Geophysical Journal International

Geophys. J. Int. (2015) **202,** 787–799
GJI Seismology

doi: 10.1093/gji/ggv183

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

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

[1] *Department of Geophysics, Faculty of Mathematics and Physics, Charles University in Prague, V Holešovičkách* 2, *CZ*-18000 *Prague, Czech Republic. E-mail: valent@karel.troja.mff.cuni.cz*
[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, *CZ*-14131 *Prague, Czech Republic*
[4] *Computer Applications in Science & Engineering, Barcelona Supercomputing Centre, Carrer del Gran Capitá* 2-4, *Barcelona, Spain*

Accepted 2015 April 27. Received 2015 April 2; in original form 2014 October 28

SUMMARY

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 models. The first test reveals the sensitivity of the result on the data noise, while the second helps to analyse the resolving power of the data set. For various noise and Gaussian smoothing levels, we analysed 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.

Key words: Numerical approximations and analysis; Tomography; Seismic tomography; Europe.

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; Fichtner *et al.* 2006; Liu & Tromp 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; Fichtner *et al.* 2013; Rickers *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.* 2007b), which is under certain conditions (e.g. large number of sources) even more efficient (Chen *et al.* 2007a).

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 & Lobkis 2002). Many studies have been devoted to the extraction of Greens' functions, see, for example, Shapiro & Campillo (2004) and 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.* 2012a,b; Lin *et al.* 2013; Lin & 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; Zhou *et al.* 2005; Boschi 2006). However, other authors (e.g. Ritzwoller *et al.* 2002; Yang & Forsyth 2006) have claimed significant improvement when using the finite-frequency sensitivity kernels. Peter *et al.* (2009) used the membrane wave approximation 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 & 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) and Gao & 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) and 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 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. 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 and in the paper by Růžek *et al.* (2012).

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. 1(a) shows the 20 s Love-wave data coverage. Each of the 568 lines connects two stations with estimated traveltime value. Fig. 1(b) 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.

3 METHOD

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. 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

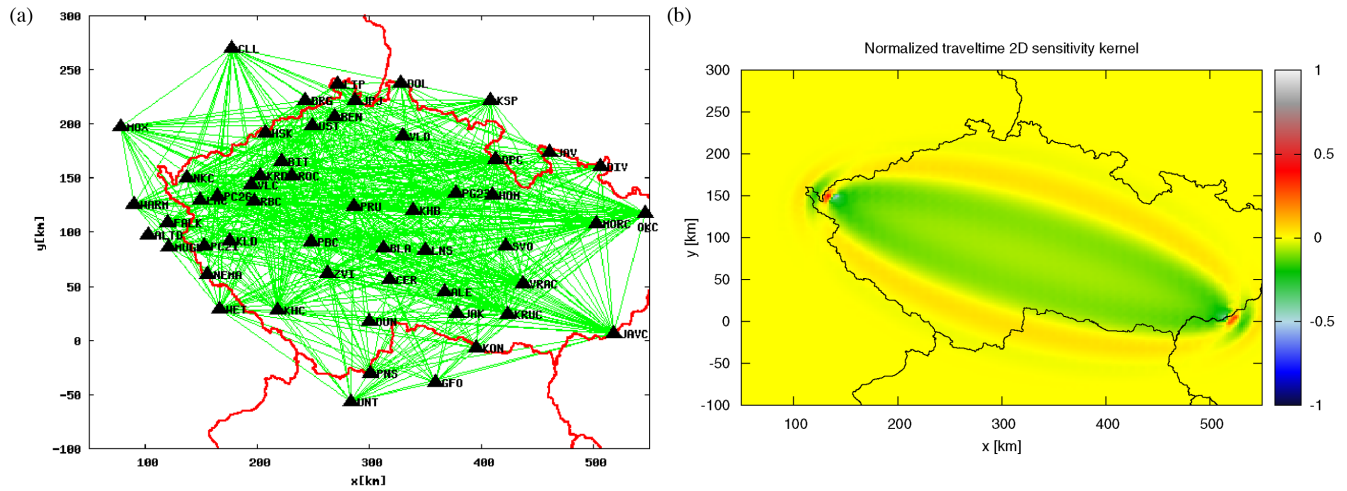

Figure 1. (a) Positions of stations used in the ambient noise processing. The stations are connected with lines along which traveltimes of the 20 s Love waves for the station pairs were obtained. The figure demonstrates the almost perfect data coverage of the studied domain. (b) Normalized traveltime sensitivity kernel for membrane waves with dominant period 20 s between two selected stations.

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.

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,

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

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:

$$\Delta T_i = \arg\max_t \int u_i(\tau) u_i^0(t+\tau)\mathrm{d}\tau. \tag{2}$$

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

$$\delta_m \chi = \sum_i h_i \Delta T_i \int_V K_i\, \delta m\, \mathrm{d}V. \tag{3}$$

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)

$$K_i = \int_t (\nabla u_i)\cdot(\nabla u_i^\dagger)\mathrm{d}t, \tag{4}$$

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 & Schuster 1991)

$$f_i^\dagger = -\frac{\dot{u}_i}{\int \dot{u}_i{}^2 \mathrm{d}t}\delta(x - x_i), \tag{5}$$

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 (Dumbser & Käser 2006; Käser & Dumbser 2006; de la Puente *et al.* 2007; Dumbser *et al.* 2007; Käser *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.

3.3 Iteration scheme and regularization

According to eq. (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. 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.* 2007a), 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.

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:

$$\zeta = \frac{||m_n - m_{targ}||}{||m_0||} \cdot 100\% \qquad (6)$$

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 and 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.

4.1 Target models

The tests are performed using two different target models:

(i) Model I: smooth model with small maximum amplitudes of the heterogeneity.
(ii) 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. 2b).

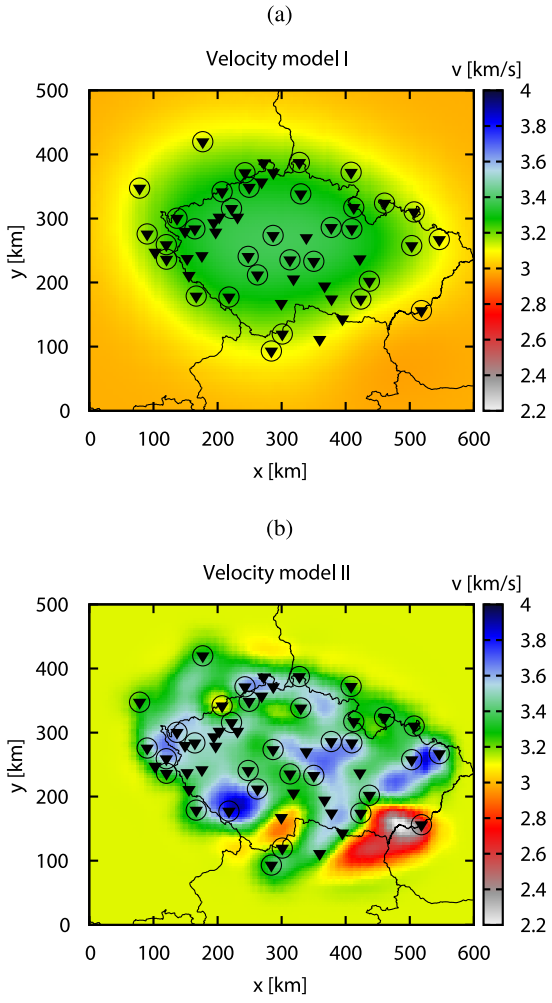

Figure 2. Velocity (target) models I (a) and II (b) for synthetic tests using 20 s data. Stations and sources are shown by inverted triangles and circles, respectively.

The two target models used to generate synthetic data for the tests are shown in Fig. 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 down-weighted by $h_i = 0.5$ in the data misfit; see Section 3, eq. (1).

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. 3(a) 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. 3(b). 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.

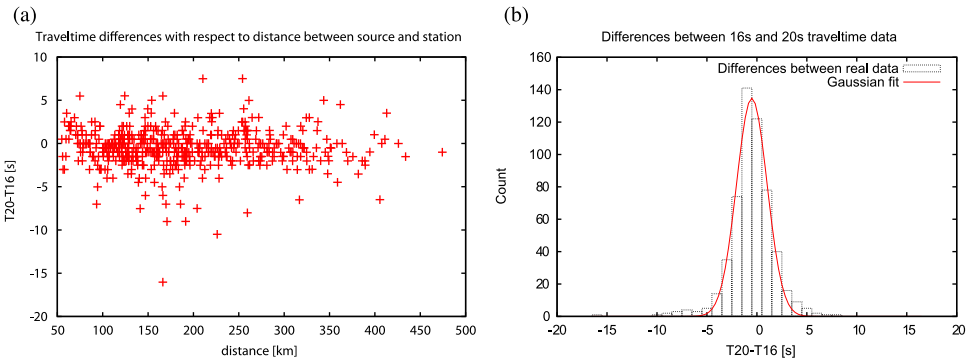

Figure 3. Differences between the 20 and 16 s Love-wave group traveltimes plotted in terms of (a) station-pair distance and (b) differences displayed as a histogram and fitted by the Gaussian distribution.

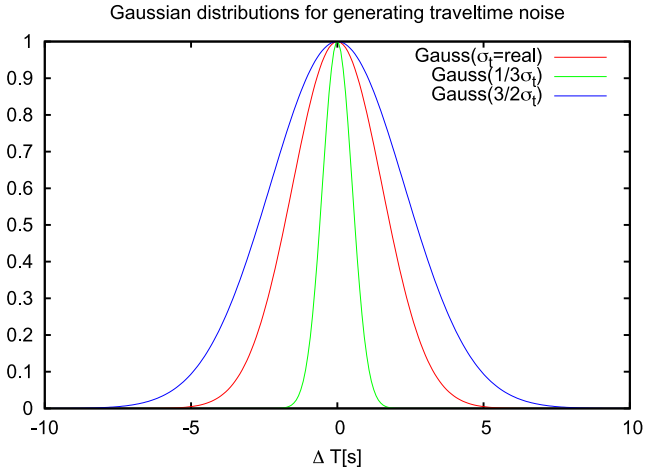

Figure 4. Gaussian functions used for generating traveltime noise in the synthetic tests: red, green and blue curves correspond to the Gaussian distributions with the standard deviations σ_t (from real data), $1/3\sigma_t$ and $3/2\sigma_t$, respectively.

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. 4. The 'accurate' synthetic seismograms calculated for models shown in Fig. 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.

4.3 Test I—inversion of the long-wavelength structures

The results for Test I are shown in the left column of Fig. 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. 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. 5a), the optimal model was achieved in 1–3 iteration steps for all noise levels. Fig. 6(a) 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. 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 6b and d), 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 5c and e). Fig. 6(e) 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. 6(f). Note that the long-wavelength structures in all the inferred models in Fig. 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. 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.

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. 5. Gradient smoothing

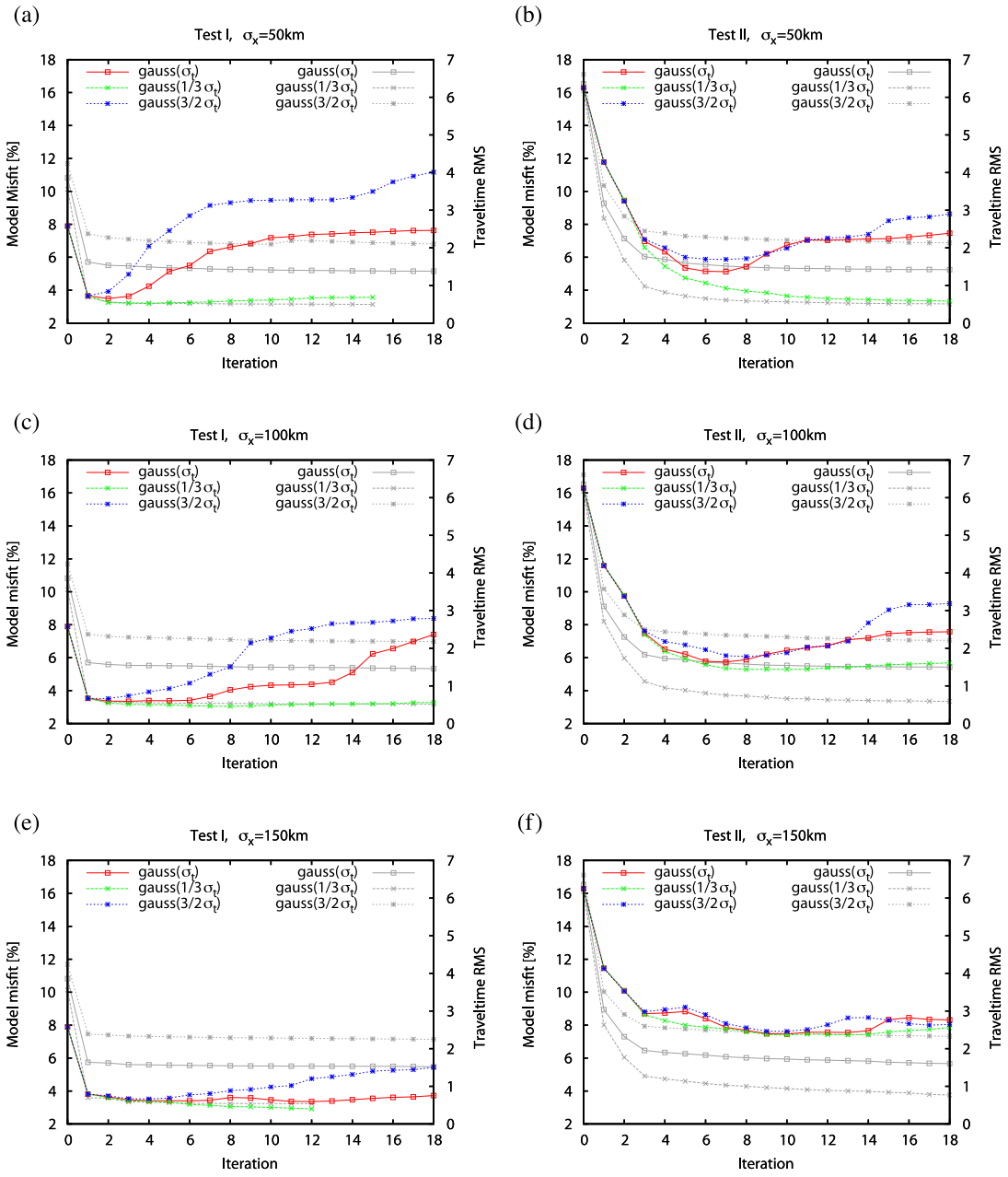

Figure 5. Results of Test I (left column) and Test II (right column). Colour curves and left vertical axis: convergence towards the target model represented by the model misfit. Grey curves and right vertical axis: data misfit in terms of traveltime residual RMS. Each row corresponds to a different level 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.

σ_x increases in Fig. 5 from the top to bottom. For illustration, Fig. 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 7(a) and (b) 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

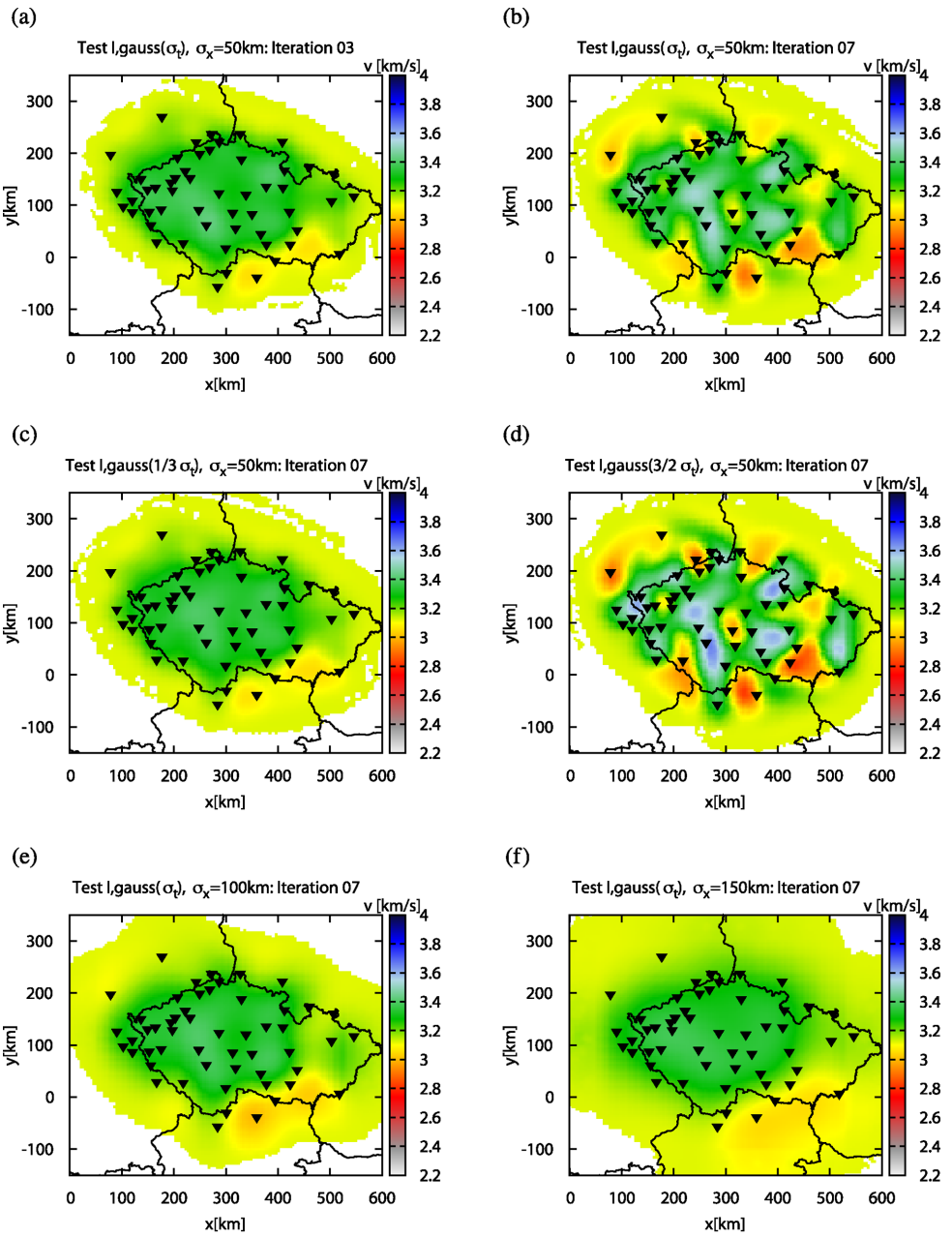

Figure 6. Models obtained in Test I for various noise levels (σ_t,$1/3\sigma_t$ and $3/2\sigma_t$; see Fig. 4), various smoothing levels σ_x and iterations.

stabilize close to the optimum for more iterations, see Fig. 5(d). 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. 7(f). This is observed as the higher value of the model misfit in the optimum as compared with other cases, Fig. 5(f).

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.

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. 8 together with the results of the inversion using original stations' positions for comparison. In particular, Fig. 8(a) shows the model misfit and the decrease of the data misfit with iterations. The decrease of

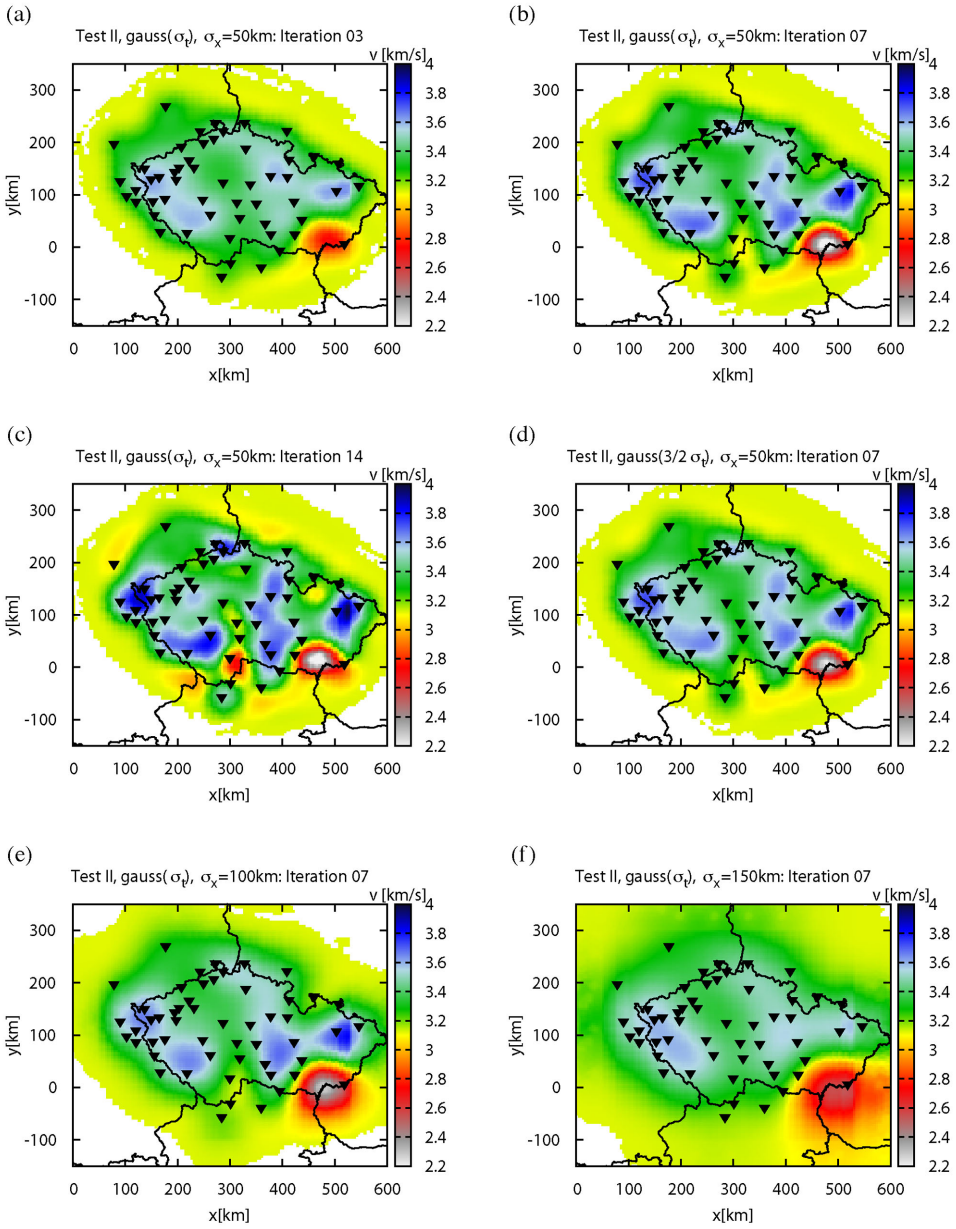

Figure 7. Same as Fig. 6, but for Test II.

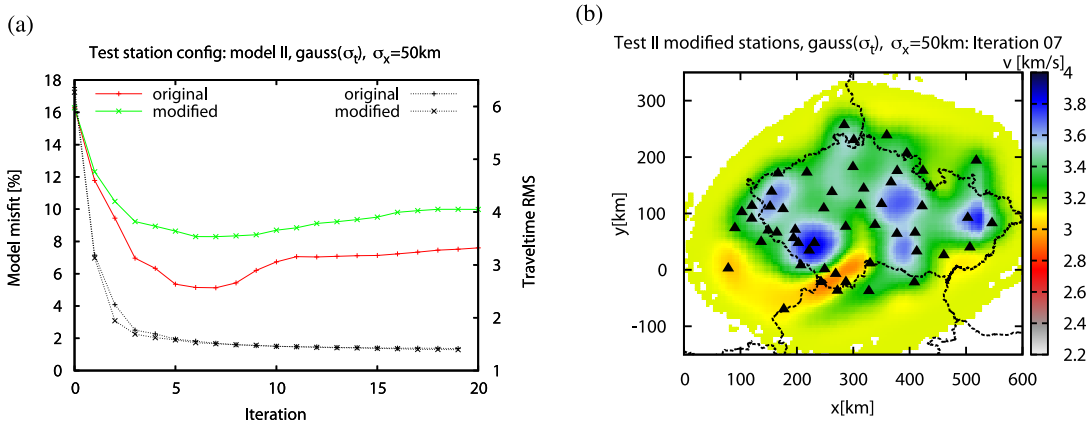

Figure 8. (a) Model and data misfit with iterations for the original and modified station configurations using Model II, noise level σ_t and gradient smoothing 50 km. (b) Model for modified station configuration in the minimum of the model misfit.

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. 8(b) 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.

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.

5 REAL DATA EXAMPLE: INVERSION OF THE 20 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 4.2). From the synthetic tests with the Gaussian errors (Section 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 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 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 9(a) and (b) 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 9(c) and (d) show the decrease of the data misfit for the two periods considered. The distribution of the traveltime residuals is also shown (Figs 9e and f). 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.

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 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 B). 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

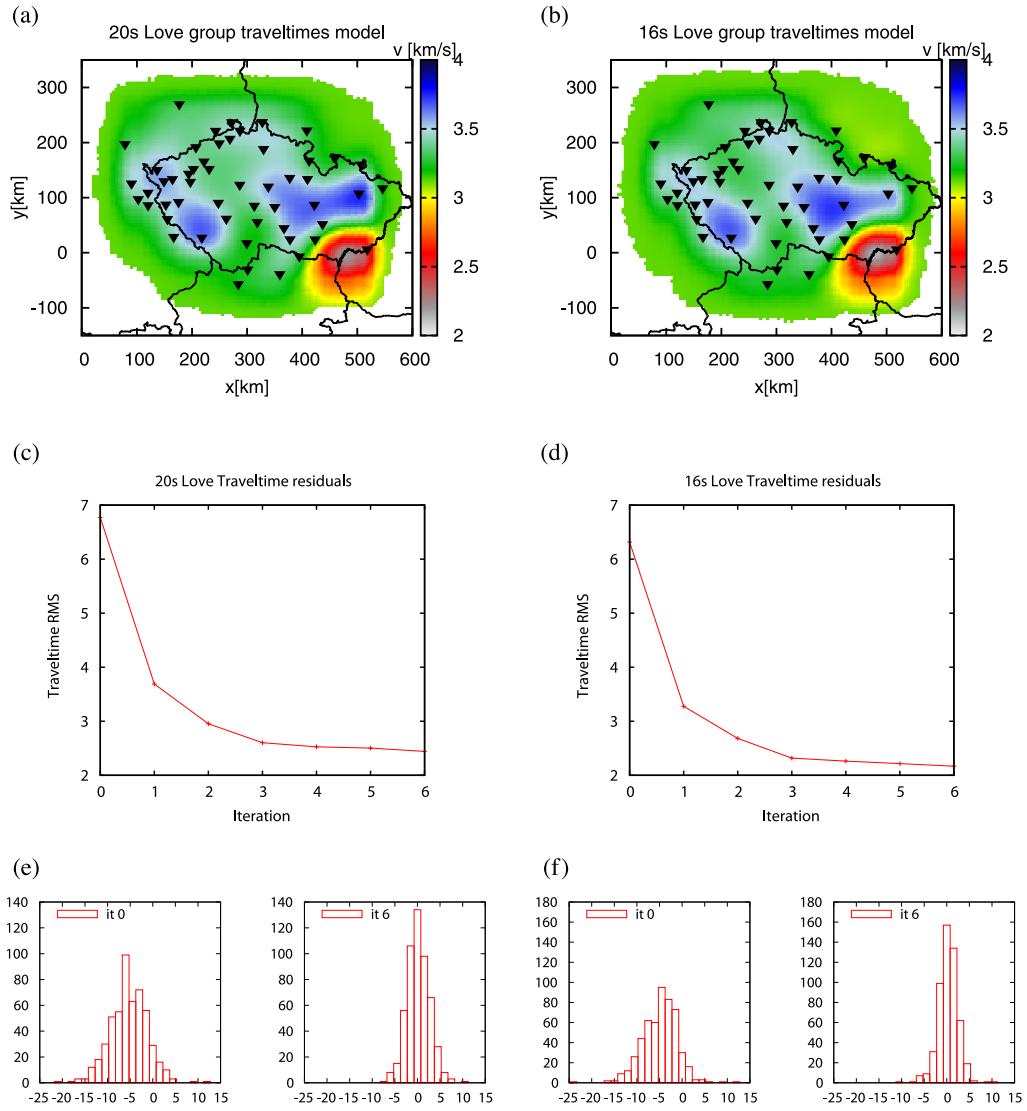

Figure 9. Results of real Love-wave data inversion: left 20 s; right 16 s. (a, b) Group velocity models; (c, d) traveltime residual RMS with iterations; (e, f) traveltime residual histograms for initial model and resulting model.

to 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 9(c) and (d) 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. 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 (≈5–10). Otherwise, the model may be spoiled by the data noise artefacts even if the data coverage seems perfect.

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.

ACKNOWLEDGEMENTS

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 also been supported by the Grant Agency of the Charles University under the project SVV-260218 and the Czech Science Foundation project 14-04372S.

REFERENCES

Babuška, V., Plomerová, J., Vecsey, L., Jedlička, P. & Růžek, B., 2005. Ongoing passive seismic experiments unravel deep lithosphere structure of the Bohemian Massif, *Stud. Geophys. Geod.,* **49**(3), 423–430.

Barmin, M.P., Ritzwoller, M.H. & Levshin, A.L., 2001. A fast and reliable method for surface wave tomography, *Pure appl. Geophys.,* **158**(8), 1351–1375.

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

Bensen, G.D., Ritzwoller, M.H. & Shapiro, N.M., 2008. Broadband ambient noise surface wave tomography across the United States, *J. geophys. Res.,* **113**(B5), doi:10.1029/2007JB005248.

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

Boué, P., Poli, P., Campillo, M., Pedersen, H., Briand, X. & Roux, P., 2013. Teleseismic correlations of ambient seismic noise for deep global imaging of the earth, *Geophys. J. Int.,* **194**(2), 844–848.

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

Chen, P., Zhao, L. & Jordan, T.H., 2007b. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97**(4), 1094–1120.

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

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

de la Puente, J., Käser, M., Dumbser, M. & Igel, H., 2007. An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes - IV. Anisotropy, *Geophys. J. Int.,* **169**(3), 1210–1228.

Dumbser, M. & Käser, M., 2006. An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes: II. The three-dimensional isotropic case, *Geophys. J. Int.,* **167**(1), 319–336.

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

Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. Theory, *Phys. Earth planet. Inter.,* **157**(1), 86–104.

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

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villaseñor, A., 2013. Multiscale full waveform inversion, *Geophys. J. Int.,* **194**(1), 534–556.

Gao, H. & Shen, Y., 2014. Upper mantle structure of the Cascades from full-wave ambient noise tomography: Evidence for 3D mantle upwelling in the back-arc, *Earth planet. Sci. Lett.,* **390**, 222–233.

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

Gouédard, P. *et al.*, 2008. Cross-correlation of random fields: mathematical approach and applications, *Geophys. Prospect.,* **56**(3), 375–393.

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

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

Lévêque, J.-J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.,* **115**(1), 313–318.

Lin, F.-C. & Tsai, V.C., 2013. Seismic interferometry with antipodal station pairs, *Geophys. Res. Lett.,* **40**(17), 4609–4613.

Lin, F.-C., Tsai, V.C., Schmandt, B., Duputel, Z. & Zhan, Z., 2013. Extracting seismic core phases with array interferometry, *Geophys. Res. Lett.,* **40**(6), 1049–1053.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.,* **96**(6), 2383–2397.

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

Moschetti, M.P., Ritzwoller, M.H. & Shapiro, N.M., 2007. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh wave group velocity maps, *Geochem. Geophys. Geosyst.,* **8**(8), Q08010, doi:10.1029/2007GC001655.

Peter, D., Tape, C., Boschi, L. & Woodhouse, J., 2007. Surface wave tomography: global membrane waves and adjoint methods, *Geophys. J. Int.,* **171**(3), 1098–1117.

Peter, D., Boschi, L. & Woodhouse, J.H., 2009. Tomographic resolution of ray and finite-frequency methods: a membrane-wave investigation, *Geophys. J. Int.,* **177**(2), 624–638.

Peter, D. *et al.*, 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.,* **186**(2), 721–739.

Plomerová, J., Achauer, U., Babuška, V. & Granet, M., 2003. BOHEMA 2001–2003: Passive seismic experiment to study lithosphere-asthenosphere system in the western part of the Bohemian Massif, *Stud. Geophys. Geod.,* **47**(3), 691–701.

Poli, P., Campillo, M., Pedersen, H. & Group, L.W., 2012a. Body-wave imaging of Earth's mantle discontinuities from ambient seismic noise, *Science,* **338**(6110), 1063–1065.

Poli, P., Pedersen, H.A. & Campillo, M. the POLENET/LAPNET Working Group, 2012b. Emergence of body waves from cross-correlation of short period seismic noise, *Geophys. J. Int.,* **188**(2), 549–558.

Rickers, F., Fichtner, A. & Trampert, J., 2013. The Iceland–Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion, *Earth planet. Sci. Lett.,* **367**, 39–51.

Ritzwoller, M.H., Shapiro, N.M., Barmin, M.P. & Levshin, A.L., 2002. Global surface wave diffraction tomography, *J. geophys. Res.,* **107**(B12), ESE 4–1–ESE 4–13.

Růžek, B., Plomerová, J. & Babuška, V., 2012. Joint inversion of teleseismic P waveforms and surface-wave group velocities from ambient seismic noise in the Bohemian Massif, *Stud. Geophys. Geod.,* **56**(1), 107–140.

Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.,* **31**(7), doi:10.1029/2004GL019491.

Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science,* **307**(5715), 1615–1618.

Sieminski, A., Lévêque, J.-J. & Debayle, E., 2004. Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophys. Res. Lett.,* **31**(24), doi:10.1029/2004GL021402.

Tanimoto, T., 1990. Modelling curved surface wave paths: membrane surface wave synthetics, *Geophys. J. Int.,* **102**(1), 89–100.

Tape, C., Liu, Q. & Tromp, J., 2007. Finite-frequency tomography using adjoint methods – Methodology and examples using membrane surface waves, *Geophys. J. Int.,* **168**(3), 1105–1129.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science,* **325**(5943), 988–992.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.,* **180**(1), 433–462.

Trampert, J. & Spetzler, J., 2006. Surface wave tomography: finite-frequency effects lost in the null space, *Geophys. J. Int.,* **164**(2), 394–400.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.

Tromp, J., Luo, Y., Hanasoge, S. & Peter, D., 2010. Noise cross-correlation sensitivity kernels, *Geophys. J. Int.,* **183**(2), 791–819.

Verbeke, J., Boschi, L., Stehly, L., Kissling, E. & Michelini, A., 2012. High-resolution Rayleigh-wave velocity maps of central Europe from a dense ambient-noise data set, *Geophys. J. Int.,* **188**(3), 1173–1187.

Weaver, R. & Lobkis, O., 2002. On the emergence of the Green's function in the correlations of a diffuse field: pulse-echo using thermal phonons, *Ultrasonics,* **40**(18), 435–439.

Wilde-Piórko, M. *et al.*, 2008. PASSEQ 2006-2008: Passive seismic experiment in Trans-European Suture Zone, *Stud. Geophys. Geod.,* **52**(3), 439–448.

Xu, Z., Chen, P. & Chen, Y., 2013. Sensitivity kernel for the weighted norm of the frequency-dependent phase correlation, *Pure appl. Geophys.,* **170**(3), 353–371.

Yang, Y. & Forsyth, D.W., 2006. Regional tomographic inversion of the amplitude and phase of Rayleigh waves with 2-D sensitivity kernels, *Geophys. J. Int.,* **166**(3), 1148–1160.

Yang, Y., Ritzwoller, M.H., Levshin, A.L. & Shapiro, N.M., 2007. Ambient noise Rayleigh wave tomography across Europe, *Geophys. J. Int.,* **168**(1), 259–274.

Zhan, Z., Ni, S., Helmberger, D.V. & Clayton, R.W., 2010. Retrieval of Moho-reflected shear wave arrivals from ambient seismic noise, *Geophys. J. Int.,* **182**(1), 408–420.

Zhao, L., Jordan, T.H., Olsen, K.B. & Chen, P., 2005. Fréchet kernels for imaging regional Earth structure based on three-dimensional reference models, *Bull. seism. Soc. Am.,* **95**(6), 2066–2080.

Zhou, Y., Dahlen, F.A., Nolet, G. & Laske, G., 2005. Finite-frequency effects in global surface-wave tomography, *Geophys. J. Int.,* **163**(3), 1087–1111.

Zhu, H., Bozdağ, E., Duffy, T.S. & Tromp, J., 2013. Seismic attenuation beneath Europe and the North Atlantic: implications for water in the mantle, *Earth planet. Sci. Lett.,* **381**(0), 1–11.

APPENDIX A: 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 bandpass 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.

APPENDIX B: 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

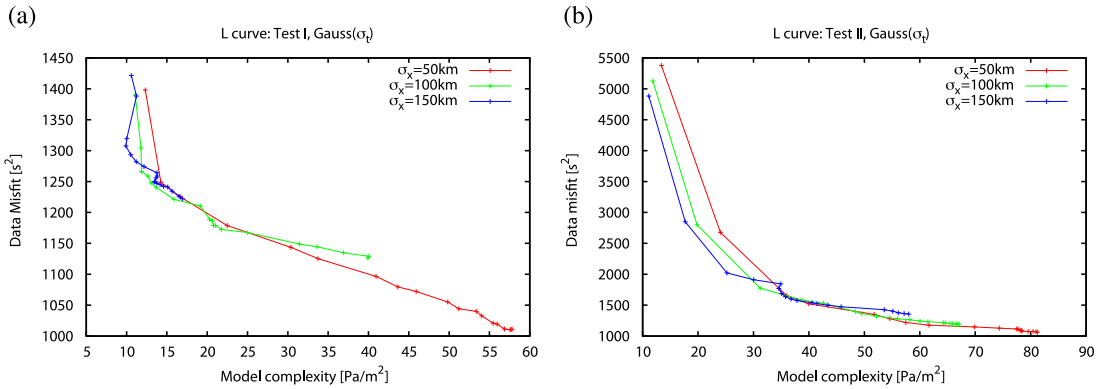

Figure B1. L-curve plot between model complexity and data misfit for Test I (a) and Test II (b) traveltime Gaussian noise with standard deviation σ_t.

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. B1. 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. B1), 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. 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. 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.