Identification of response and timing issues at permanent European broadband stations from automated data analysis

C. Weidle, R. A. Soomro, L. Cristiano, and T. Meier

Institute of Geosciences, Kiel University, Otto-Hahn-Platz 1, 24118 Kiel, Germany

Correspondence to: C. Weidle (cweidle@geophysik.uni-kiel.de)

Abstract. We apply automated two-station broadband phase velocity dispersion measurements to all available broadband data from permanent seismic stations in Europe, as available through the European Integrated Data Archive (EIDA, http://www.orfeus-eu.org/eida/) infrastructure. As part of our quality control we detect several typical patterns in our measurements that can be related to technical problems, incorrect metadata information or uncover inconsistencies in data processing routines. These effects include timing and various response issues, most prominently erroneous response information. Our procedure is thus able to identify potentially problematic (meta)data from a large set of seismic data and offers an applicable way to increase data quality at data centers.

1 Automated data analysis

To tackle the ever increasing amount of available broadband seismic data from European seismic stations for routine analysis, manual data processing and retrieval of observables needs to be replaced by automated processing tools. Furthermore, rapid response in case of seismic events requires automated data processing. While signal processing for detecting and locating seismicity is routinely done on an automated basis, seismogram analysis for structural interpretations often still requires manual screening. As locations of earthquakes as well as studies of the Earth's structure depend heavily on the quality of the data, automated routines are required for data quality control as well. While power spectral density estimates of the incoming data are well suited to detect data gaps, large errors in the gain information or noisy stations, the detection of response and timing problems is less evident

to be detected automatically and few works have addressed this problem in the past.

Gibbons (2006) identified and documented a temporary timing problem of $< 20\,\mathrm{s}$ at a single station using repeatedly occuring mining events. Identification of the issue was, however, rather accidental as the timing problem occurred at the time of a particular event of interest and was rather large ($\sim 8\,\mathrm{s}$). Although the author points out that there exist "most likely [sources of repeating seismic signals] in the vicinity of many seismic stations", a systematic implementation of his routine to all available seismic data in Europe would require a massive effort of the European seismological community.

We developed an automated routine to measure interstation phase velocity curves of fundamental mode surface waves, including automated quality checks of the raw data (Soomro et al., 2013). The measurements are done by pairwise cross-correlating instrument corrected velocity seismograms and rely on path-specific reference models.

The method is applied to all available data from permanent stations in Europe since 1990, as provided by EIDA, the European Integrated Data Archive (http://www.orfeus-eu.org/eida/). We use an automated routine based on ObsPy (Beyreuther et al., 2010) to download this massive dataset.

Besides an unprecedented amount of regional scale two-station dispersion measurements, which is to be exploited by surface wave tomography, we observe in our dataset as part of our quality control a variety of phenomena that are indicative of technical issues related to the data, metadata and/or processing. In the following we will discuss in Sect. 2 potential expectable issues by simulating them in synthetic examples. In Sect. 3 we present some data examples of these phenomena and preliminary details on few stations which either have technical issues or where the provided station metadata is incorrect. To identify stations with odd behaviour in
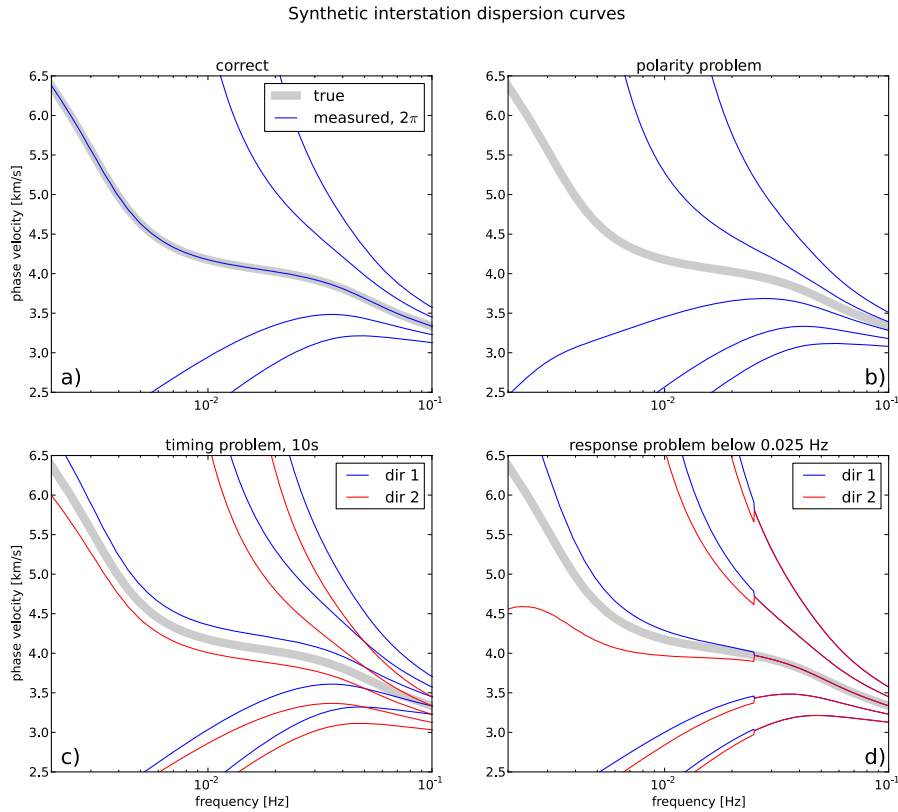
Synthetic interstation dispersion curves

Fig. 1. Simulation of potential data errors in the inter-station phase velocity measurement. Synthetic waveforms are simulated in PREM (Dziewonski and Anderson, 1981) for epicentral distances of 5000 and 6000 km. For description see text.

our automated measurements we finally present a map of stations that are suspicious in the dispersion measurements and require a manual check to exclude or identify errors in data or metadata information.

2   Method and synthetic example

Very broad-band dispersion curves of fundamental modes may be measured by cross-correlation of recordings at two stations. Frequency dependent time windows are applied to the cross-correlation function before determination of phase velocity from the phase of the cross-correlation function in order to optimize resolution in the time-frequency plane. The correct 2π phase shift is identified by comparing interstation phase velocities to path-specific reference models based on CRUST2.0 and PREM (Bassin et al., 2000; Dziewonski and Anderson, 1981). Furthermore, the smoothness of the dispersion curves is carefully checked and results for both propagation directions are averaged and compared. A detailed account of the measurement procedure is given by Soomro et al. (2013). The procedure may be applied to earthquake data as well as ambient noise.

To understand how potential data and metadata errors affect our two-station waveform analysis, we calculate syn-

thetic waveforms at 5000 and 6000 km epicentral distance in PREM (Dziewonski and Anderson, 1981) and simulate three potential issues (Fig. 1):

– If data and metadata information are correct – the standard case – one 2π branch will fall close to the reference model and will be selected to determine the acceptable bandwidth of the measurement. Other 2π branches are neglected.

– If one of the stations has wrong polarity, an additional π-phase shift is introduced, and the resulting measurements are symmetrically offset from the expected reference model. Our measurements may thus fall above and/or below and have a rather large deviation from the reference model.

– If one of the stations has a timing problem, there will be an approximately constant offset from the expected reference curve. As we measure propagation in both directions, velocities will differ for the two propagation directions.

– If one of the stations has a response problem, simulated here by a $\pi/8$ error for periods above 40 s, the measurements will branch off the expected curve for low
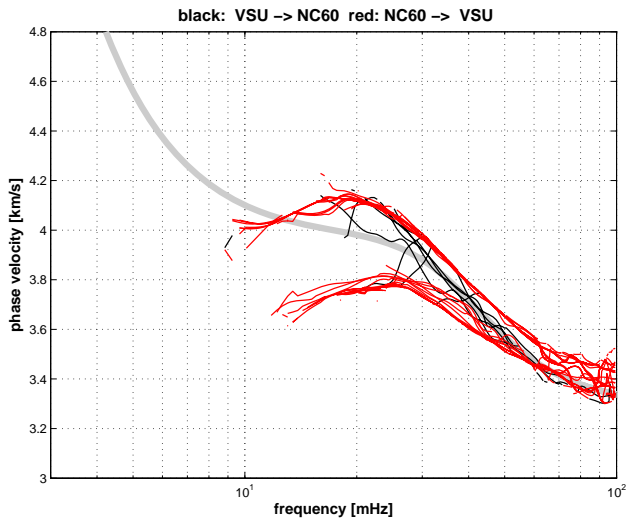
black: VSU –> NC60 red: NC60 –> VSU

Fig. 2a. Automated phase velocity measurements between station NC602 in southern Norway and VSU in Estonia. Curves show signature of a polarity issue with π offset from the expected reference curve.

black: FUR –> ASSE red: ASSE –> FUR

Fig. 2b. Automated phase velocity measurements between station FUR and ASSE in southern and northern Germany, respectively. Curves show signature of a timing issue with close to constant offset of measured curves for the two propagation directions.

frequencies, again with differing sign for the two propagation directions. The same pattern, however, may similarly reflect differing bandwidths of the two instruments involved.

A commonly debated issue in the context of two-station measurements is deviation of the wavefront propagation direction from the greatcircle by diffraction/scattering at large scale heterogeneities (e.g. Moho topographic variations, oceans, cratons). These effects are indeed observed but
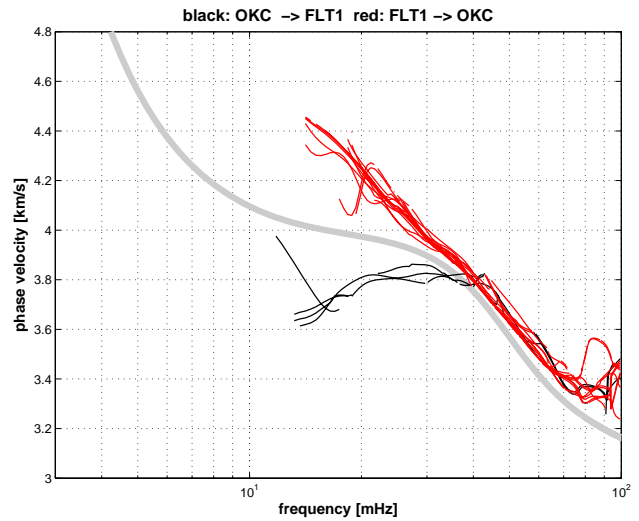

black: OKC –> FLT1 red: FLT1 –> OKC

Fig. 2c. Automated phase velocity measurements between station FLT1 in northern Germany and OKC in the eastern Czech Republic. Curves show signature of a response issue with branching offset of measured curves for the two propagation directions. This may reflect differing instrument types but could also point to wrong instrument response.
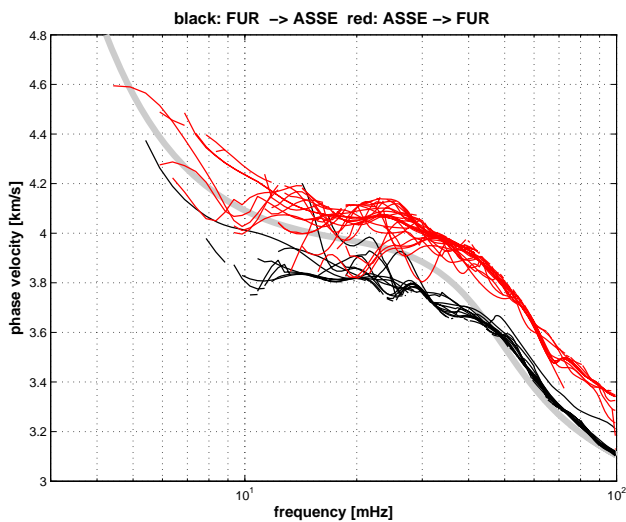
are limited to certain frequency bands. We observe such deviations as "bumps" in the dispersion curve around 20–30 s if strong variations in Moho depth are encountered, or as systematic deviations between ca. 50–80 s if contrasts in lithospheric thickness lead to deflection of the wavefield from greatcircle propagation. These structural effect will, however, always be visible as regional patterns at nearby stations and can thus be easily distinguished from technical issues as discussed here.

3 Data examples

We present some data examples with indications of polarity, timing and response issues at selected station pairs in Fig. 2.

a. Station NO.NC602 appears in our analysis as having a polarity problem (Fig. 2a). In fact, however, it turned out to be a problem with the response information and our application of it in the instrument correction. In the SEED convention (Ahern et al., 2012), the normalization factor A_0 is defined as: $A_0 = 1/|H(s)|$, where $H(s)$ is the transfer function of the instrument as a function of complex frequency s. A_0 is thus, per definition, a positive number and a positive normalization factor is provided by ORFEUS. At station NO.NC602, a Güralp CMG-3T sensor was installed in 2000, with poles and zeros for displacement that have a negative transfer function. Applying a positive normalization constant introduces thus a polarity flip in the instrument corrected data.
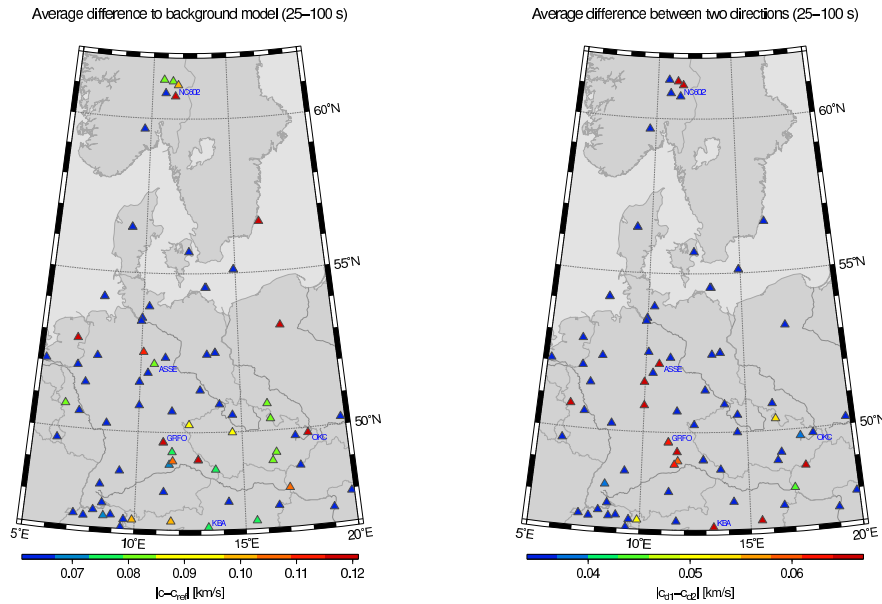
Fig. 3. Averaged deviations of measured dispersion curves in the period range 25–100 s (left) from the reference model and (right) between the two propagation directions. Stations with large standard deviations point to data inconsitencies which need to be manually checked. Stations with insignificant number of measurements are omitted.

The issue was recognized by Güralp in 2006 (Guralp Systems Ltd., 2013), however, responses for NO.NC602 provided through ORFEUS have a new, positive transfer function only for data since June 2011.

b. A timing issue could be suspected at station ASSE in northern Germany. Based on the offset of the measurements for the two directions we estimated the timing error to be on the order of 3 s (see Discussion). Manual data inspection revealed, however, an inconsistency in our data processing where a data decimation routine within ObsPy lead to a phase shift of the data at station ASSE of 3–4 s. This highlights the usefulness of this approach as part of our quality control and the important fact that not every suspiciously behaving station has a (meta)data inconsistency.

c. A response issue may reflect differing bandwidths of the two instruments involved but also indicate potentially erroneous responses. Here, responses supplied by ORFEUS for CZ.OKC and GE.FLT1 indicate both a STS2 seismometer, while in fact, CZ.OKC is a Güralp CMG-3ESP (Czech Regional Seismic Network, 2013).

4 Discussion

To systematically check our measurements for potential station issues we average deviations from the reference model and between the two directions over frequency. These stationwise averages might reflect regional structural variations where our reference models are not accurate enough but in general highlight potentially(!) problematic data. These data require then a detailed manual inspection to identify the issue that causes large deviations of our measurements from the expected reference model.

It is important to stress that this procedure is designed for quality control and an anomalous station in Fig. 3 has not necessarily a technical issue. Besides the example (b) for station ASSE above, we have found at station GRFO, for example, an inconsistency in the applied instrument correction which had a comparable signature in the measurements as a timing or response error (Fig. 2b, 2c).

In principle it is straightforward to estimate the amplitude of a timing or response/phase error. The timing/phase error is related to the offset of the measured curves for two directions by $c_{1,2}(\omega) = \frac{\delta\Delta}{t(\omega)\pm\epsilon_t} = \frac{\omega\delta\Delta}{\Phi(\omega)\pm\epsilon_\Phi(\omega)}$, where $c_{1,2}(\omega)$ is phase velocity in the two directions as function of frequency ω, $\delta\Delta$ the inter-station distance, $t(\omega)$, ϵ_t the inter-station traveltime and timing error ($\epsilon_t \ll t(\omega)$), respectively and $\Phi(\omega)$, $\epsilon_\Phi(\omega)$ the phase and phase error, respectively.

As t increases with $\delta\Delta$, the offset $|c_1 - c_2|$ is strongly distance dependent with larger offsets for shorter inter-station distance. Critical for the determination of ϵ_t is the detectability of a consistent offset between curves in two directions, which we may approximate by the standard deviation of our measurements. Soomro et al. (2013) show that the measurements have an overall standard deviation of $< 2\%$ at

shorter inter-station distance. In the main frequency band of our observation (10–50 mHz), the phase velocity c is approx. $4.0\,\mathrm{km\,s^{-1}}$, hence for an inter-station distance of $500\,\mathrm{km}$ a timing error of ca. $2.5\,\mathrm{s}$ and larger would be detectable. This value is on the same order of magnitude as suggested by Gibbons (2006) to potentially go unnoticed, as timing errors below $\sim 2\,\mathrm{s}$ may be within the range of phase identification and picking uncertainties in local to regional seismic event location.

Phase and timing errors are linked through $\Phi = \omega t$, thus the same line of arguments can be applied to phase errors. Through $\epsilon_\Phi = \omega\epsilon_t$ the traveltime error ϵ_t resulting from a given phase error increases with decreasing frequency. Therefore, the detectability of a phase error is easier at lower frequencies as the curve offset for two directions $|c_1 - c_2|$ increases with decreasing frequency (Fig. 1d).

Since identification of the origin of anomalous dispersion curve behaviour is still ongoing work we refrain from discussing other stations explicitly. However, based on the examples outlined in the previous section we point out that – as part of quality control in our automated dispersion measurement routine – we are able to easily identify stations with potentially technical issues or incorrect metadata information within a large dataset. Though a systematic analysis of all anomalous stations in our dataset, which comprises basically all currently available data in Europe through EIDA nodes, will be timeconsuming, it might be a viable way to uncover data and metadata inconsistencies in the databases.

References

Ahern, T., Casey, R., Barnes, D., Benson, R., and Knight, T.: SEED Reference Manual – Standard for the Exchange of Earthquake Data, International Federation of Digital Seismograph Networks Incorporated Research Institutions for Seismology United States Geological Survey, version 2.4 edn., http://www.iris.edu/manuals/SEED_chpt1.htm, 2012.

Bassin, C., Laske, G., and Masters, G.: The Current Limits of Resolution for Surface Wave Tomography in North America, in: EOS Trans AGU, vol. 81 (48), p. F897, http://mahi.ucsd.edu/Gabi, 2000.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J.: ObsPy: A Python Toolbox for Seismology, Seismol. Res. Lett., 81, 530–533, doi:10.1785/gssrl.81.3.530, 2010.

Czech Regional Seismic Network: website, http://www.ig.cas.cz/en/structure/observatories/czech-regional-seismological-network/okc/, 2013.

Dziewonski, A. and Anderson, D.: Preliminary reference earth model, Physics of the Earth and Planetary Interiors, 25, 297–356, doi:10.1016/0031-9201(81)90046-7, 1981.

Gibbons, S. J.: On the Identification and Documentation of Timing Errors: An Example at the KBS Station, Spitsbergen, Seismol. Res. Lett., 77, 559–571, doi:10.1785/gssrl.77.5.559, 2006.

Guralp Systems Ltd.: website, http://www.guralp.com/poles-and-zeroes-with-positive-normalization-factors/, 2013.

Soomro, R., Weidle, C., and Meier, T.: Automated fundamental mode surface wave phase velocity measurements and its application to central Europe, in preparation, 2013.