DETERMINING A 1-D VELOCITY MODEL OF THE UPPERMOST CRUST FROM *P* AND *S* ARRIVAL TIMES USING THE NEIGHBOURHOOD ALGORITHM: SYNTHETIC TEST

Jaromír JANSKÝ *, Vladimír PLICKA and Oldrich NOVOTNÝ

Department of Geophysics, Faculty of Mathematics and Physics, Charles University, V Holešovickách 2, 180 00 Praha 8, Fax +420 221911555
*Corresponding author's e-mail: jansky@seis.karlov.mff.cuni.cz

ABSTRACT
Using synthetic data we study the possibility of determining 1-D velocity models of the upper crust from P- and S-wave arrival times in the case of a narrow depth interval of seismic sources and sparse distribution of stations. The test is tailored to a similar real situation in one subregion of the western part of the Corinth Gulf, Greece. Two kinds of models are studied: (i) models composed of layers with constant velocity gradients, and (ii) models composed of homogeneous layers. To derive the structural models from arrival times, the Neighbourhood Algorithm of Sambridge (1999) is used, combined with the grid search for source locations. Weighted P- and S-wave arrival time residuals are used as the misfit function. Accurate and perturbed synthetic arrival times are used. The velocities at medium depths, with a fast velocity increase, are well determined in both models for the accurate data. However, the determination of velocity is less certain in the uppermost 5 km for the gradient model, and in the deepest layer for the model composed of homogeneous layers for the perturbed data. The presence or absence of hypocentres in the uppermost or in the second layer influences notably the obtained velocity in these layers in both models.

KEYWORDS: synthetic test, uppermost crust, arrival-time inversion, Neighbourhood Algorithm, grid search

1. INTRODUCTION

Contemporary studies of the Earth's crust by means of seismic body waves are based on a complex of methods that include the reflection method, method of refracted and overcritically reflected waves (usually referred to as deep seismic sounding), method of converted waves and seismic tomography (Pavlenkova, 1999). Artificial seismic sources are usually used in reflection and refraction methods, whereas earthquakes are usually used in the method of converted waves and in seismic tomography. In the present study we shall deal with synthetic earthquakes.

The western part of the Gulf of Corinth is very active seismically. Interesting geodynamic processes and the related seismic hazard call for detailed studies of this area. Therefore, in addition to standard monitoring (of stronger events) in this area by the all-Greece seismic network of the National Observatory of Athens, NOA (see: http://www.gein.noa.gr), and the western Greece network of the University of Patras, PATNET (see: http://seismo.geology.upatras.gr), special seismic studies have been accomplished, using temporary seismic networks.

In 1991 a seismological experiment was carried out (Rigo et al., 1996) in which 51 digital seismic stations (30 of them 3-component stations) covered a territory of about 45x45 km of the western part of the Gulf; identical with the area shown by Fig. 1. A simple crustal model, composed of homogeneous layers, was derived from these data.

The data from this experiment have recently been inverted into a 3-D tomography model of this area (Latorre et al., 2004). The results for 5 subregions of the area were presented in the form of 1-D velocity sections for depths ranging from 2 to 13 km; see Latorre et al. (2004), page 1029, Fig. 16. Subregions a) and b) in this Figure, delineated in our Fig. 1, have similar *P*-wave velocity sections. They can be represented roughly by three layers with constant velocity gradients where the layer bottoms are at depths of 5, 7 and 13 km, respectively (model TM in Fig. 2). The second layer has larger velocity gradient compared to the other two layers.

Many events occurred during the year 2001 in a subregion centered near 22.1°E and 38.3°N, i.e. in the area near to subregions a) and b) of Lattore et al. (2004). These events were recorded by the Corinth Rift Laboratory (CRL) seismic network (CRLNET, Lyon-Caen et al., 2004). The CRLNET is formed by 17 stations situated almost equally on the northern and southern coasts of the Gulf of Corinth (triangles in Fig. 1). Note that stations SER and LAK are operated by other agencies, but supply the data to the CRL. The CRLNET uses the HYPO code (Lee and Valdés, 1989) for locating in the layered model derived by
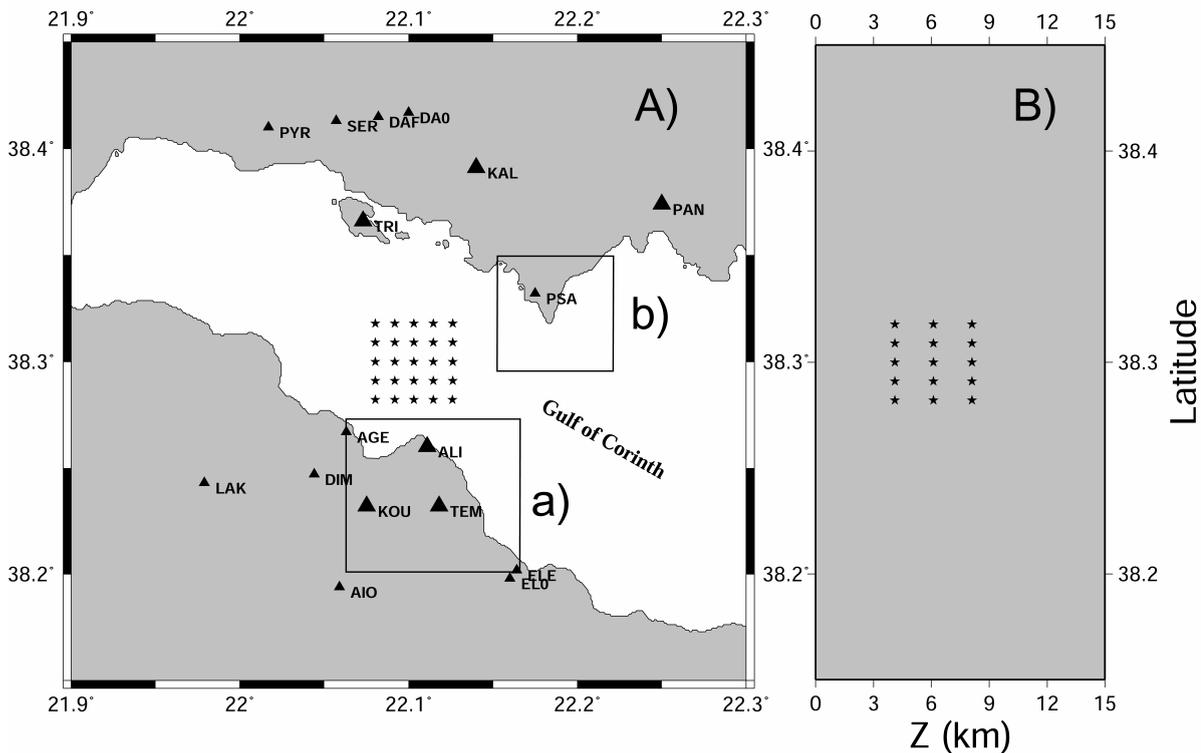
Fig. 1 Western part of the Corinth Gulf with all the CRL network stations (triangles) and the 6 stations (large triangles) used in the test. The two squares give the position of subregions a) and b) from the tomography study. The distribution of synthetic hypocentres is shown as asterisks: A) epicentres, B) depths.

Rigo et al. (1996). Hereinafter we shall refer to this model as CRLM. The upper 18 km of the model are shown in Fig. 2.

The hypocentres of the 2001 events were restricted only in a very narrow depth interval of 7 ± 1 km. Moreover, these events were usually rather small, and the majority of them was thus recorded only by a part of the CRLNET. It may thus be interesting to carry out a test of the potential of such data for determining the upper-crust model, using the arrival times from synthetic hypocentres for a similar distribution of events and stations.

Synthetic tests are often used in seismology to study the accuracy of location or model determination (e.g., Zhi Xie et al., 1996; Mäntiniemi, 1995; Crosson, 1976) and to show the influence of the individual model parameters, station distribution, etc. Our test is tailored to a similar real situation in one subregion of the western part of the Corinth Gulf, Greece (Janský et al., 2007).

2.   SYNTHETIC DATA

The proximity of our subregion to subregions a) and b) of Lattore et al. (2004) makes it possible to adopt for this area the number of layers and the layer thicknesses from the tomography as a priori information for the synthetic models. We have

generated synthetic data for two models: 1) a 3-layered model composed of layers with constant velocity gradients, without velocity jumps at the layer boundaries; 2) a 3-layered model composed of homogeneous layers. In both types of models we exclude the existence of a low-velocity channel.

In particular, we have set up this test so that the three layers of the synthetic gradient model represent roughly the average *P*-wave velocity distribution (v_P) for the tomography subregions a) and b) in Latorre et al. (2004). We only assume the upward continuation of the velocity in the first layer with the same gradient to the surface, and set the bottom of the third layer to a depth of 15 km. This synthetic model, denoted as model OG (Original Gradient model), is shown in Fig. 2. The surface velocity of this model also agrees with the older surface velocity estimation for our subregion in Le Meur et al. (1997). The second model, denoted as model OH (Original model composed of Homogeneous layers), follows the CRLM only very roughly. This model is also shown in Fig. 2.

For both synthetic models we cover the source region with a rectangular grid of 75 synthetic hypocentres (5 x 5 hypocentres with a step of 1 km in longitude and latitude at three depth levels of 4.1 km, 6.1 km and 8.1 km). They are shown in Fig. 1 as asterisks. Thus each of the three layers contains 25 hypocentres.
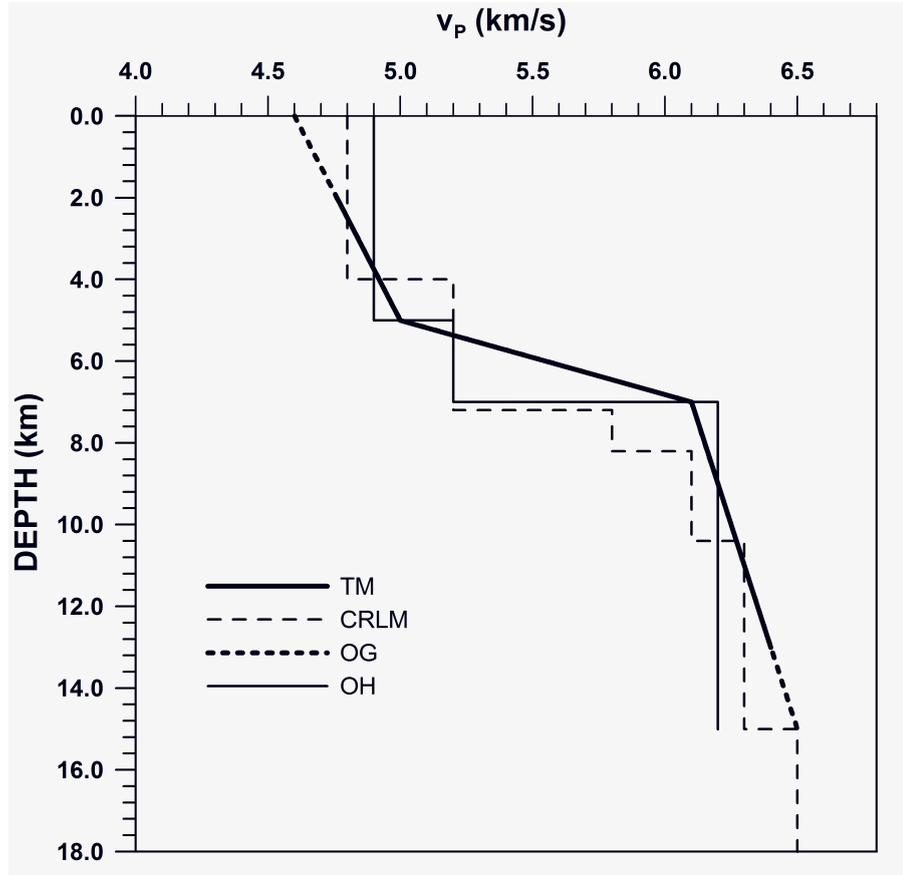
Fig. 2   Chosen velocity models: TM – tomographic model, averaged fom the gradient models of Latorre et al. (2004) for subregions a) and b) in Fig. 1; OG – original synthetic gradient model; CRLM – upper part of the model proposed by Rigo et al. (1996) and used by the CRL; OH – original synthetic model composed of homogeneous layers, roughly approximating the CRLM.

As stations for the synthetic test, we have considered three northern CRLNET stations (KAL, PAN, TRI) and three southern CRLNET stations (ALI, KOU, TEM), shown as large triangles in Fig. 1. These stations are relatively close to the given subregion and were thus able to record even weak events that occurred in our subregion in 2001. They represent a relatively sparse station network (as compared with the whole CRLNET). The azimuthal gaps for such network and our synthetic epicentres can amount to as much as 170° in the worst case.

For these models and hypocentres we have computed synthetic arrival times. The epicentral accuracy of the source-receiver two-point ray tracing iteration was set to be 0.01 km. The v_P/v_S ratio was set in both models to 1.80, i.e the same as is used in CRLM.

3. INVERSION OF THE SYNTHETIC DATA

Many different inversion techniques can be used to derive a structural model from arrival times, such as genetic algorithms (Holland, 1975), the isometric inverse algorithm (Málek et al., 2005), etc. We have selected the Neighbourhood Algorithm (NA) method (Sambridge, 1999), since in highly non-linear space it has more "power" to escape from local minima than, for example, genetic algorithms (Valleé and Bouchon, 2004; Sambridge, 1999). The main idea of the NA is the following: For each point in parameter space (i.e. for each combination of the model parameters) it is assumed that, to a first approximation, the fit is constant in some neighbourhood. To describe the neighbourhood, Voronoi cells are used (Voronoi, 1908). They are defined in the space of all sought parameters, they are unique and define a space filling, convex pavement of the space. Sambridge then describes the NA as follows: 1) Generate an initial set of ns models, each model being a certain combination of model parameters; 2) Calculate the misfit function for the (most recently generated) set of ns models and determine the nr models with the lowest misfit of all models generated so far; 3) Generate ns new models by performing a uniform random walk in the Voronoi cell of each of the chosen nr models (i.e. ns/nr samples in each cell); 4) Go to step 2. We have used *ns*=20, *nr*=10 and a total number of iterations of 80.
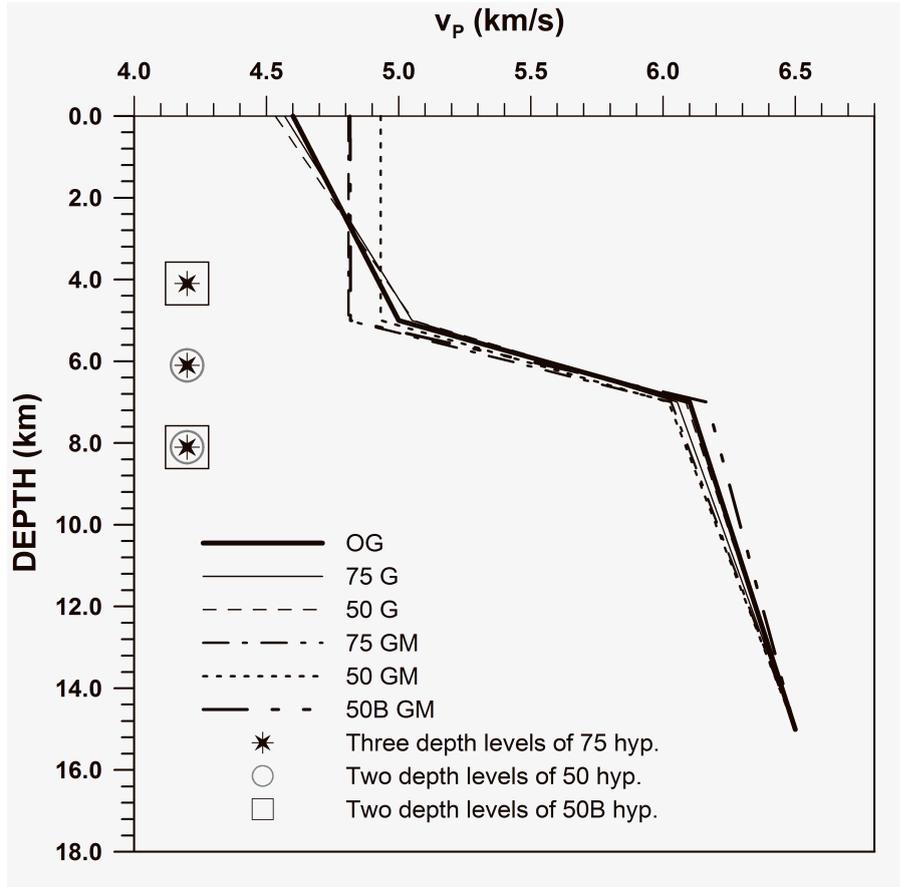
Fig. 3 Determined gradient models: OG - original synthetic gradient model used to generate the synthetic arrival data; 75G and 50G - models obtained from the inversion of synthetic accurate arrival data from 75 and 50 synthetic hypocentres, respectively; 75GM, 50GM and 50BGM - models obtained from the inversion of modified (inaccurate) synthetic arrival data from 75 and two sets of 50 synthetic hypocentres, respectively. The asterisks, circles and squares represent the depths of synthetic hypocentres, their horizontal axis coordinates being meaningless.

Consequently, a total of 1600 crustal models were generated by the NA.

We define the misfit function as the sum of squares of the weighted arrival-time residuals for the *P* and *S* waves over all stations and hypocentres of all events. We ascribe a weight of 100% to all synthetic *P*-wave onsets, and a weight of 50% to the *S*-wave onsets, to reflect the usual lower accuracy of the *S*-wave onset readings. To get the arrival times we run the event location in each generated model using a grid search and a two point ray tracing. A dense rectangular 3-D grid is used for location that covers the area of the synthetic hypocentres with a fine step of 0.2 km in all three directions. In this way, more than a thousand points are considered as a possible hypocentre for each event and generated model. In this approach the estimation of hypocentral parameters is not included directly into the NA search.

Two important factors are studied by the synthetic tests, namely the influence of the depth distribution of sources, and the influence of deviations of arrival times. To test the first factor, we compare the results using the data from all the 75 hypocentres (at depth levels of 4.1, 6.1 and 8.1 km) and the data from the 50 hypocentres only in two combinations: hypocentres at the depths of 6.1 and 8.1 km and hypocentres at the depths of 4.1 and 8.1 km.

To simulate the deviation of the medium from its 1-D representation, or the inaccuracy of time readings, we modify (perturbe) the synthetic arrival data for both models in the following way: we subtract 0.04 s from the *P* arrival time for stations ALI and KAL, and add 0.04 s for stations TEM and TRI. This value represents a time error of 4 samples for the usual sampling frequency of 100 Hz. In this way the modification of arrival times is similar for the northern and southern stations. The perturbations for the *S*-wave arrival times are that for *P* waves multiplied by 1.8, i.e. by the v_P/v_S ratio. The perturbations represent from 1% to 5% of the

individual synthetic arrival times, according to the epicentral distance of the station.

To characterize the difference between the original and inverted models, we use the integral velocity difference defined as $\int |v_P{}^o - v_P{}'|\, dz$, where $v_P{}^o$ denotes the original model and $v_P{}'$ the found model, and the integration is over depth z from 0 km to 15 km.

A. INVERSION OF THE DATA FOR THE GRADIENT MODELS

For the gradient model we determine the following 4 parameters: the v_P/v_S ratio, the velocity at the top of the first layer v_{P1}, velocity gradient B_i in the first and second layers ($v_{Pi+1}=v_{Pi}+B_id_i$, $i=1,2$), where d_i is the thickness of the i-th layer. Gradient B_3 in the third layer follows from the velocity obtained for the bottom of the second layer and the fixed velocity at the bottom of the third layer (identical to the OG model).

The results of inversion of the synthetic data under different conditions for the gradient model are given in Fig. 3. We compare the original gradient model (OG) with the results of inversion using the accurate data from 75 hypocentres (model 75G) and using the data from 50 deeper hypocentres (model 50G). The model velocity difference and other characteristics, let us call them characteristic parameters, are given in Table 1. Figure 3 (and Table 1) thus compares the influence of the hypocentre depth distribution in the case of the gradient model. The other characteristics are the obtained v_P/v_S ratio, average RMS and the average difference in epicentre and depth between the original synthetic hypocentre and the corresponding hypocentre found for the determined model.

For the gradient model and accurate data the inversion results (measured by the velocity difference) converge sufficiently to the original OG model, and are almost the same for the 50 and 75 events. The RMS is negligible and the v_P/v_S ratio was determined almost exactly to the true value of 1.80. The differences in original (synthetic) and estimated locations are negligible.

Further, we retrieve the crustal model using the modified (perturbed) synthetic arrival data again for 75 and two combinations of the 50 synthetic hypocentres. We get models denoted as 75GM, 50GM (for hypocentres at the depths of 6.1 and 8.1 km) and 50BGM (for hypocentres at the depths of 4.1 and 8.1 km). These models are shown again in Fig. 3 and their characteristic parameters are given in Table 1. These models differ more from the original model than models 75G and 50G, derived from the accurate arrival times. Large deviations from the original model are seen in the first layer, where the obtained velocities are almost constant. The models have notably larger velocity differences, larger RMS and larger differences in location. The v_P/v_S increases to 1.82 and 1.83, respectively. The difference between

the 75GM and 50GM is the largest in the first layer. The 75GM and 50BGM deviate less from the OG model than the 50GM model due to the existence of sources in the first layer. This test indicates that the inversion of inaccurate data may lead to problems in determining the velocity gradients of shallow layers.

B. INVERSION OF THE DATA FOR THE MODELS COMPOSED OF HOMOGENEOUS LAYERS

For the model composed of homogeneous layers, we determine the v_P/v_S ratio, velocity v_{P1} in the first layer and velocity increases dv_{Pi+1} ($v_{Pi+1}=v_{Pi}+dv_{Pi+1}$) in the second and third layers.

The results of the inversion of the synthetic data under different conditions for the model composed of homogeneous layers are given in Fig. 4. We compare the original homogeneous model (OH) with the result of the inversion in the case that we invert the accurate arrival data of 75 hypocentres (model 75H) and of 50 hypocentres (model 50H). The characteristic parameters for both models are given in Table 1. Here again the inverted models are very near to the OH model, especially the model 50H, and the RMS and the difference between the original and estimated locations are negligible. The v_P/v_S is again very stable.

The models obtained from the inversion of the modified arrival data for 75 hypocentres and two combinations of 50 hypocentres (the first one for hypocentres at the depths of 6.1 and 8.1 km and the second one for hypocentres at the depths of 4.1 and 8.1 km) are shown in Fig. 4 as 75HM, 50HM and 50BHM models and their characteristic parameters are given in Table 1. We see larger RMS, a larger difference in location and a larger velocity difference as compared with models 50H and 75H. The v_P/v_S ratio again increases to values of 1.82 and 1.83, respectively. The difference between the 75HM and 50HM is largest in the first layer, where the 75HM coincides practically with the OH model, as opposed to the 50HM. The largest deviation of models 75HM, 50HM and 50BHM from the OH model is in the third layer. This is due to the fact that only very short paths of rays from only 25 hypocentres propagate through the third layer. The absence of hypocentres in the second layer (model 50BHM) influenced significantly the velocity derived for this layer.

Let us note that the characteristic parameters in Table 1, i.e. the integral velocity difference, the v_P/v_S ratio, RMS and the difference between the original and estimated locations, for the 75G and 50G are rather near to those for the 75H and 50H. The same is valid for the 75GM, 50GM and 50BGM as compared with those for the 75HM, 50HM and 50BHM with some exception of low average depth difference for the model 50BHM, see Table 1.

C. PERFORMANCE OF THE NA ALGORITHM

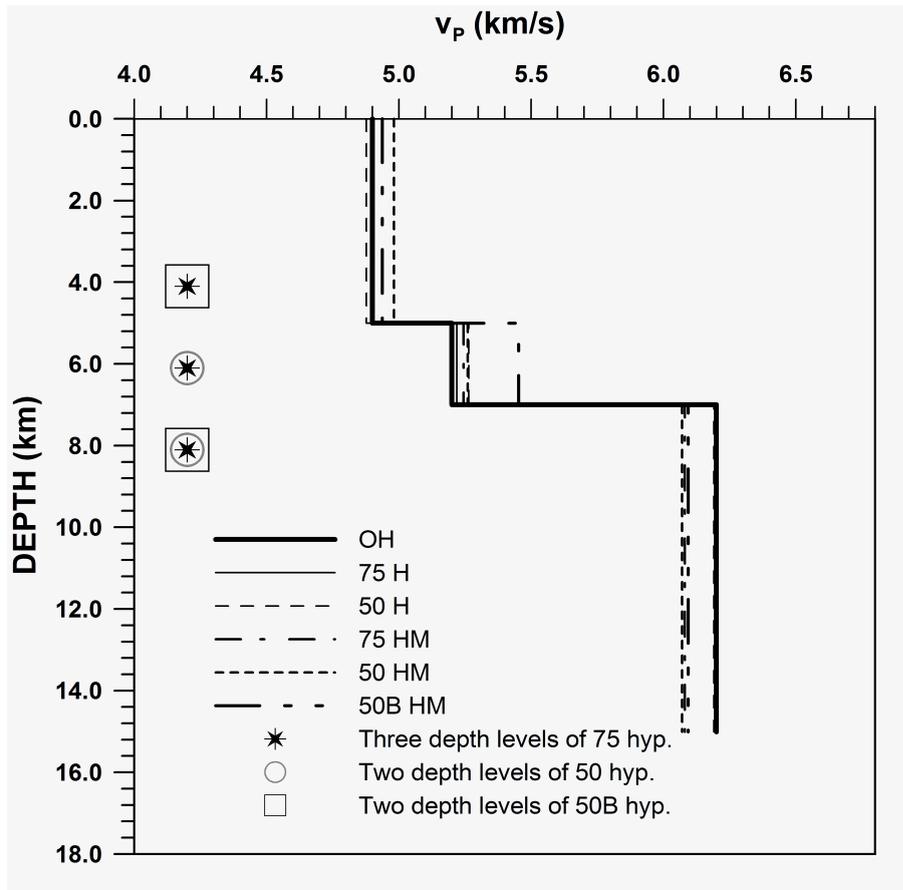To illustrate the performance of the NA, we show the convergence of the individual model

Fig. 4   Determined models composed of homogeneous layers: OH - original synthetic model composed of homogeneous layers, used to generate the synthetic arrival data; 75H and 50H - models obtained from the inversion of synthetic accurate arrival data from 75 and 50 synthetic hypocentres, respectively; 75HM, 50HM and 50BHM - models obtained from the inversion of modified synthetic arrival data from 75 and two sets of 50 synthetic hypocentres, respectively. The asterisks, circles and squares represent the depths of synthetic hypocentres.

Table 1   Characteristic parameters for the individual gradient models (G) and models composed of homogeneous layers (H) obtained by inverting the synthetic arrival data. The numbers 75 and 50 are the numbers of synthetic hypocentres used. The letter M denotes the models derived from modified (perturbed) synthetic data and B denotes the second combination of 50 hypocentres. The integral of velocity difference is related to the corresponding original model and the average location differences to the position of the original synthetic hypocentres. The original v_P/v_S was 1.80.

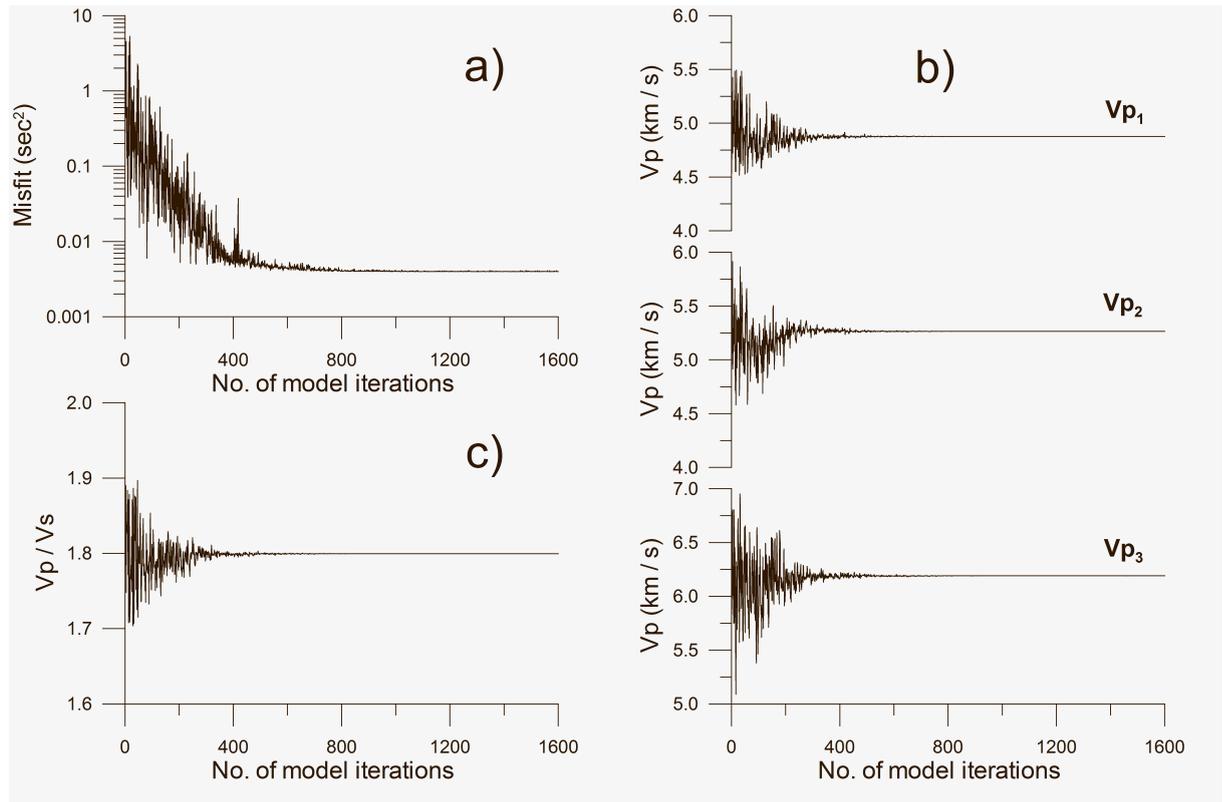| Model	Integral velocity differ. (km^2/s)	v_P/v_S	RMS (s)	Aver. epic. differ. (km)	Aver. depth differ. (km)
75G	0.34	1.80	0.006	0.01	0.02
50G	0.26	1.80	0.006	0.00	0.01
75GM	1.04	1.82	0.029	0.36	0.57
50GM	1.18	1.83	0.029	0.36	0.53
50BGM	0.92	1.83	0.029	0.27	0.52
75H	0.09	1.80	0.003	0.00	0.00
50H	0.31	1.80	0.003	0.00	0.00
75HM	1.06	1.82	0.031	0.34	0.67
50HM	1.56	1.83	0.031	0.34	0.66
50BHM	1.54	1.83	0.032	0.16	0.17

Fig. 5 Convergence of the NA algorithm in the case of deriving the model composed of homogeneous layers, using accurate data from 50 hypocentres: a) misfit function; b) v_{P1}, v_{P2} and v_{P3}; c) v_P/v_S.

parameters for model 50H in Figs. 5a – 5c. We see that less than half of the used iterations would be sufficient to get the model determination. The performance of the NA for the other sets of arrival data is similar.

4.   CONCLUSIONS

We have inverted synthetic arrival times for two 3-layered models: 1) a model composed of layers with constant velocity gradients, without velocity jumps at the layer boundaries; 2) a model composed of homogeneous layers.

The results of inversions of accurate arrival data are generally rather good and show that there is only a very small difference in inverting the data from hypocentres distributed in three or only in two layers (75 or 50 hypocentres). This is valid for inverting the data both for the gradient model and for model composed of homogeneous layers. These inversions were, more or less, run to demonstrate the inversion ability of the Neighbourhood Algorithm (in connection with the grid search location).

The most important parts of this contribution are the tests with the modified (perturbed) arrival data. Such data can simulate departures of the real medium from its 1-D representation or the influence of inaccurate onset readings. The results show a significantly lower resolution of the model parameters

and dependence of the results on the source depth distribution, as compared with the use of the accurate data set.

The obtained gradient models differ from the true model (OG model) mainly in the first layer. In this layer the deviation of the 75GM model (75 sources used) is notably lower than that for the 50GM model (only 50 deeper sources used).

The obtained models composed of homogeneous layers mutually differ again mainly in the first layer, but their larger deviation from the OH model is found in the deepest layer. We get rather significant difference from the model OH in the second layer, if there are no hypocentres in this layer (model 50BHM).

Major structural elements, i.e. the fast velocity increase between the depth of 5 and 7km in the gradient model, and the largest velocity jump at 7 km depth in the model composed of homogeneous layers, are well resolved even from the modified data.

The qualitative results of this synthetic test might help to understand better the inversion of real data in other regions, if similar conditions of limited source depth distribution and sparse station network occur. In particular, the standard velocity deviation obtained from the inversion of real data by Janský et al. (2007) for a model composed of homogeneous layers are about two times larger than the analogous differences

obtained in this paper for perturbed synthetic data. This is probably mainly due to the deviation of the local geology from the 1-D model, or overestimated accuracy of the real *S*-wave readings.

REFERENCES

Crosson, R.S.: 1976, Crustal structure modeling of earthquake data. 1. Simultaneous least squares estimation of hypocenter and velocity parameters. J. Geophys. Res., 81, 3036-3046.

Holland, J.H.: 1975, Adaptation in Natural and Artificial Systems. University of Michigan Press, Michigan.

Janský, J., Plicka, V., Lyon-Caen, H. and Novotný, O.: 2007, Estimation of velocity in the uppermost crust in a part of the western Gulf of Corinth, Greece, from the inversion of *P* and *S* arrival times using the neighbourhood algorithm. J. Seismol., 11, 199-204.

Latorre, D., Virieux, J., Monfret, T., Monteiller, V., Vanorio, T., Got, J.-L. and Lyon-Caen, H.: 2004, A new seismic tomography of Aigion area (Gulf of Corinth, Greece) from the 1991 data set. Geophys. J. Int., 159, 1013-1031.

Lee, W.H.K. and Valdés, C.M.: 1989, Hypo71PC. Toolbox for Seismic Data Acquisition, Processing, and Analysis. IASPEI & SSA.

Le Meur, H., Virieux, J. and Podvin, P.: 1997, Seismic tomography of the Gulf of Corinth: a comparison of methods. Ann. Geophys. XL, 1-25.

Lyon-Caen, H., Papadimitriou, P., Deschamps, A., Bernard, P., Makropoulos, K., Pacchiani, F. and Patau., G.: 2004, First results of the CRLN seismic array in the western Corinth rift: evidence for old fault reactivation. C. R. Geoscience, 336, 343-351.

Málek, J., Horálek, J. and Janský, J.: 2005, One-dimensional qP-wave velocity model of the upper crust for the west Bohemia/Vogtland earthquake swarm region. Studia Gephys. Geod., 49, 501-524.

Mäntiniemi, P.: 1995, Bayesian crustal structure inversion: an investigation of the feasibilities and limitations of the method. Geophys. J. Int., 123, 420-430.

Pavlenkova, N.I.: 1999, Metod glubinnogo seysmicheskogo zondirovaniya, osnovnye etapy razvitiya, dostizheniya i problemy. Fizika Zemli, No 7-8, 3-29 (in Russian).

Rigo, A., Lyon-Caen, H., Armijo, R., Deschamps, A., Hatzfeld, D., Makropoulos, K., Papadimitriou, P. and Kassaras, I.: 1996, A microseismic study in the western part of the Gulf of Corinth (Greece): implications for large scale normal faulting mechanisms. Geophys. J. Int., 126, 663-688.

Sambridge, M.: 1999, Geophysical inversion with a neighbourhood algorithm. I. Searching a parameter space. Geophys. J. Int., 138, 479-494.

Valleé, M. and Bouchon, M.: 2004, Imaging coseismic rupture in far field by slip patches. Geophys. J. Int., 156, 615-630.

Voronoi, M.G.: 1908, Nouvelles applications des paramètres continus à la thèorie des formes quadratiques, J. reine Angew. Math., 134, 198-287.

Zhi Xie, Spencer, T.W., Rabinowitz, P.D. and Fahlquist, D.A.: 1996, A new regional hypocenter location method. Bull. Seism. Soc. Am., 86, 946-958.