Robust estimation of seismic coda shape

Mikko Nikkilä[1], Valentin Polishchuk[1,2] and Dmitry Krasnoshchekov[3*]
[1]*Helsinki Institute for Information Technology, CS Dept, University of Helsinki, Finland*
[2]*Communications and Transport Systems, ITN, Linkoping University, Sweden*
[3]*Institute of Dynamics of Geospheres, Leninsky pr. 38 korp. 1, Moscow 119334 Russia*
** To whom correspondence should be addressed, E-mail: krasnd@idg.chph.ras.ru*

Summary
We present a new method for estimation of seismic coda shape. It falls into the same class of methods as non-parametric shape reconstruction with the use of neural network techniques where data are split into a training and validation datasets. We particularly pursue the well-known problem of image reconstruction formulated in this case as shape isolation in the presence of a broadly defined noise. This combined approach is enabled by the intrinsic feature of seismogram which can be divided objectively into a pre-signal seismic noise with lack of the target shape, and the remainder that contains scattered waveforms compounding the coda shape. In short, we separately apply shape restoration procedure to pre-signal seismic noise and the event record, which provides successful delineation of the coda shape in the form of a smooth almost non-oscillating function of time.
The new algorithm uses a recently developed generalization of classical computational-geometry tool of α-shape. The generalization essentially yields robust shape estimation by ignoring locally a number of points treated as extreme values, noise or non-relevant data. Our algorithm is conceptually simple and enables the desired or pre-determined level of shape detail, constrainable by an arbitrary data fit criteria.
The proposed tool for coda shape delineation provides an alternative to moving averaging and/or other smoothing techniques frequently used for this purpose.
The new algorithm is illustrated with an application to the problem of estimating the coda duration after a local event. The obtained relation coefficient between coda duration and epicentral distance is consistent with the earlier findings in the region of interest.

1 Introduction
Continuous wavetrains that follow direct P- and S-arrivals on broadband records of seismic events are called coda and generally associated with random heterogeneities in the lithosphere. It has long been known that seismic coda shape possesses several robust features; for example, temporal shape of narrow-band filtered coda is independent of epicentral distance for local events (Rautian & Khalturin 1978). Many properties of coda shape, like its decay gradient or duration, proved to be helpful in geophysical studies by providing important information on seismic source and propagating media properties (see e.g. (Sato & Fehler 1998) for a review). The shape of local earthquake coda was first formalized by Aki (1969) as a superposition of incoherent seismic waves coming from uniformly distributed discrete scatterers.
The imaginary shape of seismic coda is hard to recognize by eye on seismic records of displacement or velocity. Taking envelope of seismic record makes coda shape more explicit – the envelope is essentially a presentation of seismic record's information in the other, more amenable and visually accessible form. This operation of calculation of analytic envelope is a transfer of seismic record to another form of presentation rather than processing because it does not distort the original record. The resulting envelope more prominently displays the decay

character of the seismic signal amplitude in time. To extract useful geophysical information, various signal processing techniques are applied then to the envelopes.
Earlier studies predominantly used squared (or root-mean-squared) envelopes, while more recent works began employing other instruments too. For example, Nakahara & Carcolé (2010) used the analytic envelopes. They invoked parametric regression to give Maximum-Likelihood estimates of coda parameters under the assumption that the generic form of the decay function and noise distribution are known. One of the advantages provided by analytic envelopes is that they preserve phase information of the raw data and can be used as a measure of the time variation of the total energy (kinetic and potential) involved in the seismic response. At the same time, despite their increased amenability, analytic envelopes are still strongly oscillating functions that require smoothing in order to enable further uncomplicated processing. Without smoothing, difficulties may arise, for example, when estimating seismic attenuation by fitting a model curve to specific portions of the envelope. Indeed, the commonly used least-square fitting procedure is known to heavily weigh extreme values, which makes smoothing essential.
Among a number of smoothing procedures, the simple moving averaging remains the prevailing one. The method is easy to grasp and implement, yields fast results and has only one control parameter – sliding window length (some examples of the envelopes smoothed in various time windows 1 to 100 seconds long are given in Fig. S1). At the same time, in general there exists no single criterion for selection of the window length, so in practice the selection is done to decrease fluctuations below a certain level.
In this article we propose not to smooth the analytic envelope but robustly restore its unique form by using a method for shape reconstruction from a point cloud. For this purpose we address modern techniques for processing spatial images and multidimensional data. These methods rely on Euclidean geometry and have been proven valid in a number of substantive areas (see, e.g. (Abbott & Tsay 2000)). In particular, they become very popular in geophysics these days. Most applications of such techniques, however, are related to geophysical inversion (for example, searching the parameter space that contains models of acceptable data fit – neighborhood algorithm). On the contrary, applications to time series data haven't been widespread thus far, possibly due to difficulties with constructing relevant metrics and similarity measures.
In this article we invoke *k*-order α-shape – a new computational-geometry tool for shape restoration from a point cloud. This enables us to reconstruct the coda shape and do without smoothing. We show that application of *k*-order α-shape to seismic envelopes leads to robust delineation of the coda decay curve and yields consistent derivative estimates. On a broader scale, the proposed algorithm gives a practical means for shape recovery from any signal-plus-noise dataset given as time series.

2 The tool: *k*-order α-shape

α-shape (Edelsbrunner et al. 1983) is a classical computational-geometry tool for restoring the shape out of a point cloud. Depending on application area, the input point cloud can be obtained in various ways. One common way is sampling a physical object or an environmental phenomenon.
The definition of α-shape of a point set is as follows:
*Definition [α-shape of a point set P]***.** Let $P \subset \mathrm{R}^2$ be a set of points in the plane. The *α-shape* of P connects all pairs of points $p,q \in P$ for which there exists a disk of radius α having p,q on the boundary and having no points of P inside. For example, with $\alpha = 0$ no pairs of points get

connected in the α-shape. In the other extreme case, when $\alpha \to \infty$, the α-shape is the convex hull of *P*.

α-shape formally captures the intuitive notion of "shape" of a point set (Fig. 1, left). The value of α controls the level of details with which the shape is restored. Over the years, α-shapes were used in numerous domains (see e.g. Edelsbrunner & Mucke 1994; Teichmann & Capps 1998; Albou et al. 2009; Zhou W & Yan H. 2012). Many useful properties of α-shapes have been established theoretically and confirmed in practice in diverse applications ranging from graphics to biology.

2.1 *k*-order α-shape: handling outliers

Real-world data are often noisy and contain outliers that distort the α-shape (Fig. 1, middle). To address this issue, Krasnoshchekov & Polishchuk (2014) have recently introduced *k-order α-shape* – an extension of α-shape, capable of ignoring a certain amount of outliers and providing a more robust shape reconstruction. Figure 1(left) shows α-shape of a point set, nicely reconstructing the cloud shape. In Fig. 1(middle) two outliers are added. Now the α-shape appears to miss an important feature of the set – the inner hole. The idea of *k*-order α-shape was prompted by trying to cope with outliers by allowing few points to reside inside the α-disks defining the shape. In Fig 1(right) 1-order α-shape of the set with the outliers is shown. It can be seen how *k*-order α-shape with $k > 0$, being less sensitive to the outliers, provides a more robust result.

Formally, *k*-order α-shapes are defined similarly to α-shapes using disks of radius *α*. Unlike with the standard α-shapes defined by empty disks, the disks that define *k*-order α-shapes contain *k* points of *P*. Specifically, a disk of radius α is called *k-full* if exactly *k* points of *P* reside inside the disk. The points of *P*, contained in *k*-full disks are called *outside*; the points that are not outside are called *inside*.

Definition [k-order α-shape of a point set P]. *k*-order α-shape is the α-shape of the inside points.

With this definition the α-shape is the *k*-order α-shape with $k = 0$. Algorithmic and combinatorial properties of *k*-order α-shapes are given in (Krasnoshchekov & Polishchuk 2014). There we show that *k*-order α-shapes are connected to the *k*-order Voronoi diagram in the same way in which α-shapes are connected to the Voronoi diagram. Our video demonstrating properties of *k*-order α-shapes "in action" can be viewed from (Krasnoshchekov et al. 2010). By playing with the interactive web-applet at http://www.cs.helsinki.fi/group/compgeom/kapplet/ one can get a feel of how *k*-order α-shapes work.

2.2 Reconstructing inner shape

The α-shape was originally designed to delineate the *outer* shape of the point set; following the outer boundary is enabled by using α-disks empty of points of *P*. In particular, extreme points in the data may significantly influence the α-shape. With positive *k*, *k*-order α-shape locally "shaves off" *k* extreme points of *P* – specifically, the points that are outside. As a result, the *inner* shape of the point set is obtained. See Figure 2 for an example.

We remark that α-shapes (and *k*-order α-shapes) are *non-parametric* techniques. Indeed, no *a priori* assumption is made about the reconstructed shape. That is, it is not assumed that the shape is a graph of a linear, exponential, logarithmic, or any other function. In contrast, in *parametric* regression the general form of functional dependence between the input and output is assumed to be known *in advance* – it is postulated that the dependence belongs to some known family of functions. The regression task is then to find the best values for the parameters that describe the function (i.e., the search is performed in the parameters space).

3 Algorithm and processing

The previous section described advantages of using *k*-order α-shape for shape reconstruction purposes. Here we present an algorithm for analyzing time series data with *k*-order α-shapes. We report on an application of our algorithm to processing seismic envelopes. Our main purpose is to reconstruct a unique smooth curve representing the continuous shape of seismic coda from the shape's discrete representation given in the form of seismic envelope. That is, in terms of spatial reconstruction we view this discrete dataset as a result of sampling of the original coda shape. Basing on fundamental assumptions in body wave propagation, we also assume that the sampling is sufficient to enable reconstruction of the desired shape with reasonable error. In contrast to moving averaging that smooths envelope by averaging constant number of nearby points, the presented algorithm allows one to infer spatial relationships among the dataset points. And the whole reconstruction procedure thus consists of finding spatial relationships between points of an unorganized dataset and estimating the relevant local element of the target shape.

The technical task is thus to restore the sought shape out of the point cloud that also includes noise. Our approach treats seismic noise separately from the target coda shape. *k*-order α-shape has wide variety of capabilities including the above-demonstrated options of strict rejection of extreme values. Figuring on these capabilities, we employ *k*-order α-shape to robustly extract seismic coda shape from analytic envelope serving as the input dataset to the shape reconstruction procedure. The dataset is essentially time series with the following features. First, the dataset includes a passage with seismic noise, where the sought shape is knowingly absent. Second, it has a passage with assumed presence of the target shape complicated by presence of extreme points and seismic noise. We remark that straightforward application of the classical α-shape is evidently not effective because the resulting shape would be seriously distorted by extreme points and characterize basically the outer shape of the point cloud, while we are intuitively searching for the "inner" one. This prompted us to invoke *k*-order α-shape – the generalization of α-shape, capable of handling extreme values and reconstructing the inner shape of a dataset.

3.1 Processing of time series

The algorithm: We apply the *k*-order α-shape to time series analysis as follows. For every time tick we have two disks of radius α – the *upper* disk and the *lower* disk. The upper disk is put above the data points and is moved down; the disk is stopped as soon as it contains *k* data points (Fig. 3, left). Similarly, the lower disk is put below the data and is moved up until gathering *k* points inside. Our output at the time tick is the midpoint of the segment that connects the stopped disks' centers.

To build up intuition about the algorithm's output, consider the extreme cases of $\alpha = 0$ and $\alpha = \infty$. In the former case, the output will be just the original time series; in the latter case the output will be the straight line at the average level between the k^{th} max and k^{th} min of the time series. Therefore, the algorithm should be run with reasonable choice of α and k. A practical method for choosing α and k for analytic envelope of seismic trace is presented in Appendix.

The proposed algorithm is applicable and yields an output shape on time series of any nature. We remark that seismograms are particularly amenable to the algorithm due to the presence of seismic noise. This unique feature of such data allows us to adjust α and k separately for each trace. The selection is based on a general presumption that output from the shape reconstruction procedure applied to seismic noise must be close to the straight line.

Similarly to standard time series processing methods, the output of *k*-order α-shape at a time tick is influenced only by values at time ticks that are close (within plus minus α) to the time tick. However, traditional time series methods treat the local data in a predefined manner, e.g., by weighting the points so that further points have smaller weights. *k*-order α-shape, on the contrary, has no predefined weighting scheme and produces the output at every time tick based on interaction of the α-disks with the actual spatial distribution of the local data. *k*-order α-shape method is robust: small variations in the location of input points do not influence the resulting shape (Fig. 3, right). Indeed, at each time tick the output of our algorithm is defined by the points that made the upper and the lower disks stop. Variations in positions of the other points (both inside and outside the disks) are not crucial for the output at the time tick.

The overall algorithmic flow of the processing is presented in Fig. 4. Its individual steps are discussed above. We made available our MATLAB code for the above algorithm at http://www.cs.helsinki.fi/group/compgeom/seism.zip. The code reads local bulletin, loads raw data files in mseed format, computes envelopes of records for the specified time intervals, runs the *k*-order α-shape algorithm and outputs the restored coda shape of the local event. All the procedures are fully automated; the user can also adjust the code parameters to suit her particular application. More detailed instructions are available in the code archive or by contacting the authors. Finally, an example for visual comparison of averaged and α-shaped envelopes is given in Online Supplementary Materials (Fig. S1).

3.2 Application to synthetic data

We tested the performance of *k*-order α-shape on a synthetic problem instance in which the true shape is known in advance and can be compared with the output of the algorithm. We designed the dataset to resemble real seismic data. The dataset consisted of two parts: the straight line at a constant amplitude level followed by the 3-D solution of time-dependent Boltzmann equation taken as a realistic example of synthetic coda envelope (Paasschens 1997). The latter part is the true shape that *k*-order α-shape has to reconstruct. It essentially consists of the ballistic peak followed by a tail due to waves which have undergone a single forward scattering event, and the diffusion-type tail (see e.g. (Hoshiba 1995) or (Calvet & Margerin 2013) for illustrations of the coda shape just after the direct wave). We then superposed Rayleigh-distributed fluctuations (Nakahara & Carcolé 2010; Anache-Ménier et al. 2009) all through the dataset. The created synthetic trace was 45 min long at sampling frequency of 100 Hz giving a total of 270000 data points (just like real data we will consider below).

By way of proof of concept, we applied the above *k*-order α-shape algorithm to the created synthetic trace and obtained the output curve. The results are presented in Fig. 5. It can be seen the output curve almost coincides with the true shape without prominent shifts or excursions all through the preceding noise, single scattering region and the diffusion tail (Fig. 5, left). The only shape detail somehow distorted by the processing is the ballistic peak due to direct arrival – we observe a sort of "shoulder" instead (Fig. 5, right). To quantify the difference between the output curve and the true shape we calculated the root-mean-square error of the reconstruction, which made few tenths of percent.

3.3 Real data trial

More testing was performed then on real data. It was necessary because the above generated synthetic dataset was missing some important features of the real seismic noise. For example, it does not take into account any temporal correlation in the noise inherently present in any seismic data, or low frequency variations in the mean noise level. To evaluate our algorithm on real data

we processed a broadband record of local Fennoscandian event. The crust earthquake with M_L = 2.4 was registered 109 km away by temporal station LP23 of LAPNET network. Prior to application of the algorithm, we increased signal to noise ratio by frequency filtering between 1 and 15 Hz, evaluated envelope of the filtered trace and took logarithm of the envelope. Figure 6 presents the shape reconstruction output from the trace. In contrast to wild oscillations of the input envelope, the output curve is a function smoothly decaying from the first arrival peak to the level of seismic noise and drawing low frequency waves around this level before and after the event.

3.4 Error analysis

To evaluate standard error and stability of the output curve we used a bootstrap-type resampling algorithm (Efron & Tibshirani 1991). For this purpose we first created multiple (N) realizations of the original envelope. In each realization we randomly removed 25 per cent of data points while keeping synchronization of the bootstrapped pseudo-envelope. The latter means that some time ticks become "empty", i.e. have no corresponding envelope value. Such lacunas momentarily and locally decrease the original sampling rate and make the re-sampled bootstrapped dataset unevenly spaced.

Each of the N created pseudo-envelopes was processed with the above algorithm, which yielded the population of N output curves. As a compromise between statistical significance and computational efficiency we used N = 50; using a somewhat different value of N does not significantly affect our results. Figure 6 plots overlaid the whole population of 50 bootstrapped solutions and the unperturbed output curve. Visually, all through the curve the data points of the processed pseudo-envelopes are evenly distributed around the original output curve and make no local clusters; the unperturbed curve doesn't systematically deviate from the assumed imaginary center of the "tube" formed by the population.

To quantify error, at each time tick we calculated the relative standard deviation of the whole population of output datapoints. Except for the time period corresponding to arrival of primary waves, the error was well below 1% (Fig. 7). In Figure 8 we present distributions of error below and above 5%. The total number of time ticks where the error was above 5% made 767, which was just 0.5% of the total length of the trace. All these points occurred in the region of the sharp growth of envelope amplitude associated with first arrival of P. Restoring coda shape in this region of abrupt changes is beyond the scope of this paper and irrelevant for coda studies.

The above considerations justify the use of *k*-order α-shape for extraction of coda shape from seismograms. In the following section we report on the results of application of *k*-order α-shape reconstruction tool to determining one of important coda characteristics - the coda duration.

4 Coda duration measurement: a concrete application

In this section we report on an application of *k*-order α-shape algorithm to measurement of total duration of seismic coda after a local event. The total coda duration in seconds is usually measured between the P-wave arrival and the time when S-coda amplitudes decrease to the level of microseisms (Bath 1981). The relation between logarithm of coda duration, epicentral distance and local magnitude is important in studies of local seismicity and is usually derived for each region individually (Soloviev 1965). Here we measure coda durations of five local events occurred in Fennoscandia and compare the outputs with previously obtained results for this region.

The list of analyzed events with relevant source parameters from Institute of Seismology of the University of Helsinki (www.seismo.helsinki.fi/english/bulletins) is given in Table 1. Duration of

seismic coda after a local earthquake was measured on vertical records of broadband instruments of permanent and temporal Finnish stations installed under auspices of LAPNET/POLENET project. Seismic stations were equipped with the same unified instruments, which eliminated inconsistency due to equipment diversity. The epicentral distance to seismic stations varied from 5 to 516 km. Each event was registered by at least 10 nearby stations, and all traces were processed in exactly the same way independently of epicentral distance or magnitude.

Each record was processed as follows. To increase signal to noise ratio we bandpass filtered each vertical trace between 1 and 15 Hz. Such frequency filtering enabled to visualize the picture of coda decay. Then we hand-pick the onset of the first arrival of P (alternatively, the onsets can be taken from the source bulletin). Each onset was used as a start time from which the event coda duration was measured on the record. We then calculated analytic envelopes and applied the *k*-order α-shape algorithm to measure the total duration of seismic coda. The analyzed time period of traces spanned 15 minutes prior to first arrival of P and half an hour past it, which is sure to include the whole coda of any local event.

Our dataset includes records of weaker events whose seismic coda length is notoriously hard to measure. Indeed, the analyzed events with $M_L < 2.0$ from the above list were almost unobservable by eye on raw records of more distant stations, so application of moving averaging to such small events may virtually smooth out the weak signal that barely stands above the background seismic noise even after frequency filtering.

We identified the coda end as the time instant when the output curve first drops below the mean level of the preceding background noise (although background seismic noise is not ideal and reveals manifold biases and anomalies, its mean level over the time period immediately preceding the event is frequently used as a natural benchmark). That is, we used the simplest definition of coda duration as the time elapsed from first arrival of primary waves until the coda falls beneath some absolute amplitude level. In general such approach is not easy to implement, in particular due to rapid oscillations of the analyzed time series (either the original seismogram or its envelope). To overcome it, there have been presented a number of tricks involving signal-to-noise amplitude ratios, relative amplitude measurements, etc. In contrast, our output curve representing the coda shape is a smoothly decaying almost non-oscillating function that supports uncomplicated time measurements on the base of simple criteria.

An example of output curves along with measured coda durations are shown in Fig. 9. The 14 restored shapes of seismic codas recorded at epicentral distances between 5 and 308 km from a local event (#1 in Table 1) are displayed overlaid and aligned onto the origin time. The presented shapes are consistent with theoretical predictions as to constant decay gradient and the total time required to dump seismic energy radiated from a source. According to Fig. 9, variability in this latter parameter estimated by our automatic procedure is roughly about 25 seconds, which is not large for a weak event.

To compare total coda durations with previous results we turned to the notion of duration magnitude. The linear relationship between logarithm of seismic coda length $\boldsymbol{\tau}$ and epicentral distance $\boldsymbol{\Delta}$ in km to the local event with magnitude M is most commonly given by

$$\boldsymbol{M} = a*\boldsymbol{\Delta} + b*\log_{10}\boldsymbol{\tau} + c \qquad (1)$$

or

$$\boldsymbol{M} = a*\boldsymbol{\Delta} + b*\log^2{}_{10}\boldsymbol{\tau} + c \qquad (2)$$

where *a*, *b* and *c* are constants. The values of *a*, *b* and *c* for a given station or region are usually estimated using the regression analysis of extensive datasets of τ versus Δ and are subject to significant changes from one region to another. For a given magnitude the linear relationship between logarithm of τ and Δ is defined by the ratio *a/b*. This ratio for formula (1) was estimated

as -1.2×10^{-3} for Japan (Tsumura 1967) and -3.4×10^{-4} for Northern California (Hirshorn et al. 1987). The Fennoscandian magnitude scale on the base of coda duration was derived by Wahlstrom (1980) who used formula (2). The Fennoscandian linear relationship coefficient *a/b* in the formula was estimated as -2.9×10^{-3}, which is not far from the analogous scale for Central California yielding -4.4×10^{-3} (Bakun 1984).
In order to estimate output quality of our automatic coda duration measurements, for each event from our dataset we plotted squared logarithm of total coda length in seconds versus epicentral distance in km. The results for two events with the smallest and the largest magnitudes (M = 1.4 and M = 2.4, respectively) are given in Fig. 10. The plots clearly display the expected dependence of coda duration decrease with epicentral distance. The last column of Table 1 shows the relevant linear regression estimates for all 5 events. The presented values of linear relationship coefficient *a/b* exhibit some variation, but the average value made $(-3.6 \pm 1.0) \times 10^{-3}$, which is very close to the previous estimate by Wahlstrom (1980).

5 Conclusion and future work

In this paper we presented a new algorithm for restoring seismic coda shape from analytic envelopes. We successfully applied the algorithm to measuring total coda length on local records of weak events. The relation coefficient between coda duration and epicentral distance is consistent with the earlier findings for the region.
The new algorithm is based on α-shape - a well established tool for shape reconstruction, that previously has been successfully applied in various domains. We demonstrate how the new tool provides robust non-parametric estimation of shape of a point cloud given as time series. A useful feature of our algorithm is that it is not oriented towards any specific absolute criterion of output quality, so in any particular application the user is free to choose the most suitable quality measure.
We envision that *k*-order α-shapes will be applied to other seismological tasks such as duration and local magnitude estimations, studies of seismic attenuation derived from coda decay and fine temporal variations in coda shape. It would also be interesting to see if *k*-order α-shape is applicable to other geophysical datasets (e.g. geomagnetic).

Appendix

To choose α and k for each trace we use the part of the record that contains background noise only (Fig. 4). We set α equal the standard deviation of the noise. To choose k we take *1000* random time ticks and at each tick place two α-disks – one with the center 2α above the mean level of noise, and the other with the center 2α below the mean level; k is taken as the average number of data points over the *2000* disks. We then apply the *k*-order α-shape algorithm to the noise, and measure the similarity between the output shape and the mean level using root-mean-square deviation (RMSD). Intuitively, the α-disks are expected to "shave off" the noise so that a straight line is obtained in the output. If the variability of the output curve is too high, we scale the time axis and recompute k with the same procedure (averaging over the 2000 disks at random time ticks). When larger scaling is applied to the time axis, the chosen value for k also becomes larger, which implies a smaller-variability output curve (decreased RMSD). One stops at the smallest scale for which the RMSD gets below the pre-defined level. RMSD between 0.05α and 0.005α usually suffices. Using higher scales would lead to an oversmoothed output.
Overall, the described method follows the standard paradigm in artificial intelligence and machine learning: splitting data into training and validation sets. We emphasize that α and k are chosen separately for each trace, i.e. these operational parameters are adjusted individually for

each record in question. Furthermore, for any particular record, the choice of α is independent of the time axis scaling, while k is chosen separately for each scale. We also remark that RMSD is not the only possible measure of quality of noise fit; any other goodness measure can be used just as well.

Acknowledgements

The authors wish to thank LAPNET/POLENET Workgroup headed by Elena Kozlovskaya from Sodankylä Geophysical Observatory for providing high-quality broadband seismic records of LAPNET seismic array. We also thank the anonymous reviewers for their helpful input and researchers at Helsinki Institute of Seismology for discussions. This work was supported by the Academy of Finland grants 1138520 and 261019 and by Research Funds of University of Helsinki.

The POLENET/LAPNET project was a part of the International Polar Year 2007-2009 and a part of the POLENET IPY consortium. Organisations participated in the POLENET/LAPNET project are: (1) Sodankylä Geophysival Observatory of the University of Oulu(Finland), (2) Institute of Seismology of the University of Helsinki (Finland), (3) University of Grenoble (France), (4) University of Strasbourg (France), (5) Institute of Geodesy and Geophysics, Vienna University of Technology (Austria)j, (6) Geophysical Institute of the Czech Academy of Sciences, Prague (Czech republic), (7) Institute of Geophysics ETH Zürich (Switzerland), (8) Institute of Geospheres Dynamics of the Russian Academy of Sciences, Moscow (Russia), (9) The Kola Regional Seismological Centre, of the Russian Academy of Sciences (Russia), (10) Geophysical Centre of the Russian Academy of Sciences, Schmidt Institute of Physics of the Earth of the Russian Academy of Sciences (Russia), (11) Swedish National Seismological Network, University of Uppsala (Sweden), (12) Institute of Solid Earth Physics, University of Bergen (Norway), (13)NORSAR (Norway), (14) University of Leeds (UK). Equipment for the temporary deployment was provided by RESIFSISMOB, FOSFORE, EOST-IPG Strasbourg Equipe sismologie (France), seismic pool (MOBNET) of the Geophysical Institute of the Czech Academy of Sciences (Czech republic), instrument pool of Sodankylä Geophysical Observatory (Finland), Institute of Geosphere Dynamics of RAS (Russia), Institute of Geophysics ETH Zürich (Switzerland), Institute of Geodesy and Geophysics, Vienna University of Technology (Austria), University of Leeds (UK). Funding agencies provided support for organisation of the experiment are: Finland : The Academy of Finland (grant No. 122762) and University of Oulu; France: BEGDY program of the Agence Nationale de la Recherche, Institut Paul Emil Victor and ILP (International Lithosphere Program). task force VIII.; Czech Republic: grant No. IAA300120709 of the Grant Agency of the Czech Academy of Sciences; Russian Federation : Russian Academy of Sciences (programs No 5 and No 9). POLENET/LAPNET Working Group: Elena Kozlovskaya (1), Helle Pedersen (3), Jaroslava Plomerova (6), Ulrich Achauer (4), Eduard Kissling (7), Irina Sanina (8), Teppo Jämsen (1), Hanna Silvennoinen (1), Catherine Pequegnat (3), Riitta Hurskainen (1), Robert Guiguet (3), Helmut Hausmann (5), Petr Jedlicka (6), Igor Aleshin (10), Ekaterina Bourova (3), Reynir Bodvarsson (11), Evald Brückl (5), Tuna Eken (6), Pekka Heikkinen (2) Gregory Houseman (14), Helge Johnsen (12), Elena Kremenetskaya (9), Kari Komminaho (2), Helena Munzarova (6) Roland Roberts (11), Bohuslav Ruzek (6), Hossein Shomali (11), Johannes Schweitzer (13), Artem Shaumyan (8), Ludek Vecsey (6), Sergei Volosov (8)

Table 1. Source parameters of analyzed events and relation coefficients (*a/b*) between coda duration and epicentral distance estimated by least squares for each event.

#	Origin		Source Location and Magnitude				# of stations	a/b
	Date yyyy/mm/dd	Time hh:mm:ss.0	latitude (degrees)	longitude (degrees)	depth (km)	M_L		
1	2008/09/27	02:20:11.9	65.987	29.963	18	2.4	31	$(-2.2 \pm 0.4)x10^{-3}$
2	2007/11/13	03:29:07.8	66.041	29.519	17.1	1.7	13	$(-5.2 \pm 1.0)x10^{-3}$
3	2008/01/17	03:46:43.2	65.796	29.866	14.2	1.7	12	$(-3.9 \pm 1.1)x10^{-3}$
4	2008/08/10	19:02:28.9	65.592	27.866	14.8	1.6	10	$(-3.9 \pm 1.4)x10^{-3}$
5	2008/04/06	05:12:05.1	65.876	29.868	9.8	1.4	10	$(-3.0 \pm 0.6)x10^{-3}$

References

Abbott, A., Tsay, A., 2000. Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect, *Sociological Methods & Research*, **29**, 3–33.

Albou, L. P., Schwarz, B., Poch, O., Wurtz, J.M., Moras, D., 2009. Defining and characterizing protein surface using alpha shapes, *Proteins*, **76**(1), 1-12.

Aki, K., 1969. Analysis of seismic coda of local earthquakes as scattered waves, *J. Geophys. Res.,* **74**, 615—631.

Anache-Ménier, D., van Tiggelen, B.A. & Margerin, L., 2009. Phase Statistics of Seismic Coda Waves, *Phys. Rev. Lett.,* **102**, #24, 248501.

Bakun, W. H., 1984. Magnitudes and Moments of Duration, *Bull. Seismol. Soc. Am.,* **74** (6), 2335–2356.

Bakun, W. H., 1984. Seismic Moments, Local Magnitudes, and Coda-Duration Magnitudes for Earthquakes in Central California, *Bull. Seismol. Soc. Am.,* **74**, 439–458.

Bath, M., 1981. Earthquake magnitude - recent research and current trends, *Earth Science Rev*., **17**, 315-398.

Calvet, M. & Margerin, L., 2013. Lapse- Time Dependence of Coda Q: Anisotropic Multiple- Scattering Models and Application to the Pyrenees, *Bull. Seism. Soc. Am.*, **103** (3), 1993-2010.

Edelsbrunner, H., Kirkpatrick, D. G. & Seidel, R., 1983. On the shape of a set of points in the plane, *IEEE Trans. Inform. Theory*, **29**, 551–559.

Edelsbrunner, H., Mucke, E. P., 1994. Three-dimensional alpha shapes, *ACM Trans. Graph.* **13**(1), 43—72.

Efron, B., Tibshirani, R., 1991. Statistical data analysis in the computer age, *Science*, **253**, 390–395.

Hirshorn, B., Lindh, A., Allen, R., 1987. Real Time Signal Duration Magnitudes from Low-gain Short Period Seismometers, *U.S. Geol. Surv., Open-File Rep,* 87–630.

Hoshiba, M., 1995. Estimation of nonisotropic scattering in western Japan using coda wave envelopes: Application of a multiple nonisotropic scattering model, *J. Geoph. Res.,* **100**, B1, 645-657.

Krasnoshchekov, D., Polishchuk, V., 2014. Order-k α-hulls and α-shapes, *Information*

Processing Letters, **114**, 76—83.
Krasnoshchekov, D., Polishchuk, V., Vihavainen, A., 2010. Shape approximation using k-order alpha-hulls, *Symposium on Computational Geometry*, 109-110, http://youtu.be/KeRKgv4Slts
Nakahara, H., Carcolé, E., 2010. Maximum-Likelihood Method for Estimating Coda Q and the Nakagami-m Parameter, *Bull. Seismol. Soc. Am.*, **100** (6), 3174–3182, doi: 10.1785/0120100030.
Paasschens, J.C.J., 1997. Solution of the time-dependent Boltzmann equation, *Phys. Rev.*, E56, 1135–1141.
Rautian, T.G., Khalturin, V.G., 1978. The use of the coda for the determination of the earthquake source spectrum, *Bull. Seismol. Soc. Am.*, **68**, 923–948.
Sato, H., Fehler, M.C., 1998. *Seismic wave propagation and Scattering in the homogeneous Earth*, AIP Press, New York.
Soloviev, S.L., 1965. Seismicity of Sakhalin, *Bull. Earthq. Res. Inst.*, **43**, 95—102.
Teichmann, M., Capps, M. V., 1998. Surface reconstruction with anisotropic density-scaled alpha shapes, in *IEEE Visualization*, 67—72.
Tsumura, K., 1967. Determination of Earthquake Magnitude from Total Duration of Oscillation, *Bull. Earthquake Res. Inst.,* Tokyo Univ., **15**, 7–18.
Wahlström, R., 1980. Duration Magnitudes for Swedish Earthquakes, *Geophysica*, **16**, #2, 171-183.
Zhou.W., Yan, H., 2012 Alpha shape and Delaunay triangulation in studies of protein-related interactions, *Briefings in Bioinformatics*, doi: 10.1093/bib/bbs077.

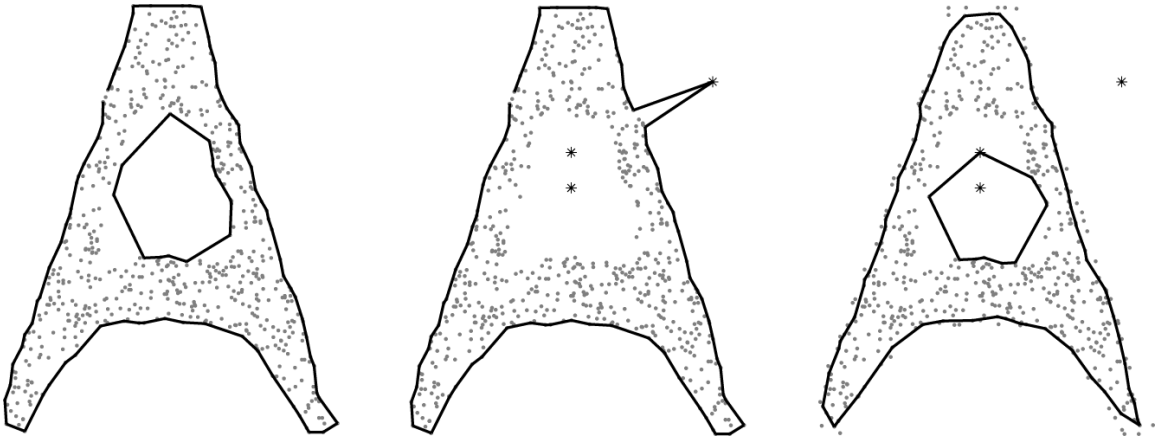

Figure 1. Left: A point set and its α-shape. Middle: Few outliers (shown with asterisks) distort the restored shape. Right: *k*-order α-shape ($k = 1$) of the set with outliers brings the shape back.

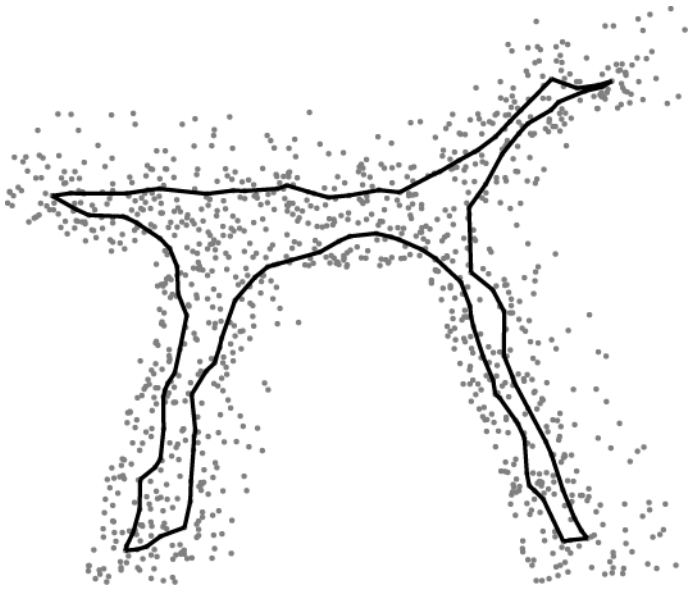

Figure 2. The *k*-order α-shape (k = 15) effectively restores the inner shape of the data cloud.

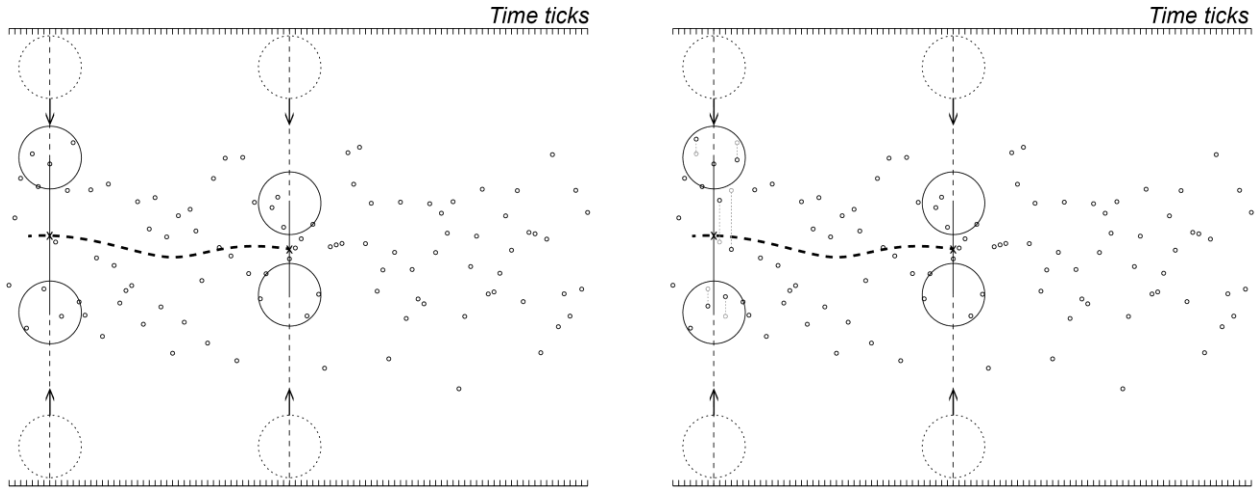

Figure 3. Application of *k*-order α-shape to time series. Left: The upper (resp. lower) disk is pushed down (resp. up) from above (resp. below) the data; the disk stops when there is exactly *k = 4* points inside it. The output at the time tick is the average of the disks' centers. Right: some points (both inside and outside the disks) have changed their locations slightly; nevertheless. the output at the time tick does not change.

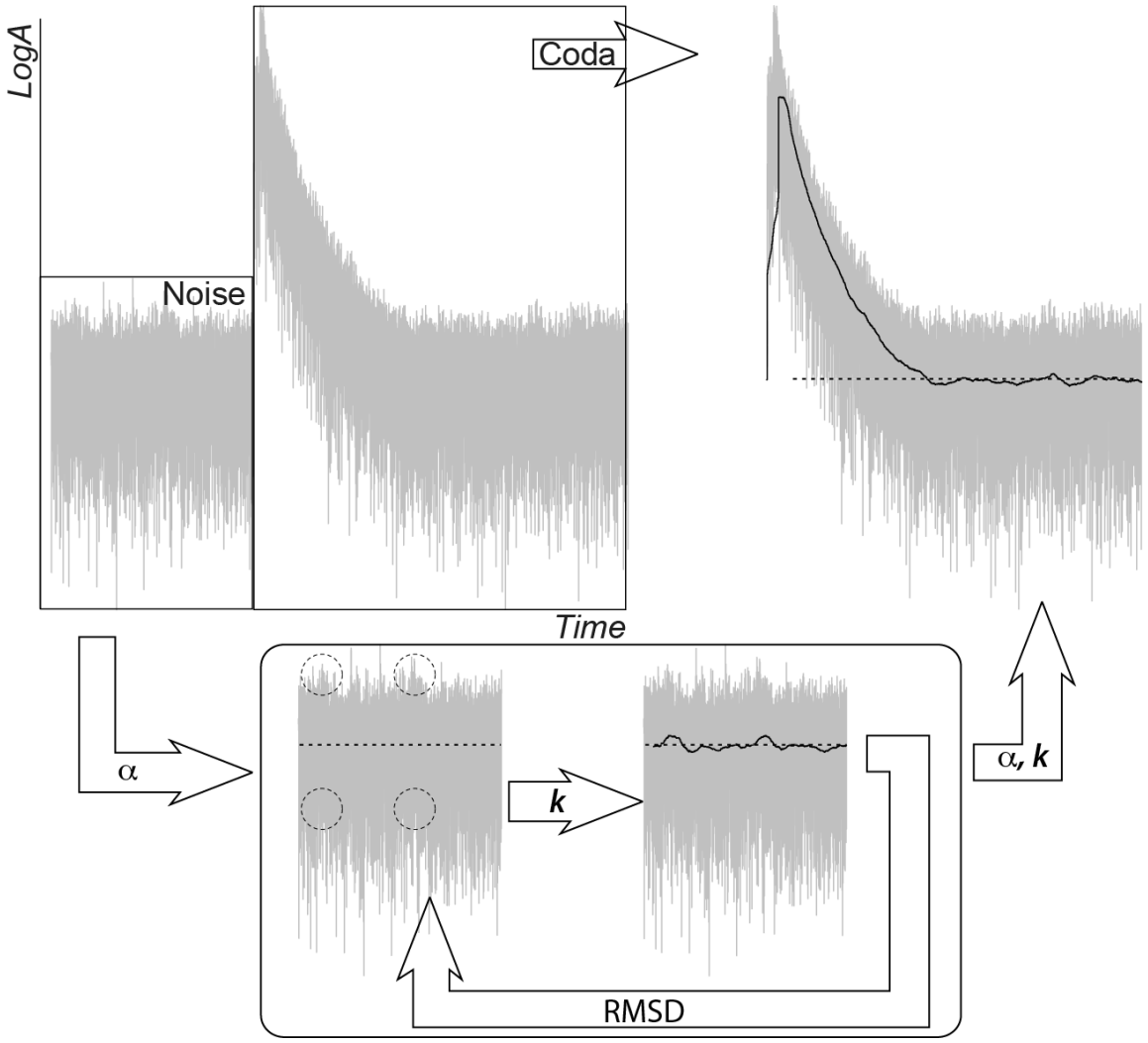

Figure 4. The overall algorithmic flow: Seismic record is split into the noise and coda. The noise is used to choose α and k (e.g. with the help of RMSD – see Appendix). The chosen α and k are applied to coda to restore its shape. The dashed line is the mean level of background noise (also shown are four disks at random time ticks); the output of the algorithm is in black.

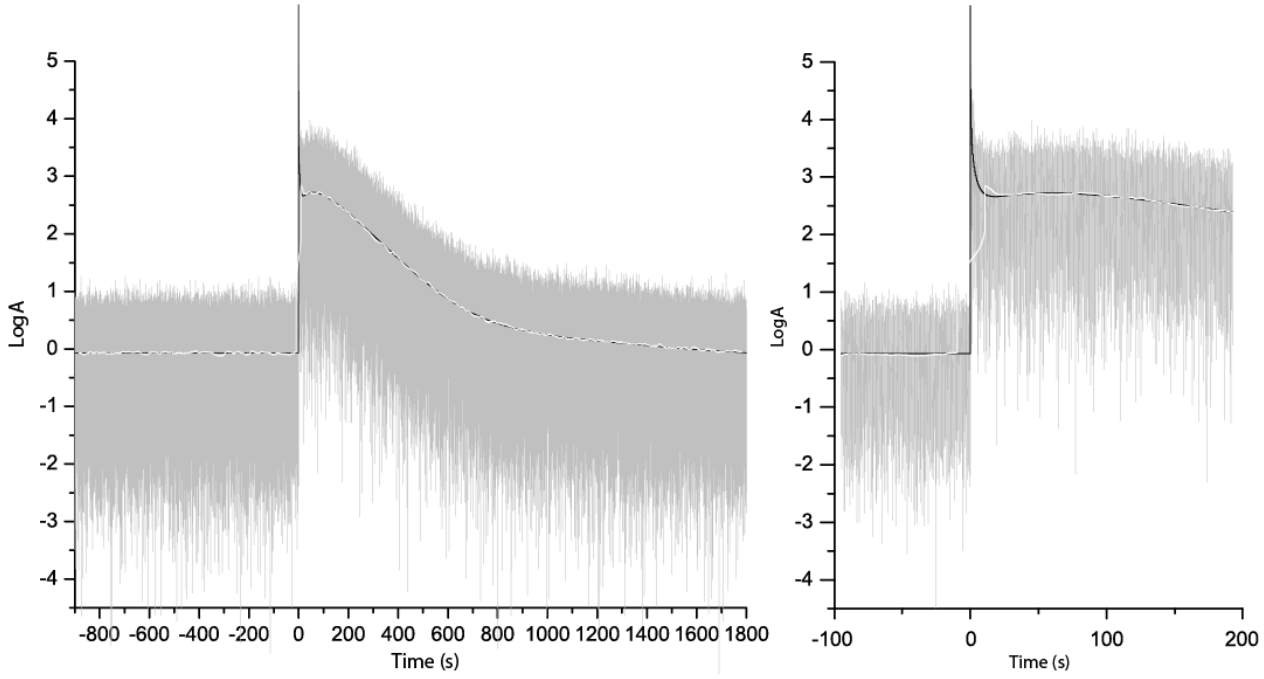

Figure 5. Application of *k*-order α-shape to synthetic example: the result of processing the whole 45-min long synthetic dataset (left), and the 5-min excerpt corresponding to the passage that represents the first arrival of P (right). In both panes the true shape (black) and the restored shape (white) are overlaid on the synthetic dataset (grey) with Rayleigh distributed noise.

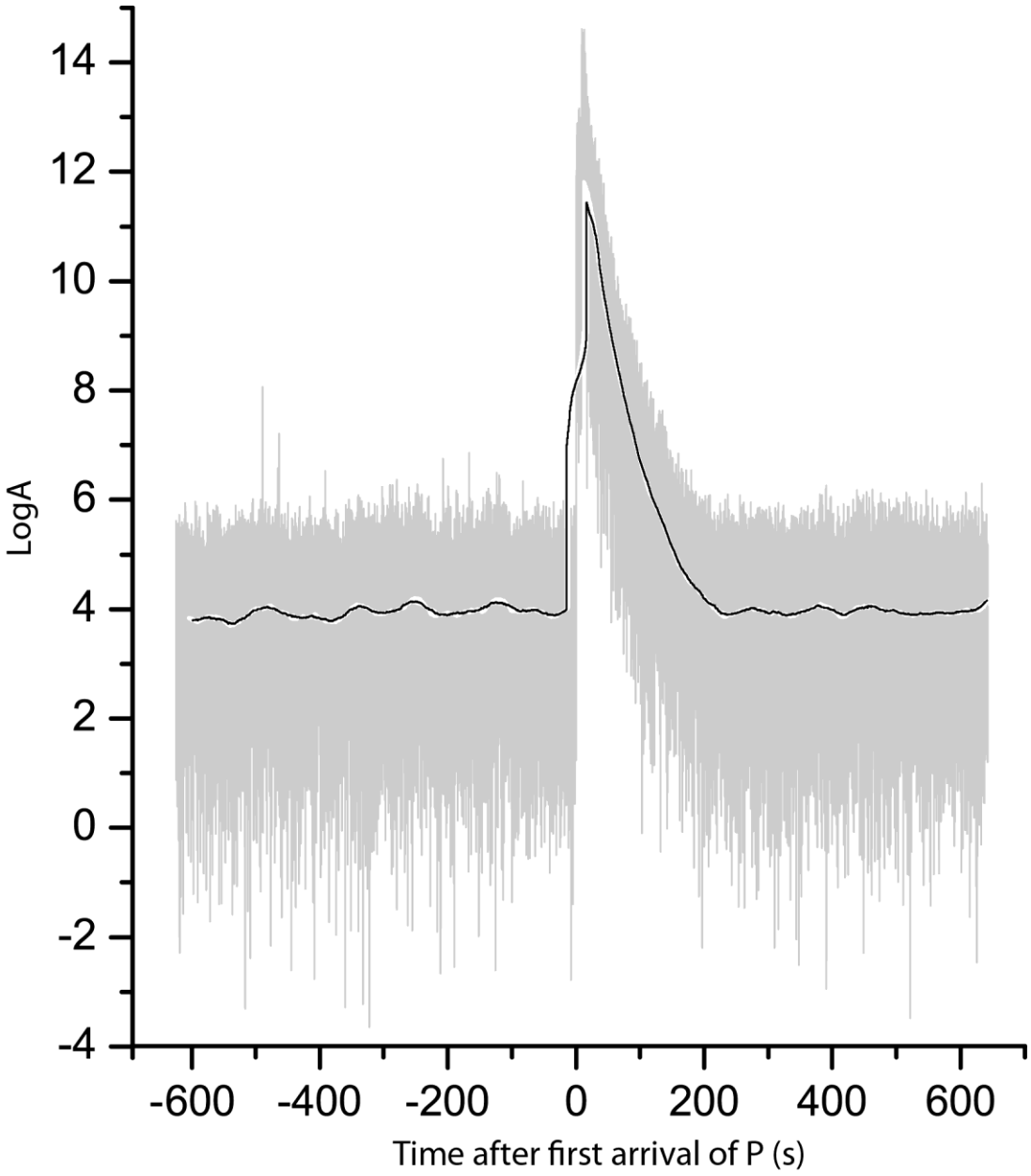

Figure 6. Error analysis: the unperturbed output (black) and 50 bootstrapped solutions (white) overlaid on the original analytic envelope (grey).

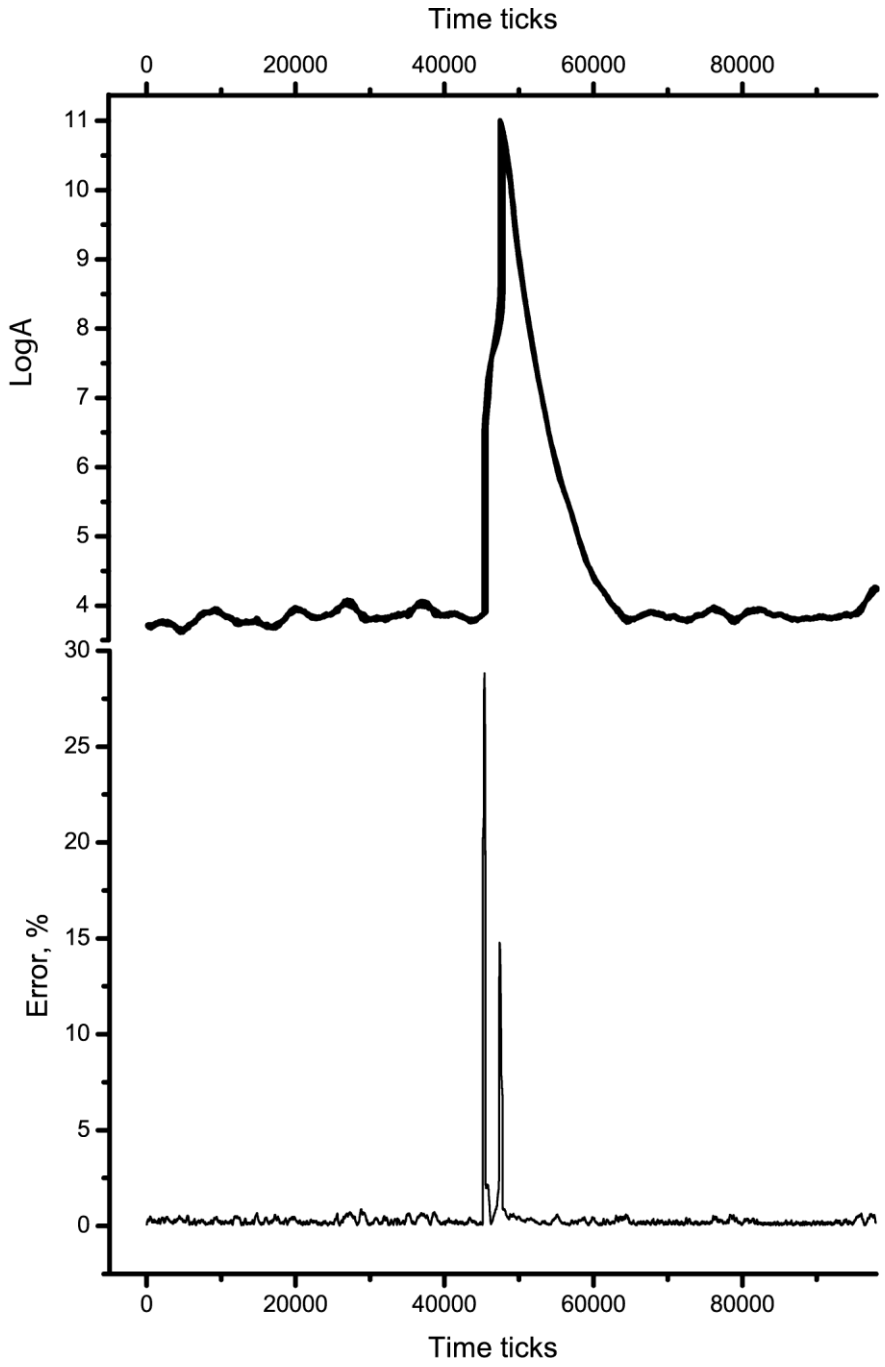

Figure 7. Quantifying the error: synchronized plot of bootstrapped population of output curves (above) and relative standard deviation (below) versus time ticks. The range of higher error correlates with first arrival of P-waves and comes as a "shoulder" on the output curve.

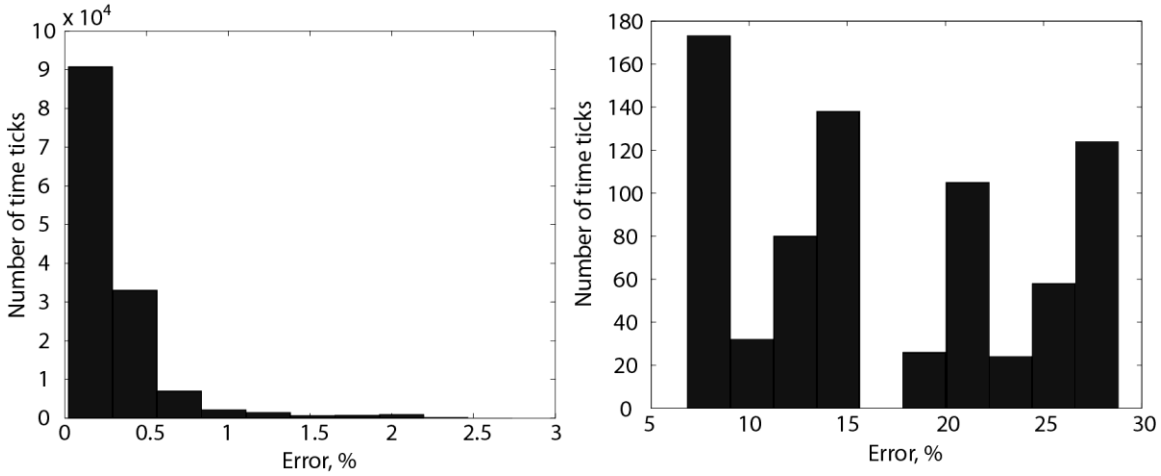

Figure 8. Histogram of error below (left) and above (right) 5%.

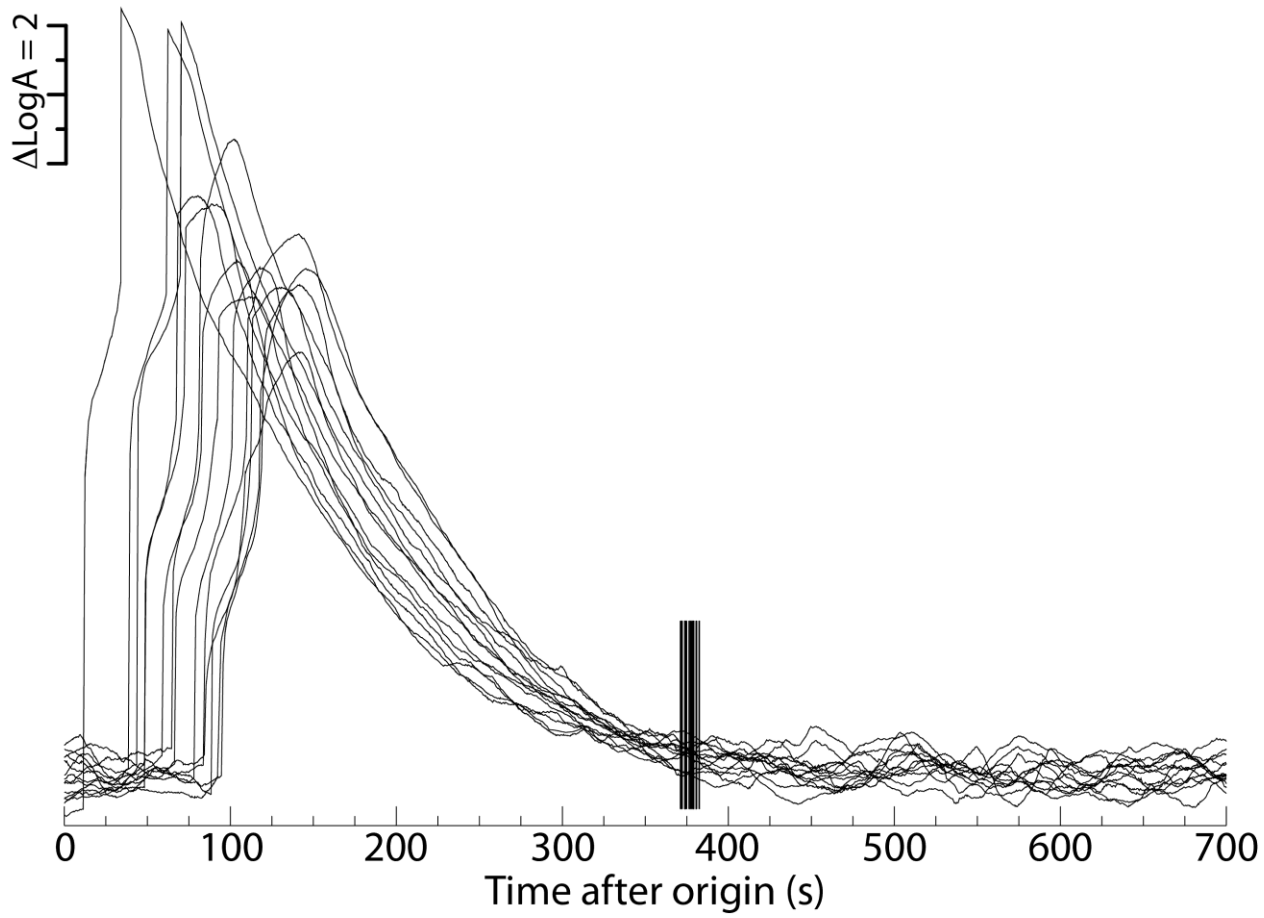

Fig. 9. Results of coda shape restoration procedure applied to 14 envelopes of a local event (see event #1 in Table 1 for seismic source parameters).

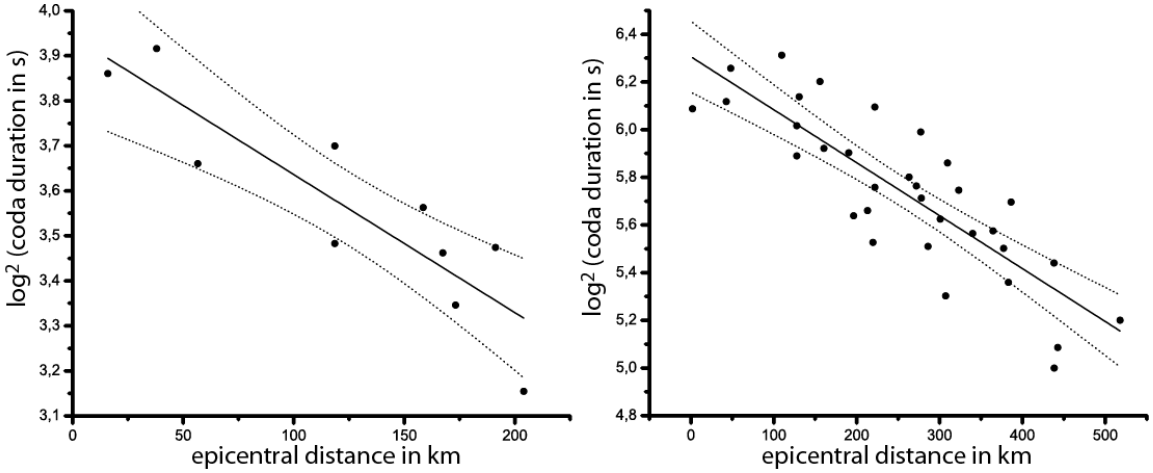

Fig. 10. Linear regression results for events #5 (left) and #1 (right) from Table 1.