Geophysical Research: Solid Earth*, *113* **(B5)**, 2008.

Beránek, B., and V. Tobyáš, Study of the velocity conditions in the Earth's crust in the regions of the Bohemian massif and the Carpathian system along International Profiles VI and VII, *Studia Geophysica et Geodaetica*, *15* **(3)**, 316–330, Sep, 1971.

Bodin, T., M. Sambridge, H. Tkalčić, P. Arroucau, K. Gallagher, and N. Rawlinson, Transdimensional inversion of receiver functions and surface wave dispersion, *Journal of Geophysical Research: Solid Earth*, *117* **(B2)**, 2012, B02301.

Boschi, L., Global multiresolution models of surface wave propagation: comparing equivalently regularized Born and ray theoretical solutions, *Geophysical Journal International*, *167* **(1)**, 238–252, 2006.

Boué, P., P. Poli, M. Campillo, H. Pedersen, X. Briand, and P. Roux, Teleseismic correlations of ambient seismic noise for deep global imaging of the Earth, *Geophysical Journal International*, *194* **(2)**, 844–848, 2013.

Bozdağ, E., J. Trampert, and J. Tromp, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophysical Journal International*, *185* **(2)**, 845–870, 2011.

Brenguier, F., N. M. Shapiro, M. Campillo, A. Nercessian, and V. Ferrazzini, 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations, *Geophysical Research Letters*, *34* **(2)**, 2007, L02305.

Campillo, M., and A. Paul, Long-Range Correlations in the Diffuse Seismic Coda, *Science*, *299* **(5606)**, 547–549, 2003.

Čermák, V., Results of heat flow studies in Czechoslovakia, in *Crustal Structure of the Bohemian Massif and the West Carpathians*, edited by V. Bucha, and M. Blížkovský, 85–118, Springer, Berlin, 1994.

Chen, M., H. Huang, H. Yao, R. van der Hilst, and F. Niu, Low wave speed zones in the crust beneath SE Tibet revealed by ambient noise adjoint tomography, *Geophysical Research Letters*, *41* **(2)**, 334–340, 2014.

Chen, P., L. Zhao, and T. H. Jordan, Full 3D Tomography for the Crustal Structure of the Los Angeles Region, *Bulletin of the Seismological Society of America*, *97* **(4)**, 1094–1120, 2007*a*.

Chen, P., T. H. Jordan, and L. Zhao, Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods, *Geophysical Journal International*, *170* **(1)**, 175–181, 2007*b*.

Chen, P., T. H. Jordan, and E.-J. Lee, Perturbation kernels for generalized seismological data functionals (GSDF), *Geophysical Journal International*, *183* **(2)**, 869–883, 2010.

Cho, K. H., R. B. Herrmann, C. J. Ammon, and K. Lee, Imaging the Upper Crust of the Korean Peninsula by Surface-Wave Tomography, *Bulletin of the Seismological Society of America*, *97* **(1B)**, 198–207, 2007.

Cloetingh, S., J. vanWees, P. Ziegler, L. Lenkey, F. Beekman, M. Tesauro, A. Förster, B. Norden, M. Kaban, N. Hardebol, D. Bonté, A. Genter, L. Guillou-Frottier, M. T. Voorde, D. Sokoutis, E. Willingshofer, T. Cornu, and G. Worum, Lithosphere tectonics and thermo-mechanical properties: An integrated modelling approach for Enhanced Geothermal Systems exploration in Europe, *Earth-Science Reviews*, *102* **(3–4)**, 159 – 206, 2010.

Colli, L., A. Fichtner, and H.-P. Bunge, Full waveform tomography of the upper mantle in the South Atlantic region: Imaging a westward fluxing shallow asthenosphere?, *Tectonophysics*, *604*, 26–40, 2013.

Dahlen, F. A., S.-H. Hung, and G. Nolet, Fréchet kernels for finite-frequency traveltimes–I. Theory, *Geophysical Journal International*, *141* **(1)**, 157–174, 2000.

Dèzes, P., S. Schmid, and P. Ziegler, Evolution of the European Cenozoic Rift System: interaction of the Alpine and Pyrenean orogens with their foreland lithosphere, *Tectonophysics*, *389* **(1–2)**, 1 – 33, 2004.

Dörr, W., and G. Zulauf, Elevator tectonics and orogenic collapse of a Tibetan-style plateau in the European Variscides: the role of the Bohemian shear zone, *International Journal of Earth Sciences*, *99* **(2)**, 299–325, 2010.

Drost, K., U. Linnemann, N. McNaughton, O. Fatka, P. Kraft, M. Gehmlich, C. Tonk, and J. Marek, New data on the Neoproterozoic – Cambrian geotectonic setting of the Teplá-Barrandian volcano-sedimentary successions: geochemistry, U-Pb zircon ages, and provenance (Bohemian Massif, Czech Republic), *International Journal of Earth Sciences*, *93* **(5)**, 742–757, 2004.

Dumbser, M., M. Käser, and E. F. Toro, An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes - V. Local time stepping and p-adaptivity, *Geophysical Journal International*, *171* **(2)**, 695–717, 2007.

Dumbser, M., and M. Käser, An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes – II. The three-dimensional isotropic case, *Geophysical Journal International*, *167* **(1)**, 319–336, 2006.

Dziewonski, A. M., B. H. Hager, and R. J. O'Connell, Large-scale heterogeneities in the lower mantle, *Journal of Geophysical Research*, *82* **(2)**, 239–255, 1977.

Earl, D. J., and M. W. Deem, Parallel tempering: Theory, applications, and new perspectives, *Physical Chemistry Chemical Physics*, *7*, 3910–3916, 2005.

Fichtner, A., H.-P. Bunge, and H. Igel, The adjoint method in seismology: I. Theory, *Physics of the Earth and Planetary Interiors*, *157* **(1)**, 86–104, 2006.

Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge, Theoretical background for continental- and global-scale full-waveform inversion in the time–frequency domain, *Geophysical Journal International*, *175* **(2)**, 665–685, 2008.

Fichtner, A., B. L. Kennett, H. Igel, and H.-P. Bunge, Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophysical Journal International*, *179* **(3)**, 1703–1725, 2009.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. Villaseñor, Multiscale full waveform inversion, *Geophysical Journal International*, , ggt118, 2013.

Finger, F., A. Gerdes, M. René, and G. Riegler, The Saxo-Danubian Granite Belt: magmatic response to post-collisional delamination of mantle lithosphere below the southwestern sector of the Bohemian Massif (Variscan orogen), *Geologica Carpathica*, *60* **(3)**, 205–212, 2009.

Fischer, T., J. Horálek, P. Hrubcová, V. Vavryčuk, K. Bräuer, and H. Kämpf, Intra-continental earthquake swarms in West-Bohemia and Vogtland: A review, *Tectonophysics*, *611* **(C)**, 1 – 27, 2014.

Franke, W., The mid-European segment of the Variscides: tectonostratigraphic units, terrane boundaries and plate tectonic evolution, *Geological Society, London, Special Publications*, *179* **(1)**, 35–61, 2000.

Galetti, E., A. Curtis, G. A. Meles, and B. Baptie, Uncertainty Loops in Travel-Time Tomography from Nonlinear Wave Physics, *Phys. Rev. Lett.*, *114*, 148501, Apr, 2015.

Galetti, E., A. Curtis, B. Baptie, D. Jenkins, and H. Nicolson, Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry, *Geophysical Journal International*, *208* **(1)**, 36, 2017.

Gao, H., E. D. Humphreys, H. Yao, and R. D. van der Hilst, Crust and lithosphere structure of the northwestern U.S. with ambient noise tomography: Terrane accretion and Cascade arc development, *Earth and Planetary Science Letters*, *304* **(1–2)**, 202 – 211, 2011.

Gao, H., and Y. Shen, Upper mantle structure of the Cascades from full-wave ambient noise tomography: Evidence for 3D mantle upwelling in the back-arc, *Earth and Planetary Science Letters*, *390* **(0)**, 222 – 233, 2014.

Gauthier, O., J. Virieux, and A. Tarantola, Two-dimensional nonlinear inversion of seismic waveforms; numerical results, *Geophysics*, *51* **(7)**, 1387–1403, 1986.

Geissler, W. H., H. Kämpf, R. Kind, K. Bräuer, K. Klinge, T. Plenefisch, J. Horálek, J. Zedník, and V. Nehybka, Seismic structure and location of a CO2 source in the upper mantle of the western Eger (Ohře) Rift, central Europe, *Tectonics*, *24* **(5)**, 2005, TC5001.

Geissler, W. H., H. Kämpf, Z. Skácelová, J. Plomerová, V. Babuška, and R. Kind, Lithosphere structure of the NE Bohemian Massif (Sudetes) – A teleseismic receiver function study, *Tectonophysics*, *564–565*, 12 – 37, 2012.

Gouédard, P., L. Stehly, F. Brenguier, M. Campillo, Y. Colin deVerdière, E. Larose, L. Margerin, P. Roux, F. J. Sánchez-Sesma, N. M. Shapiro, and R. L. Weaver, Cross-correlation of random fields: mathematical approach and applications, *Geophysical Prospecting*, *56* **(3)**, 375–393, 2008.

Grad, M., A. Guterch, S. Mazur, G. R. Keller, A. Špičák, P. Hrubcová, and W. H. Geissler, Lithospheric structure of the Bohemian Massif and adjacent Variscan belt in central Europe based on profile S01 from the SUDETES 2003 experiment, *Journal of Geophysical Research: Solid Earth*, *113* **(B10)**, 2008, B10304.

Guo, Z., Y. J. Chen, J. Ning, Y. Feng, S. P. Grand, F. Niu, H. Kawakatsu, S. Tanaka, M. Obayashi, and J. Ni, High resolution 3-D crustal structure beneath NE China from joint inversion of ambient noise and receiver functions using NECESSArray data, *Earth and Planetary Science Letters*, *416*, 1–11, 2015.

Guo, Z., Y. J. Chen, J. Ning, Y. Yang, J. C. Afonso, and Y. Tang, Seismic evidence of on-going sublithosphere upper mantle convection for intra-plate volcanism in Northeast China, *Earth and Planetary Science Letters*, *433*, 31 – 43, 2016.

Guo, Z., X. Gao, W. Wang, and Z. Yao, Upper- and mid-crustal radial anisotropy beneath the central Himalaya and southern Tibet from seismic ambient noise tomography, *Geophysical Journal International*, *189* **(2)**, 1169–1182, 2012.

Gutenberg, B., and C. Richter, On seismic waves, *Gerlands Beiträge zur Geophysik*, *54*, 94–136, 1939.

Guy, A., J.-B. Edel, K. Schulmann, Č. Tomek, and O. Lexa, A geophysical model of the Variscan orogenic root (Bohemian Massif): implications for modern collisional orogens, *Lithos*, *124* **(1–2)**, 144–157, 2011.

Haned, A., E. Stutzmann, M. Schimmel, S. Kiselev, A. Davaille, and A. Yelles-Chaouche, Global tomography using seismic hum, *Geophysical Journal International*, *204* **(2)**, 1222, 2015.

Haskell, N. A., The dispersion of surface waves on multilayered media, *Bulletin of the Seismological Society of America*, *43* **(1)**, 17–34, 1953.

Heuer, B., W. Geissler, R. Kind, and H. Kämpf, Seismic evidence for asthenospheric updoming beneath the western Bohemian Massif, central Europe, *Geophysical research letters*, *33* **(5)**, 2006, L05311.

van der Hilst, R., S. Widiyantoro, and E. Engdahl, Evidence for deep mantle circulation from global tomography, *Nature*, *386*, 578–584, 1997.

Hrubcová, P., P. Środa, A. Špičák, A. Guterch, M. Grad, G. Keller, E. Brueckl, and H. Thybo, Crustal and uppermost mantle structure of the Bohemian Massif based on CELEBRATION 2000 data, *Journal of Geophysical Research: Solid Earth*, *110* **(B11)**, 2005, B11305.

Hrubcová, P., P. Środa, C. . W. Group et al., Crustal structure at the easternmost termination of the Variscan belt based on CELEBRATION 2000 and ALP 2002 data, *Tectonophysics*, *460* **(1)**, 55–75, 2008.

Hrubcová, P., P. Środa, M. Grad, W. Geissler, A. Guterch, J. Vozár, E. Hegedűs, and S. . W. Group, From the Variscan to the Alpine Orogeny: crustal structure of the Bohemian Massif and the Western Carpathians in the light of the SUDETES 2003 seismic data, *Geophysical Journal International*, *183* **(2)**, 611–633, 2010.

Hrubcová, P., and P. Środa, Complex local Moho topography in the Western Carpathians: Indication of the ALCAPA and the European Plate contact, *Tectonophysics*, *638*, 63–81, 2015.

Hung, S.-H., F. Dahlen, and G. Nolet, Fréchet kernels for finite-frequency traveltimes–II. Examples, *Geophysical Journal International*, *141* **(1)**, 175–203, 2000.

Janoušek, V., D. Bowes, G. Rogers, C. M. Farrow, and E. Jelínek, Modelling Diverse Processes in the Petrogenesis of a Composite Batholith: the Central Bohemian Pluton, Central European Hercynides, *Journal of Petrology*, *41* **(4)**, 511, 2000.

Jeffreys, H., and K. Bullen, *Seismological tables*. Office of the British Association for the Advancement of Science, 1940.

Jiang, C., Y. Yang, and Y. Zheng, Penetration of mid-crustal low velocity zone across the Kunlun Fault in the NE Tibetan Plateau revealed by ambient noise tomography, *Earth and Planetary Science Letters*, *406*, 81–92, 2014.

Jiang, C., Y. Yang, N. Rawlinson, and W. L. Griffin, Crustal structure of the Newer Volcanics Province, SE Australia, from ambient noise tomography, *Tectonophysics*, *683*, 382–392, 2016.

Kalvoda, J., O. Babek, O. Fatka, J. Leichmann, R. Melichar, S. Nehyba, and P. Spacek, Brunovistulian terrane (Bohemian Massif, Central Europe) from late Proterozoic to late Paleozoic: a review, *International Journal of Earth Sciences*, *97* **(3)**, 497–518, 2008.

Karousová, H., J. Plomerová, and V. Babuška, Three-dimensional velocity model of the crust of the Bohemian Massif and its effects on seismic tomography of the upper mantle, *Studia Geophysica et Geodaetica*, *56* **(1)**, 249–267, 2012.

Käser, M., and M. Dumbser, An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes – I. The two-dimensional isotropic case with external source terms, *Geophysical Journal International*, *166* **(2)**, 855–877, 2006.

Käser, M., M. Dumbser, J. De La Puente, and H. Igel, An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes – III. Viscoelastic attenuation, *Geophysical Journal International*, *168* **(1)**, 224–242, 2007.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, Optimization by Simulated Annealing, *Science*, *220* **(4598)**, 671–680, 1983.

Kolínský, P., J. Málek, and J. Brokešová, Shear wave crustal velocity model of the western Bohemian Massif from Love wave phase velocity dispersion, *Journal of seismology*, *15* **(1)**, 81–104, 2011.

Kristeková, M., J. Kristek, and P. Moczo, Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophysical Journal International*, *178* **(2)**, 813–825, 2009.

Lévêque, J.-J., L. Rivera, and G. Wittlinger, On the use of the checker-board test to assess the resolution of tomographic inversions, *Geophysical Journal International*, *115* **(1)**, 313–318, 1993.

Li, H., W. Su, C.-Y. Wang, Z. Huang, and Z. Lv, Ambient noise Love wave tomography in the eastern margin of the Tibetan plateau, *Tectonophysics*, *491* **(1–4)**, 194 – 204, 2010, Great 12 May 2008 Wenchuan Earthquake (Mw7.9), China.

Lin, F.-C., M. H. Ritzwoller, J. Townend, S. Bannister, and M. K. Savage, Ambient noise Rayleigh wave tomography of New Zealand, *Geophysical Journal International*, *170* **(2)**, 649–666, 2007.

Lin, F.-C., M. P. Moschetti, and M. H. Ritzwoller, Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophysical Journal International*, *173* **(1)**, 281, 2008.

Lin, F.-C., V. C. Tsai, B. Schmandt, Z. Duputel, and Z. Zhan, Extracting seismic core phases with array interferometry, *Geophysical Research Letters*, *40* **(6)**, 1049–1053, 2013*a*.

Lin, F.-C., D. Li, R. W. Clayton, and D. Hollis, High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic array, *Geophysics*, *78* **(4)**, Q45–Q56, 2013*b*.

Lin, F.-C., and V. C. Tsai, Seismic interferometry with antipodal station pairs, *Geophysical Research Letters*, *40* **(17)**, 4609–4613, 2013.

Liu, Q., and J. Tromp, Finite-Frequency Kernels Based on Adjoint Methods, *Bulletin of the Seismological Society of America*, *96* **(6)**, 2383–2397, 2006.

Lomax, A., and R. Snieder, The contrast in upper mantle shear-wave velocity between the East European Platform and tectonic Europe obtained with genetic algorithm inversion of Rayleigh-wave group dispersion, *Geophysical Journal International*, *123* **(1)**, 169–182, 1995.

Luo, Y., and G. Schuster, Wave-equation traveltime inversion, *Geophysics*, *56* **(5)**, 645–653, 1991.

Luo, Y., Y. Xu, and Y. Yang, Crustal structure beneath the Dabie orogenic belt from ambient noise tomography, *Earth and Planetary Science Letters*, *313–314*, 12 – 22, 2012.

Majdański, M., M. Grad, A. Guterch, S. . W. Group et al., 2-D seismic tomographic and ray tracing modelling of the crustal structure across the Sudetes Mountains basing on SUDETES 2003 experiment data, *Tectonophysics*, *413* **(3)**, 249–269, 2006.

Majdański, M., E. Kozlovskaya, M. Grad, and S. . W. Group, 3D structure of the Earth's crust beneath the northern part of the Bohemian Massif, *Tectonophysics*, *437* **(1–4)**, 17 – 36, 2007.

Málek, J., J. Janský, O. Novotný, and D. Rössler, Vertically Inhomogeneous Models of the Upper Crustal Structure in the West-Bohemian Seismoactive Region Inferred from the Celebration 2000 Refraction Data, *Studia Geophysica et Geodaetica*, *48* **(4)**, 709–730, Oct, 2004.

Málek, J., J. Horálek, and J. Janský, One-Dimensional qP-Wave Velocity Model of the Upper Crust for the West Bohemia/Vogtland Earthquake Swarm Region, *Studia Geophysica et Geodaetica*, *49* **(4)**, 501–524, Oct, 2005.

Málek, J., J. Janský, and J. Horálek, Layered Velocity Models of the Western Bohemia Region, *Studia Geophysica et Geodaetica*, *44* **(4)**, 475–490, Oct, 2000.

Marquering, H., F. Dahlen, and G. Nolet, Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophysical Journal International*, *137* **(3)**, 805–815, 1999.

Matos, C., G. Silveira, L. Matias, R. Caldeira, M. L. Ribeiro, N. A. Dias, F. Krüger, and T. B. dosSantos, Upper crustal structure of Madeira Island revealed from ambient noise tomography, *Journal of Volcanology and Geothermal Research*, *298*, 136 – 145, 2015.

Matte, P., The Variscan collage and orogeny (480–290 Ma) and the tectonic definition of the Armorica microplate: a review, *Terra Nova*, *13* **(2)**, 122–128, 2001.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, *21* **(6)**, 1087–1092, 1953.

Montagner, J.-P., Regional three-dimensional structures using long-period surface waves, *Annales Geophysicae*, *4* **(B3)**, 283–294, 03, 1986.

Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung, Finite-Frequency Tomography Reveals a Variety of Plumes in the Mantle, *Science*, *303* **(5656)**, 338–343, 2004.

Mordret, A., M. Landès, N. Shapiro, S. Singh, and P. Roux, Ambient noise surface wave tomography to determine the shallow shear velocity structure at Valhall: depth inversion with a Neighbourhood Algorithm, *Geophysical Journal International*, *198* **(3)**, 1514, 2014.

Moschetti, M. P., M. H. Ritzwoller, and N. M. Shapiro, Surface wave tomography of the western United States from ambient seismic noise: Rayleigh wave group velocity maps, *Geochemistry, Geophysics, Geosystems*, *8* **(8)**, Q08010, 2007.

Mosegaard, K., and A. Tarantola, Monte carlo sampling of solutions to inverse problems, *Journal of Geophysical Research: Solid Earth*, *100* **(B7)**, 12431–12447, 1995.

Mousavi, S., K. Bauer, M. Korn, and B. Hejrani, Seismic tomography reveals a mid-crustal intrusive body, fluid pathways and their relation to the earthquake swarms in West Bohemia/Vogtland, *Geophysical Journal International*, *203* **(2)**, 1113, 2015.

Nakata, N., J. P. Chang, J. F. Lawrence, and P. Boué, Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *Journal of Geophysical Research: Solid Earth*, *120* **(2)**, 1159–1173, 2015, 2015JB011870.

Nataf, H.-C., I. Nakanishi, and D. L. Anderson, Measurements of mantle wave velocities and inversion for lateral heterogeneities and anisotropy: 3. Inversion, *Journal of Geophysical Research: Solid Earth*, *91* **(B7)**, 7261–7307, 1986.

Nicolson, H., A. Curtis, B. Baptie, and E. Galetti, Seismic interferometry and ambient noise tomography in the British Isles, *Proceedings of the Geologists' Association*, *123* **(1)**, 74 – 86, 2012.

Nishida, K., J.-P. Montagner, and H. Kawakatsu, Global Surface Wave Tomography Using Seismic Hum, *Science*, *326* **(5949)**, 112–112, 2009.

Novotný, M., Depth-Recursive Tomography of the Bohemian Massif at the CEL09 Transect—Part B: Interpretation, *Surveys in Geophysics*, *33* **(2)**, 243–273, 2012.

Novotný, M., Z. Skácelová, J. Mrlina, B. Mlčoch, and B. Růžek, Depth-recursive tomography along the Eger Rift using the S01 profile refraction data: tested at the KTB super drilling hole, structural interpretation supported by magnetic, gravity and petrophysical data, *Surveys in Geophysics*, *30* **(6)**, 561, Nov., 2009.

Novotný, O., and L. Urban, Seismic models of the Bohemian Massif and of some adjacent regions derived from deep seismic soundings and surface wave investigations: A review, *Induced Seismicity and Associated Phenomena*, , 227–249, 1988.

Novotný, O., A preliminary seismic model for the region of the West-Bohemian earthquake swarms, *Studia Geophysica et Geodaetica*, *40* **(4)**, 353–366, Oct, 1996.

Novotný, O., Seismic surface waves, http://geo.mff.cuni.cz/vyuka/Novotny-SeismicSurfaceWaves-ocr.pdf, 1999.

Obermann, A., M. Lupi, A. Mordret, S. S. Jakobsdóttir, and S. A. Miller, 3D-ambient noise Rayleigh wave tomography of Snæfellsjökull volcano, Iceland, *Journal of Volcanology and Geothermal Research*, *317*, 42–52, 2016.

Ouyang, L., H. Li, Q. Lü, Y. Yang, X. Li, G. Jiang, G. Zhang, D. Shi, D. Zheng, S. Sun et al., Crustal and uppermost mantle velocity structure and its relationship with the formation of ore districts in the Middle–Lower Yangtze River region, *Earth and Planetary Science Letters*, *408*, 378–389, 2014.

Pang, G., J. Feng, and J. Lin, Crust structure beneath Jilin Province and Liaoning Province in China based on seismic ambient noise tomography, *Journal of Volcanology and Geothermal Research*, *327*, 249–256, 2016.

Peter, D., C. Tape, L. Boschi, and J. Woodhouse, Surface wave tomography: global membrane waves and adjoint methods, *Geophysical Journal International*, *171* **(3)**, 1098–1117, 2007.

Peter, D., L. Boschi, and J. H. Woodhouse, Tomographic resolution of ray and finite-frequency methods: a membrane-wave investigation, *Geophysical Journal International*, *177* **(2)**, 624–638, 2009.

Peter, D., D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz et al., Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophysical Journal International*, *186* **(2)**, 721–739, 2011.

Piana Agostinetti, N., G. Giacomuzzi, and A. Malinverno, Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophysical Journal International*, *201* **(3)**, 1598–1617, 2015.

Pilia, S., N. Rawlinson, N. Direen, A. Reading, R. Cayley, L. Pryer, P. Arroucau, and M. Duffett, Linking mainland Australia and Tasmania using ambient seismic noise tomography: Implications for the tectonic evolution of the east Gondwana margin, *Gondwana Research*, *28* **(3)**, 1212 – 1227, 2015.

Plomerová, J., U. Achauer, V. Babuška, and M. Granet, BOHEMA 2001-2003: Passive Seismic Experiment to Study Lithosphere-Asthenosphere System in the Western Part of the Bohemian Massif, *Studia Geophysica et Geodaetica*, *47* **(3)**, 691–701, 2003.

Plomerová, J., L. Vecsey, V. Babuška, M. Granet, and U. Achauer, Passive Seismic Experiment Mosaic - A Pilot Study of Mantle Lithosphere Anisotropy of the Bohemian Massif, *Studia Geophysica et Geodaetica*, *49* **(4)**, 541–560, 2005.

Plomerová, J., U. Achauer, V. Babuška, L. Vecsey, and B. W. Group, Upper mantle beneath the Eger Rift (Central Europe): plume or asthenosphere upwelling?, *Geophysical Journal International*, *169* **(2)**, 675–682, 2007.

Poli, P., M. Campillo, H. Pedersen, and L. W. Group, Body-Wave Imaging of Earth's Mantle Discontinuities from Ambient Seismic Noise, *Science*, *338* **(6110)**, 1063–1065, 2012*a*.

Poli, P., H. A. Pedersen, M. Campillo, and t. W. Group, Emergence of body waves from cross-correlation of short period seismic noise, *Geophysical Journal International*, *188* **(2)**, 549–558, 2012*b*.

Porritt, R. W., M. S. Miller, L. J. O'Driscoll, C. W. Harris, N. Roosmawati, and L. T. da Costa, Continent–arc collision in the Banda Arc imaged by ambient noise tomography, *Earth and Planetary Science Letters*, *449*, 246–258, 2016.

Pros, Z., T. Lokajíček, R. Přikryl, A. Špičák, V. Vajdová, and K. Klíma, Elastic Parameters of West Bohemian Granites under Hydrostatic Pressure, *Pure and Applied Geophysics*, *151* **(2)**, 631–646, Mar, 1998.

de la Puente, J., M. Käser, M. Dumbser, and H. Igel, An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes - IV. Anisotropy, *Geophysical Journal International*, *169* **(3)**, 1210–1228, 2007.

Pullammanappallil, S. K., and J. N. Louie, A generalized simulated-annealing optimization for inversion of first-arrival times, *Bulletin of the Seismological Society of America*, *84* **(5)**, 1397, 1994.

Rawlinson, N., S. Pilia, M. Young, M. Salmon, and Y. Yang, Crust and upper mantle structure beneath southeast Australia from ambient noise and teleseismic tomography, *Tectonophysics*, *689*, 143 – 156, 2016.

Rawlinson, N., and W. Spakman, On the use of sensitivity tests in seismic tomography, *Geophysical Journal International*, *205* **(2)**, 1221–1243, 2016.

Rickers, F., A. Fichtner, and J. Trampert, The Iceland–Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion, *Earth and Planetary Science Letters*, *367*, 39–51, 2013.

Ritzwoller, M. H., N. M. Shapiro, M. P. Barmin, and A. L. Levshin, Global surface wave diffraction tomography, *Journal of Geophysical Research: Solid Earth*, *107* **(B12)**, ESE 4–1–ESE 4–13, 2002.

Ritzwoller, M. H., F.-C. Lin, and W. Shen, Ambient noise tomography with a large seismic array, *Comptes Rendus Geoscience*, *343* **(8)**, 558 – 570, 2011.

Romanowicz, B., 3D structure of the Earth's lower mantle, *Comptes Rendus Geoscience*, *335* **(1)**, 23 – 35, 2003.

Růžek, B., P. Hrubcová, M. Novotný, A. Špičák, and O. Karousová, Inversion of travel times obtained during active seismic refraction experiments CELEBRATION 2000, ALP 2002 and SUDETES 2003, *Studia Geophysica et Geodaetica*, *51* **(1)**, 141–164, 2007.

Růžek, B., K. Holub, and J. Rušajová, Three-dimensional Crustal model of the Moravo-Silesian region obtained by seismic tomography, *Studia Geophysica et Geodaetica*, *55* **(1)**, 87–107, 2011.

Růžek, B., J. Plomerová, and V. Babuška, Joint inversion of teleseismic P waveforms and surface-wave group velocities from ambient seismic noise in the Bohemian Massif, *Studia Geophysica et Geodaetica*, *56* **(1)**, 107–140, 2012.

Růžek, B., L. Valentová, and F. Gallovič, Significance of Geological Units of the Bohemian Massif, Czech Republic, as Seen by Ambient Noise Interferometry, *Pure and Applied Geophysics*, *173* **(5)**, 1663–1682, 2016.

Růžek, B., and J. Horálek, Three-dimensional seismic velocity model of the West Bohemia/Vogtland seismoactive region, *Geophysical Journal International*, *195* **(2)**, 1251–1266, 2013.

Ryberg, T., U. Muksin, and K. Bauer, Ambient seismic noise tomography reveals a hidden caldera and its relation to the Tarutung pull-apart basin at the Sumatran Fault Zone, Indonesia, *Journal of Volcanology and Geothermal Research*, *321*, 73 – 84, 2016.

Sambridge, M., A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization, *Geophysical Journal International*, *196* **(1)**, 357, 2014.

Saygin, E., and B. L. N. Kennett, Crustal structure of australia from ambient seismic noise tomography, *Journal of Geophysical Research: Solid Earth*, *117* **(B1)**, 2012, B01304.

Saygin, E., and B. L. Kennett, Ambient seismic noise tomography of Australian continent, *Tectonophysics*, *481* **(1–4)**, 116–125, 2010, Insights into the Earth's Deep Lithosphere.

Schulmann, K., J. Konopásek, V. Janoušek, O. Lexa, J.-M. Lardeaux, J.-B. Edel, P. Štípská, and S. Ulrich, An Andean type Palaeozoic convergence in the Bohemian Massif, *Comptes Rendus Geoscience*, *341* **(2–3)**, 266 – 286, 2009.

Schulmann, K., O. Lexa, V. Janoušek, J. M. Lardeaux, and J. B. Edel, Anatomy of a diffuse cryptic suture zone: An example from the Bohemian Massif, European Variscides, *Geology*, *42* **(4)**, 275–278, 2014.

Shapiro, N. M., and M. Campillo, Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophysical Research Letters*, *31* **(7)**, 2004, L07614.

Shapiro, N. M., M. Campillo, L. Stehly, and M. H. Ritzwoller, High-Resolution Surface-Wave Tomography from Ambient Seismic Noise, *Science*, *307* **(5715)**, 1615–1618, 2005.

Shen, W., M. H. Ritzwoller, V. Schulte-Pelkum, and F.-C. Lin, Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophysical Journal International*, *192* **(2)**, 807, 2012.

Shirzad, T., and Z. H. Shomali, Shallow crustal radial anisotropy beneath the Tehran basin of Iran from seismic ambient noise tomography, *Physics of the Earth and Planetary Interiors*, *231*, 16–29, 2014.

Sieminski, A., J.-J. Lévêque, and E. Debayle, Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophysical Research Letters*, *31* **(24)**, 2004.

Snieder, R., Large-Scale waveform inversions of surface waves for lateral heterogeneity: 2. Application to surface waves in Europe and the Mediterranean, *Journal of Geophysical Research: Solid Earth*, *93* **(B10)**, 12067–12080, 1988.

Spica, Z., V. M. Cruz-Atienza, G. Reyes-Alfaro, D. Legrand, and A. Iglesias, Crustal imaging of western Michoacán and the Jalisco Block, Mexico, from ambient seismic noise, *Journal of Volcanology and Geothermal Research*, *289*, 193–201, 2014.

Spica, Z., C. Caudron, M. Perton, T. Lecocq, T. Camelbeeck, D. Legrand, J. Piña-Flores, A. Iglesias, and D. K. Syahbana, Velocity models and site effects at Kawah Ijen volcano and Ijen caldera (Indonesia) determined from ambient noise cross-correlations and directional energy density spectral ratios, *Journal of Volcanology and Geothermal Research*, *302*, 173–189, 2015.

Tanimoto, T., Modelling curved surface wave paths: membrane surface wave synthetics, *Geophysical Journal International*, *102* **(1)**, 89–100, 1990.

Tape, C., Q. Liu, and J. Tromp, Finite-frequency tomography using adjoint methods–Methodology and examples using membrane surface waves, *Geophysical Journal International*, *168* **(3)**, 1105–1129, 2007.

Tape, C., Q. Liu, A. Maggi, and J. Tromp, Adjoint tomography of the southern California crust, *Science*, *325* **(5943)**, 988–992, 2009.

Tape, C., Q. Liu, A. Maggi, and J. Tromp, Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophysical Journal International*, *180* **(1)**, 433–462, 2010.

Tarantola, A., *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

Tarantola, A., and B. Valette, Inverse problems = Quest for Information, *Journal of Geophysics*, *50* **(3)**, 159–170, 1982.

Thomson, W. T., Transmission of Elastic Waves through a Stratified Solid Medium, *Journal of Applied Physics*, *21* **(2)**, 89–93, 1950.

Tierney, L., Markov Chains for Exploring Posterior Distributions, *The Annals of Statistics*, *22* **(4)**, 1701–1728, 12, 1994.

Trampert, J., and J. Spetzler, Surface wave tomography: finite-frequency effects lost in the null space, *Geophysical Journal International*, *164* **(2)**, 394–400, 2006.

Tromp, J., C. Tape, and Q. Liu, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophysical Journal International*, *160* **(1)**, 195–216, 2005.

Tromp, J., Y. Luo, S. Hanasoge, and D. Peter, Noise cross-correlation sensitivity kernels, *Geophysical Journal International*, *183* **(2)**, 791–819, 2010.

Uličný, D., L. Špičáková, R. Grygar, M. Svobodová, S. Čech, and J. Laurin, Palaeodrainage systems at the basal unconformity of the Bohemian Cretaceous Basin: roles of inherited fault systems and basement lithology during the onset of basin filling, *Bulletin of Geosciences*, *84* **(4)**, 577–610, 2009.

Valentová, L., F. Gallovič, B. Růžek, J. de laPuente, and P. Moczo, Choice of regularization in adjoint tomography based on two-dimensional synthetic tests, *Geophysical Journal International*, *202* **(2)**, 787–799, 2015.

Valentová, L., F. Gallovič, and P. Maierová, Three-dimensional S-wave velocity model of the Bohemian Massif from Bayesian ambient noise tomography, *Tectonophysics*, *717* **(C)**, 484 – 498, 2017.

Verbeke, J., L. Boschi, L. Stehly, E. Kissling, and A. Michelini, High-resolution Rayleigh-wave velocity maps of central Europe from a dense ambient-noise data set, *Geophysical Journal International*, *188* **(3)**, 1173–1187, 2012.

Virieux, J., and S. Operto, An overview of full-waveform inversion in exploration geophysics, *Geophysics*, *74* **(6)**, WCC1–WCC26, 2009.

Weaver, R., and O. Lobkis, On the emergence of the Green's function in the correlations of a diffuse field: pulse-echo using thermal phonons, *Ultrasonics*, *40* **(1–8)**, 435 – 439, 2002.

Wilde-Piórko, M., J. Saul, and M. Grad, Differences in the Crustal and Uppermost Mantle Structure of the Bohemian Massif from Teleseismic Receiver Functions, *Studia Geophysica et Geodaetica*, *49* **(1)**, 85–107, 2005.

Wilde-Piórko, M., W. Geissler, J. Plomerová, M. Grad, V. Babuška, E. Brückl, J. Cyziene, W. Czuba, R. England, E. Gaczyński, R. Gaždová, S. Gregersen, A. Guterch, W. Hanka, E. Hegedüs, B. Heuer, P. Jedlička, J. Lazauskiene, G. Keller, R. Kind, K. Klinge, P. Kolínský, K. Komminaho, E. Kozlovskaya, F. Krüger, T. Larsen, M. Majdański, J. Málek, G. Motuza, O. Novotný, R. Pietrasiak, T. Plenefisch, B. Růžek, S. Sliaupa, P. Środa, M. Świeczak, T. Tiira, P. Voss, and P. Wiejacz, PASSEQ 2006-2008: Passive seismic experiment in Trans-European Suture Zone, *Studia Geophysica et Geodaetica*, *52* **(3)**, 439–448, 2008.

Xu, Z., P. Chen, and Y. Chen, Sensitivity Kernel for the Weighted Norm of the Frequency-Dependent Phase Correlation, *Pure and Applied Geophysics*, *170* **(3)**, 353–371, 2013.

Yang, Y., M. H. Ritzwoller, A. L. Levshin, and N. M. Shapiro, Ambient noise Rayleigh wave tomography across Europe, *Geophysical Journal International*, *168* **(1)**, 259–274, 2007.

Yang, Y., M. H. Ritzwoller, F.-C. Lin, M. P. Moschetti, and N. M. Shapiro, Structure of the crust and uppermost mantle beneath the western united states revealed by ambient noise and earthquake tomography, *Journal of Geophysical Research: Solid Earth*, *113* **(B12)**, 2008, B12310.

Yang, Y., and D. W. Forsyth, Regional tomographic inversion of the amplitude and phase of Rayleigh waves with 2-D sensitivity kernels, *Geophysical Journal International*, *166* **(3)**, 1148–1160, 2006.

Young, M. K., R. A. Cayley, M. A. McLean, N. Rawlinson, P. Arroucau, and M. Salmon, Crustal structure of the east Gondwana margin in southeast Australia revealed by transdimensional ambient seismic noise tomography, *Geophysical Research Letters*, *40* **(16)**, 4266–4271, 2013*a*.

Young, M. K., N. Rawlinson, and T. Bodin, Transdimensional inversion of ambient seismic noise for 3D shear velocity structure of the Tasmanian crust, *Geophysics*, *78* **(3)**, WB49–WB62, 2013*b*.

Zha, Y., S. C. Webb, S. S. Wei, D. A. Wiens, D. K. Blackman, W. Menke, R. A. Dunn, and J. A. Conder, Seismological imaging of ridge–arc interaction beneath the Eastern Lau Spreading Center from OBS ambient noise tomography, *Earth and Planetary Science Letters*, *408*, 194–206, 2014.

Zhan, Z., S. Ni, D. V. Helmberger, and R. W. Clayton, Retrieval of Moho-reflected shear wave arrivals from ambient seismic noise, *Geophysical Journal International*, *182* **(1)**, 408–420, 2010.

Zhao, L., T. H. Jordan, and C. H. Chapman, Three-dimensional Fréchet differential kernels for seismic delay times, *Geophysical Journal International*, *141* **(3)**, 558–576, 2000.

Zhao, L., T. H. Jordan, K. B. Olsen, and P. Chen, Fréchet Kernels for Imaging Regional Earth Structure Based on Three-Dimensional Reference Models, *Bulletin of the Seismological Society of America*, *95* **(6)**, 2066–2080, 2005.

Zhou, Y., F. A. Dahlen, G. Nolet, and G. Laske, Finite-frequency effects in global surface-wave tomography, *Geophysical Journal International*, *163* **(3)**, 1087–1111, 2005.

Zhou, Y., G. Nolet, F. A. Dahlen, and G. Laske, Global upper-mantle structure from finite-frequency surface-wave tomography, *Journal of Geophysical Research: Solid Earth*, *111* **(B4)**, 2006, B04304.

Zhu, H., E. Bozdağ, T. S. Duffy, and J. Tromp, Seismic attenuation beneath Europe and the North Atlantic: Implications for water in the mantle, *Earth and Planetary Science Letters*, *381* **(0)**, 1 – 11, 2013.

Included papers

Valentová, L., F. Gallovič, B. Růžek, J. de laPuente, and P. Moczo, Choice of regularization in adjoint tomography based on two-dimensional synthetic tests, *Geophysical Journal International*, *202* **(2)**, 787–799, 2015.

Valentová, L., F. Gallovič, and P. Maierová, Three-dimensional S-wave velocity model of the Bohemian Massif from Bayesian ambient noise tomography, *Tectonophysics*, *717* **(C)**, 484 – 498, 2017.