Data quality control and tools in passive seismic experiments exemplified on the Czech broadband seismic pool MOBNET in the AlpArray collaborative project

Luděk Vecsey¹, Jaroslava Plomerová¹, Petr Jedlička¹, Helena Munzarová¹, Vladislav Babuška¹, and the AlpArray working group

¹Institute of Geophysics of the Czech Academy of Sciences, 14131 Prague, Czech Republic

Abstract. This paper focuses on major issues related to the data reliability and network performance of 20 broadband (BB) stations of the Czech (CZ) MOBNET (MOBILE NETwork) seismic pool within the AlpArray seismic experiments. Currently used high-resolution seismological applications require high-quality data recorded for a sufficiently long time interval at seismological observatories and during the entire time of operation of the temporary stations. In this paper we present new hardware and software tools we have been developing during the last two decades while analysing data from several international passive experiments. The new tools help to assure the high-quality standard of broadband seismic data and eliminate potential errors before supplying data to seismological centres. Special attention is paid to crucial issues like the detection of sensor misorientation, timing problems, interchange of record components and/or their polarity reversal, sensor mass centring, or anomalous channel amplitudes due to, for example, imperfect gain. Thorough data quality control should represent an integral constituent of seismic data recording, preprocessing, and archiving, especially for data from temporary stations in passive seismic experiments. The presented tools can be applied manually or automatically on data from any seismic network.

1 Introduction

The long-term experience of some of the authors with passive-experiment data processing encourages the team to summarize the tools we have been using for testing and improving seismic data quality and which might be of interest to a broader community. The necessity of data quality control is evident nowadays and several procedures are applied automatically in data centres, e.g. MUSTANG software in IRIS, which identify data errors in the centre databases. Our endeavour is to identify data errors and correct them when possible, before supplying data to the centres. Data quality control before experimental data archiving is of great importance.

Data from passive seismic experiments of different lateral extent and a dense station distribution became a crucial source of information for the modern Earth interior researches. USArray (www.usarray.org) or IberArray (http://iberarray.ictja.csic.es; Díaz et al., 2010) represent the large-scale temporary networks, whereas for example TRANSALP (Lippitsch et al., 2003) or BOHEMA (Plomerová et al., 2007) belong to small-scale passive experiments in central Europe. Participants of the AlpArray project, the European collaborative geoscience initiative, deployed the largest network of temporary broadband (BB) stations ever per-
Figure 1. Broadband stations of the MOBNET pool involved in the AlpArray project: (a) in the network of the AlpArray-EASI complementary project (2014–2015) and (b) in the AlpArray Seismic Network (AASN). The broadband stations of the complete AASN cover the area within 250 km distance (outer white curve) from the smoothed 800 m altitude line of the Alps (inner white curve). The array consists of 623 stations including 30 ocean-bottom seismometers (OBS) in the Ligurian Sea.

Figure 2. Schematic map of major tectonic units of the Bohemian Massif and seismic stations of the MOBNET pool involved in the network of the AlpArray-EASI complementary project (blue circles) and the AlpArray Seismic Network (AASN, green circles) of the AlpArray project. CRSN – Czech Regional Seismic Network (Institute of Geophysics, 1973); WEBNET – West Bohemia Seismic Network (Institute of Geophysics, 1991).
considering anisotropy in all three types of investigations, as well as in shear-wave splitting analyses, will be applied. The depth range of scientific investigations encompasses the crust and the mantle lithosphere, down to the lithosphere–asthenosphere boundary (LAB), as well as the sub-lithospheric upper mantle.

To achieve the objectives of the project, it is necessary to apply various geological–geophysical imaging methods on data recorded by a homogeneous network of broadband seismic stations in the greater Alpine area (Fig. 1). Although the area is in some parts densely covered by permanent seismic observatories, their distribution is far from homogeneous. Therefore, the distribution of ~360 existing permanent stations has been complemented by ~260 temporary BB stations to create a relatively dense network of unprecedented large scale in Europe, with a homogeneous station spacing of about 52 km. The station spacing and station location is designed in such a way that for any site in the Alpine region there is always a station of the AlpArray Seismic Network (AASN) at a distance up to ~30 km. The temporary seismic network of such a large extent requires intensive collaboration between many institutions (currently more than 45 institutions from 17 countries), the combination of individual national–institutional seismic pools of temporary stations, and coordination of their deployment, following the high-level maintenance and experienced handling. Thanks to the large extent of the array and density of the stations, the results from seismic tomography and several other techniques applied to data collected during the unique passive experiment will shed light on the detailed 3-D architecture of the crust and upper mantle from the Earth’s surface down to ~600 km of this extremely complicated orogenic region.

The AlpArray area, set as a region delimited by a 250 km distance from the 800 m altitude isoline surrounding the Alps, covers a large portion of the Czech (CZ) part of the Bohemian Massif. Ten BB observatories of the Czech Regional Seismological Network (CRSN), one permanent BB station of the West Bohemia Seismic Network (WEBNET) along with 20 temporary BB stations from the pool of seismic stations MOBNET (MOBILE NETwork) of the Institute of Geophysics of the Czech Academy of Sciences (IG CAS), cover the area with the spacing required (Fig. 1). Apart of the AASN, which is the backbone of the AlpArray project, the MOBNET stations were involved in the AlpArray Eastern Alpine Seismic Investigation (EASI) complementary project (coded XT in the European Integrated Data Archive, EIDA; AlpArray Seismic Network, 2014). The additional 10 stations have been operational since June 2017 in another complementary project, AlpArray IVREA (coded XK in the European Integrated Data Archive). The Czech team is responsible for the deployment and maintenance of the MOBNET stations included in the Czech part of the AASN (coded Z3 in the European Integrated Data Archive system, www.fdsn.org/networks/detail/Z3_2015/), as well as for the completeness and correctness of the seismic data recorded in the AlpArray-EASI complementary experiment and transferred to the EIDA centres.

The main purpose of this paper is to describe technical parameters of the MOBNET stations, to present newly developed control units for setting sensor and data acquisition systems (DAS), and to document the significance of careful data quality control, which could help other groups in addition to those involved in the AlpArray project in preparing their seismic data for archiving. Special attention is paid to the detection of sensor misorientation, timing problems, interchange of components and/or their polarity reversal, mass centring problems, or anomalous channel amplitudes due to an imperfect gain. Maintaining the high quality of archived seismological data is crucial for the success of the AlpArray project, as well as of any passive seismic experiment.

2 Deployment of MOBNET stations within the AlpArray project

The first integration of the 20 MOBNET stations into the AlpArray project had been realized during the Eastern Alpine Seismic Investigation project, which was the first implemented AlpArray complementary experiment (Table 1; see Fig. 1). The AlpArray-EASI transect was composed of 55 broadband seismic stations. They operated from July 2014 to October 2015 and were configured in a zigzag pattern on either side of the central longitudinal line of 13.35° E, with a north–south distance between stations of 10 km. The transect spanned a 540 km long region, between the Ore Mountains at the Czech–German border in the north and the Adriatic Sea, near Trieste in the south. The distance from each station to either side of the central line was ~6 km. We followed the general recommendations of the technical strategy of the AlpArray project (www.alparray.ethz.ch), including the temperature insulation of sensors. We kept the stations within 1.5 km of the target location if topographic, field, and infrastructure conditions allowed.

The northernmost stations AAE01–AAE20 (Fig. 2) of the MOBNET pool involved in the AlpArray-EASI network were equipped mostly with the Streckeisen STS-2 seismometers, with two CMG-3T and three CMG-3ESP seismometers, and with the GAIA DAS. The stations were installed preferably in vaults of castles/chateauxs, churches, or suitable abandoned buildings. Figure 3 shows an example of a station location, seismometer installation, the quality of the site, noise level, etc. Supplement Figs. S1–S19 give the same detailed information for the remaining 19 MOBNET stations employed in the AlpArray-EASI network. Following the note of Molinari et al. (2016), we can characterize the locations as an urban free-field site and only exceptionally as a building site (Table 1). The stations ran at the autonomous regime and reported their state of health (SOH) daily via the SMS messages. Altogether, we recorded 280 GB of data
The station is located in the former stables of the Manětín chateau. The seismometer is built on a concrete floor. The GPS antenna is brought out through the window: 5 m length, S direction, and partially obscured view. Geomorphology: Rakovník uplands. Subsoil: phyllite.

Figure 3. Example of installation of one of the broadband MOBNET stations in the AlpArray-EASI complementary experiment. In addition to the photo-documentation of the installation of the AAE08 station (Manětín), we show its location (upper right), station coordinates and period of recordings (top panel), and we describe the installation of accessories (GPS antenna) and the bedrock (left panel in the middle). The central panel shows the probabilistic power spectral density of ambient noise in the vertical component during the first 5 months of the station registration.

The stations exhibit distinct seasonal variations of noise level, which results in the exceeding of the noise limit in the long-period range of the horizontal components (Fig. 5; Wolin et al., 2015).

Data from the Czech temporary stations, with the access restricted according to the AlpArray rules, are transferred to ODC (www.orfeus-eu.org/data/eida/nodes/), while data from the Czech permanent stations with open access are continuously being stored in GEOFON (http://geofon.gfz-potsdam.de/). Figure 6 shows the current status of data availability from the MOBNET stations included in the AlpArray passive field experiments. In the case of the AlpArray-EASI complementary project, which ran for 15 months in 2014–2015, we retrieved 96 % of the data at each station on average (Fig. 6a). As concerns the ongoing AlpArray project, the data completeness is 99 % for the MOBNET stations included in the AASN whose data have been downloaded by September 2017. Several gaps in data were caused by summer thunderstorms that damaged electrical supplies (Fig. 6b). Although almost all our stations operate offline, the data completeness for the MOBNET stations in the AASN is similar to that for stations in the Austrian or Swiss parts of the AASN with an
online data transmission (Fuchs et al., 2016; Molinari et al., 2016).

3 Seismometer and GAIA control and calibration devices

Our broadband temporary stations involved in the AlpArray project are equipped mostly with broadband seismometers STS-2, several Guralp CMG sensors (Table 1), and with GAIA data acquisition systems developed by the Vistec company (www.vistec.cz). The hardware of the GAIA data acquisition systems supports seismometer control, but the firmware we use does not allow it. First, our AlpArray stations operate autonomously (they are not connected to the internet) and, second, a technician servicing a station often does not need to use a computer for data collection. For these reasons and to assure a high-degree reliability of the seismometer–DAS pairs’ performance, we have developed four special control devices for seismometers of different types and one for the GAIA DAS. In general, these boxes generate pulses into the systems and compare the amplitudes of the input and output signals. The devices enable the calibration of the sensors and data acquisition systems, as well as the checking of the in situ gain of all the individual components and polarity of the recorded signal. The hardware check facilitates the identification/verification of any malfunction of the systems and enables their immediate treatment, often directly in the field. These devices have been developed in response to our experience, which was gained during the preceding experiments. We applied the devices during station installations, regular servicing, or during detection of station malfunction by a software quality check. In the future, they can be used before station deployment together with the Huddle test of the instruments. Nevertheless, some malfunctioning can occur during station operation and, therefore, regular checks during station services are recommended.

3.1 Guralp host box (CMG-3T, CMG-3ESP, and CMG-3ESPC)

The Guralp host box developed in our laboratory (Fig. 7a) becomes an integral constituent of our CMG-3T, CMG-3ESP, and CMG-3ESPC seismometers. It connects the seismometer and the GAIA DAS and it is an analogy of the standard handheld unit produced by the Guralp company, or the host box of the STS-2 seismometer. The standard Guralp host box allows fundamental handlings of the seismometer, which are
the lock–unlock function of pendulums, their centring, and a calibration with the use of an external signal. On top of the fundamental handling, the Guralp host box developed in our laboratory enables the application of the built-in source control and calibration unit (CMG-40T).

This device (Fig. 7b) enables the display of the positions of the pendulums and the calibration of the seismometer by the unit rectangular pulse signal or the Dirac delta pulse. It also has an input for an external calibrating signal of an arbitrary shape. The polarity of the calibrating signal can be changed and the signal size can be altered in two levels. There is a rotary switch between the calibration mode and the display mode of pendulum positions of the Z, NS, and EW components. A push button centres the pendulums. The Guralp control and calibration unit is plugged into the Guralp host box connector.

### 3.2 Guralp control and calibration unit (CMG-3T, CMG-3ESP, and CMG-ESPC)

This device (Fig. 7b) enables the display of the positions of the pendulums and the calibration of the seismometer by the unit rectangular pulse signal or the Dirac delta pulse. It also has an input for an external calibrating signal of an arbitrary shape. The polarity of the calibrating signal can be changed and the signal size can be altered in two levels. There is a rotary switch between the calibration mode and the display mode of pendulum positions of the Z, NS, and EW components. A push button centres the pendulums.
ponents are calibrated together, the calibration currents and of an arbitrary shape, e.g. of a sinusoidal signal. If the com-

pulse. There is also a switch for an external calibrating signal separately with the unit rectangular signal or the Dirac delta

Figure 5. Probabilistic power spectral density (PPSD) on the HHE horizontal component of the AAE01 station of the AlpArray-EASI experiment. Branching of the PPSD reflects both diurnal and seasonal variations of noise, and this is clearly visible in both the high-frequency and in the low-frequency bands, respectively. Noise level during the summer period is low and far below the required noise level (20 dB below the New High Noise Model, NHNM, upper grey curve; Peterson, 1993) even for periods larger than 10 s, whereas during winter time noise exceeds the limit of 10 dB below NHNM for periods larger than 30 s.

...for the pendulum position check of the seismometers with electronic control, but in such cases, the centring unit does not enable the correction of the pendulum position.

3.4 STS-2 control and calibration unit

The STS-2 control and calibration unit (Fig. 7d) has been developed for centring pendulums and for calibrating the Streckeisen STS-2 seismometers. The device is connected to the host box provided by the seismometer producer. The host box forms the integral part of the system, through which the STS-2 seismometers are controlled and powered. The host box has two connectors. The first one is used to connect the digitizer; the second one (marked as “monitor”) serves to remotely control and monitor the seismometer via the STS-2 control and calibration unit. This unit displays the positions of the pendulums for the U, V, and W components, or it can be switched to show offsets of the standard Z, NS, and EW components of the output signal. The unit is equipped with a button to automatically centre the pendulum position (auto-zero push button) connected in parallel to a similar button in the host box. The 120 s/1 s switch of the control and calibration unit changes modes between the broadband and short-period regimes.

Each of the U, V, and W components can be calibrated separately with the unit rectangular signal or the Dirac delta pulse. There is also a switch for an external calibrating signal of an arbitrary shape, e.g. of a sinusoidal signal. If the components are calibrated together, the calibration currents and their polarities are chosen so that the output signals (components Z, NS, and EW) have the same amplitudes and polarities. This procedure guarantees the correct functioning of the seismometer.

3.5 GAIA gain and calibration unit

The GAIA gain and calibration unit (Fig. 7e) checks and calibrates inputs into the GAIA DAS, but it can be used for the calibration of any type of digitizer as well (Kinemetrics, Nanometrics, Ref Tek, Guralp, etc.) after being equipped with the corresponding connector reductions. The unit enables the calibration of analogue inputs to check the correct order of the channels, to determine intensity of cross talks between the channels, and to measure channel amplification and sensitivity (a voltage corresponding to the LSB – least significant bit). The number of channels undergoing calibration and channel polarity can be changed. The calibration is done by a defined voltage jump. For the calibration of the analog inputs, we can use a differential or a single-ended mode. In the differential mode, the voltage is connected between inputs marked as +IN and −IN. In the single-ended mode, the calibrating voltage is connected to ground (GND) and +IN or −IN. The built-in generator of sawtooth-shape calibrating voltage serves for an assessment of the linearity of the analog-to-digital conversion.

4 Data quality control and assurance

The high level of data quality has to be stable during a long time interval for seismological observatories and for the entire time of operation of the temporary stations within the passive experiments. Data quality control represents the necessary steps to achieve and maintain the high quality level of recorded data. We differentiate between (1) in situ controls with technical equipment, applied during station installation and servicing, and (2) subsequent software controls, applied to downloaded data.

4.1 Seismic noise

The measure of seismic ambient noise level is nowadays a standard procedure when searching and selecting sites that are suitable for station installation. Therefore, we measured noise at each site before a station installation for a short time. Once a station is installed, the noise level has to be frequently checked to monitor potential changes in conditions of the recordings or to detect technical problems at the station. According to the AlpArray working group requirements, the average noise level should be 20 dB lower than the New High Noise Model (NHNM; Peterson, 1993) on all components within the 1–10 Hz frequency range. In the long-period range (30–200 s), the same noise level is required only for the vertical component. Because ambient noise is usually higher in the horizontal components, the average noise level is recom-

Figure 6. Data completeness of the MOBNET stations in the AlpArray projects: (a) in the AlpArray-EASI complementary field measurements and (b) in the ongoing AlpArray experiment.

Figure 7. Control and calibration units developed for the broadband seismometers and the GAIA DAS to guarantee the high quality of recorded data.

Figure 8 shows the PPSD medians for all MOBNET stations deployed in the AlpArray-EASI and AASN networks. While the noise level for periods below 1 s fulfils the noise requirements for most of the stations and for the three components, noise in the horizontal components for periods longer than 10 s is often higher, especially in winter, but still acceptable for temporary deployments. One has to bear in mind that a compromise between optimal site conditions and the required array geometry has to be accomplished. Thus, at the short-period range, we have to accept higher noise level at some sites, where human activity is higher (e.g. AAE03 located in an administrative building in a village). Microseisms dominate a period interval of 1–10 s in central Europe and also increase in winter. The broadband seisome-
Figure 8. Medians of the probabilistic power spectral density (PPSD) of seismic noise at the MOBNET stations involved (a) in the AlpArray-EASI complementary experiment and (b) in the AlpArray Seismic Network (AASN). The grey thick lines mark the New Low and New High Noise models (NLNM and NHNM; Peterson, 1993). The thick red line (AlpArray high noise limit, AAHNL) marks the upper limit of the recommended noise level below the NHNM.

...ters are sensitive to several external effects, especially in the range of longer periods. The most significant of these are the temperature changes, either diurnal or seasonal, and pressure fluctuations. An enhanced insulation of seismometers might decrease the effects, particularly on the horizontal components. Therefore, seismometer insulation plays an important role in ensuring the high-quality data. The majority of our stations are installed in vaults with only small temperature variations, which could directly influence the seismometer pendulums. On the other hand, there are also indirect effects of temperature, particularly an inclination of bedrocks or buildings. Most of our stations are equipped with...
Figure 9. The waveform similarity method showing the interchange of the EW and Z components (a) at the AAE05 station. Daily amplitude means (b) of all three components at the AAE05 station during June–July 2015 indicate the interchange of all three components. Vertical dashed lines show dates of station servicing.

the STS2 seismometers with three pendulums in 120° orientation. The 120° orientation means that there are no vertical and two horizontal orientations of pendulums. These are reconstructed from the original 120° pendulums. There is no reason why two reconstructed horizontal components should be more affected by direct temperature changes than the Z component reconstructed from movements of the three pendulums. Only two of our stations, A090A and A078A, are located outside a building, where temperature changes can be significant. These stations exhibit the highest noise at the long-period range along with the A082A station located above loess around the Elbe river. However, we also observe a similarly high long-period noise at station A081A, which is in a vault and well insulated. The lowest noise at longer periods is observed at station A076A, where the seismometer is in a 3.5 m deep shaft dug in a rock below the building and thus experiences minimal temperature variations or other jamming. On the other hand, the seismometer of station A084A is also located in almost ideal conditions inside a castle, in a vault space near a rock outcrop, but in spite of that the station exhibits relatively high noise in the long-period range. Noise is generally higher in the EW components than in the NS components in our region (Fig. 8). The sources of these effects are complex and not all of them are clear. Dominant western winds could probably increase the noise registered in the EW components. Although we are not able to determine all sources of long-period noise, we can exploit the difference in noise levels in the horizontal and vertical components as one of the tools to decipher the potential interchange of the components, as we describe below.

4.2 Sensor orientation

The exact orientation of seismometers in the geographic coordinate system is one of the most important tasks during station installations. Misoriented sensors negatively affect the results of the procedures based on modern three-component seismological observations and can lead to false interpretations (Ekström and Busby, 2008; Vecsey et al., 2014; Wang et al., 2016). The determination of the northward direction has been routinely performed for years with the use of standard compass, with the best accuracy being ±5° in the case of no magnetic disturbances in the nearby surroundings. However, such accuracy is no longer sufficient. The top-level current practice is the orientation of seismometers with the use of the high-precision optical gyrocompass measurements during a station installation and the repetition of the measure-
The Rayleigh-wave polarization-angle method showing the differences between the Rayleigh-wave polarizations and theoretical back azimuths (BAZ) depending on the BAZ of earthquakes; (a) linear dependence of the difference identifies the polarity reversal of the NS component. The polarization of waves arriving from the east or west is not affected by the reversed NS component. Station PA10 of the PASSEQ experiment 2006–2008 (Wilde-Piórko et al., 2006; Vecsey et al., 2014) is used as an example; (b) the constant difference of $\sim 180^\circ$ corresponds to the polarity reversal of both horizontal components at the A077A station of AASN.

Figure 10. The Rayleigh-wave polarization-angle method showing the differences between the Rayleigh-wave polarizations and theoretical back azimuths (BAZ) depending on the BAZ of earthquakes; (a) linear dependence of the difference identifies the polarity reversal of the NS component. The polarization of waves arriving from the east or west is not affected by the reversed NS component. Station PA10 of the PASSEQ experiment 2006–2008 (Wilde-Piórko et al., 2006; Vecsey et al., 2014) is used as an example; (b) the constant difference of $\sim 180^\circ$ corresponds to the polarity reversal of both horizontal components at the A077A station of AASN.

To determine the correct sensor orientations, one can use the Rayleigh-wave polarization-angle method (e.g. Stachnik et al., 2012), in which differences between the Rayleigh-wave polarizations and their theoretical back azimuths are plotted depending on event origin times. Of course, this method cannot be as precise as measurements with the optical gyrocompass. Rueda and Mezcua (2015) found only $1–5^\circ$ differences between the northern estimate by the Rayleigh-wave polarization method and the gyrocompass measurements of the north for long-term data series at observatories. In the case of shorter time intervals the accuracy of the Rayleigh-wave polarization method is low (can exceed $10^\circ$) and thus only large sensor misorientation can be detected. For the determination of the exact moment of the change in sensor orientation, the Rayleigh-wave polarization-angle method can be combined with plots of daily amplitude means. After determining a day when the sensor happens to be accidentally misoriented, one has to search for sudden amplitude changes in the data, which cannot be connected with the seismic signal.

When installing our stations for the AlpArray-EASI transect, we oriented the seismometers carefully, but only with the use of a standard compass. Later we checked the orientation of all sensors with an optical gyrocompass. We have found deviations larger than $5^\circ$ from the true north at 9 of the 20 stations (Table 2, first measurements) and extremely large deviations in orientation at two of them (AAE13 $N = 282^\circ$ and AAE04 $N = 341^\circ$). The first deviation of $-78^\circ$ in the north determination can be attributed to an error which occurs when the STS2 sensor-orienting rod is installed in the north-facing direction instead of the east-facing direction. The reason for the wrong seismometer orientation in the AAE04 station is unknown. The seismometers at two other stations, AAE13 and AAE18, significantly changed their orientation during the experiment by 8 and $7^\circ$ (the fifth column in Table 2), respectively. We have used the Rayleigh-wave polarization-angle method for a rough estimate of a moment when the orientation of the sensors has been changed and of daily means and signal plots for setting the exact time of the sensor reorientations.

The sensors in all our stations involved in the currently running AASN network (A071-A090) have been installed with the use of our gyrocompass and their orientation is regularly checked. During about a 1-year period of the array operation, we recorded three unwanted changes in sensor orientation due to human intervention. In addition to the necessary sensor reorientation on the spot, previous inaccuracies in sensor orientations have been corrected in the metadata. In the case where the deviation in the seismometer orientation is larger than $5^\circ$, the horizontal N and E components are renamed to components 2 and 3, respectively, according to the SEED reference manual (FDSN, 2012, http://www.fdsn.org/media/_s/publications/SEEDManual_V2.4.pdf).

4.3 Timing issues

Correct timing is crucial for studies based on exact arrival times of seismic waves. Incorrect time decreases the accuracy of picking arrival times of individual phases and causes a false phase identification or a complete loss of data. Here we address three important timing problems: the leap sec-
Table 2. Sensor orientation during the AlpArray-EASI complementary experiment measured by gyrocompass. The largest differences are in bold.

<table>
<thead>
<tr>
<th>EASI stations (CZ)</th>
<th>First measurement (degree)</th>
<th>Second measurement (degree)</th>
<th>Third measurement (degree)</th>
<th>Difference (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAE01</td>
<td>359.9</td>
<td>359.9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>AAE02</td>
<td><strong>7.4</strong></td>
<td>0.4</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>AAE03</td>
<td>4.0</td>
<td>359.8</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>AAE04</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>AAE05</td>
<td>357.1</td>
<td>359.3</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>AAE06</td>
<td>3.2</td>
<td>0.9</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>AAE07</td>
<td><strong>355.6</strong></td>
<td>0.4</td>
<td>2.6</td>
<td>2.2</td>
</tr>
<tr>
<td>AAE08</td>
<td>358.0</td>
<td>0.6</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>AAE09</td>
<td>2.1</td>
<td>0.3</td>
<td>359.6</td>
<td>–0.7</td>
</tr>
<tr>
<td>AAE10</td>
<td><strong>8.7</strong></td>
<td>0.4</td>
<td>3.1</td>
<td>2.7</td>
</tr>
<tr>
<td>AAE11</td>
<td>5.2</td>
<td>0.7</td>
<td>359.5</td>
<td>–1.2</td>
</tr>
<tr>
<td>AAE12</td>
<td>2.9</td>
<td>0.7</td>
<td>359.5</td>
<td>–1.2</td>
</tr>
<tr>
<td>AAE13</td>
<td><strong>282.0</strong></td>
<td>0.7</td>
<td>352.3</td>
<td>–8.4</td>
</tr>
<tr>
<td>AAE14</td>
<td>2.3</td>
<td>359.5</td>
<td>357.2</td>
<td>–2.3</td>
</tr>
<tr>
<td>AAE15</td>
<td>3.2</td>
<td>359.9</td>
<td>359.4</td>
<td>–0.5</td>
</tr>
<tr>
<td>AAE16</td>
<td>2.2</td>
<td>359.4</td>
<td>0.5</td>
<td>1.1</td>
</tr>
<tr>
<td>AAE17</td>
<td><strong>6.2</strong></td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>AAE18</td>
<td><strong>7.2</strong></td>
<td>0.4</td>
<td>6.9</td>
<td><strong>6.5</strong></td>
</tr>
<tr>
<td>AAE19</td>
<td><strong>6.0</strong></td>
<td>359.8</td>
<td>359.6</td>
<td>–0.2</td>
</tr>
<tr>
<td>AAE20</td>
<td>3.5</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

First measurements: checking within towards north orientation of seismometers determined during station installation with the use of a standard compass. Second measurements: towards the north reorientation of sensors with the gyrocompass. Third measurements: check of sensor orientation at the end of registration. Difference: difference between the third and second measurements.

Second recorded with a delay, switch between the UTC and GPS times, and malfunction of an oscillator tuning the station time or loss of time synchronization for a time period.

The first item – the leap second – is introduced into the Coordinated Universal Time (UTC) usually once or twice per year in order to keep the UTC time close to the mean solar time. The leap second is usually applied at midnight while clocks in data acquisition systems are being synchronized later, for example, with a 30–90 min delay. Moreover, the leap-second correction is applied at individual stations differently, because the times of their synchronizations differ. It is thus necessary to apply the leap second exactly at midnight (00:00) for all temporary stations before data archiving. Surprisingly, we have found a case when even a permanent observatory, out of Bohemian Massif, kept the uncorrected time for about 1 month.

The second item – the switch between the UTC and GPS times – can arise due to the wrong synchronization of the inner time (UTC) of a station and the GPS time. This can happen when the coordinated universal time in the “almanac”, transmitted by satellites, disappears from the memory of a station due to a number of reasons (e.g. low voltage of inner battery, incorrect satellite signal recorded). Existing time gaps and overlaps in miniSEED data can be calculated from the time of the first sample, number of samples, and sampling rate in each miniSEED block. Then the appropriate time shift is applied in miniSEED data for the identified time interval. Currently, the UTC and GPS times differ by 18 s. Such time shift can last for several hours or a full day and thus needs to be corrected.

Timing errors of 1 s or smaller are not clearly evident during routine seismological analyses but can be revealed from station “log” files, if provided by the registration system. Small time shifts can occur as a result of improper time synchronization due to the loss of the GPS signal or due to the failure of the oscillator tuning the station time. This third item is a more complicated issue and it allows only an approximate time reconstruction. A failure of the oscillator tuning can cause a jump or a linear increase in timing errors in data. However, such difficulties should occur only exceptionally. If they happen and we are able to identify such problems and reconstruct the real timing, it is necessary to correct times directly in the miniSEED data, which is a more complex task than applying corrections in the metadata. The same concerns a reconstruction of the correct time after a loss of the time synchronization. When checking our data, we have found an oscillator failure at station A087A, which resulted in a final time error of 0.18 s during 8 days in October 2015.

Keeping exact time in seismic data is a delicate issue. However, severe errors due to asynchronous application of the leap seconds or due to switches between the UTC and GPS times can be identified and corrected automatically in
Figure 11. Seismometer gain check by the ambient noise gain method. Gain imperfection can be identified from monthly averages of normalized power spectral ratios. Both panels show that either gain of the EW component or gains of both the NS and Z components are not correct. To decide which amplitudes are correct and which are not, one has to check the gain of each component of the sensor–DAS pair with the calibration units, as described in Sect. 3.

any data set, including the entire AlpArray data set. Small time changes must be solved individually.

4.4 Interchange of components and polarity reversal

Sometimes results from different studies dealing with waveforms raise a suspicion that the components of seismograms are interchanged and/or the polarities reversed. Although the case is rare, we found it several times in different data sets, including data from permanent observatories. Surprisingly, the component interchange can occur during station operation, e.g. twice in the AlpArray stations until now. The simplest way to verify the correct identification of the three components is a comparison of waveforms for a selected strong teleseismic event recorded on several nearby stations, which we call the waveform similarity method (Fig. 9a). Several other methods can be used as well, e.g. a visualization of daily means of signal amplitudes, sometimes called offsets (Fig. 9b), or a comparison of noise levels in the vertical and horizontal components in PPSD. In the case of correct component identification, the noise level in the vertical component should be lower than that in the horizontal components. Correction of interchanged components can be done either in the metadata or preferably directly in the miniSEED data.

Reversed polarity of components, arising from different technical reasons, is not as rare as one would expect. We identified polarity reversals using the manual waveform similarity method for nearby stations. We can also use a single-station method that is based on a semi-automatic search of Rayleigh-wave polarization (the Rayleigh-wave polarization-angle method) originally developed for verification of sensor orientation. Differences between the Rayleigh-wave polarization and the theoretical back azimuths are plotted against the theoretical back azimuths (Fig. 10). If only one horizontal component is reversed, the differences change linearly between $-180$ and $+180^\circ$. The zero difference re-
Figure 12. Error in sensor mass centring. Flat curves in the probabilistic power spectral density (PPSD; a) reflect an amplitude saturation due to insufficient mass centring and identify the existence of a problem. The time range of the failure of the automatic mass recentring is determined by changes in the daily means of the amplitudes (b), their standard deviations (c), and their absolute values of daily amplitude extremes (d).

reflects the fact that the reversed component does not play any role in the component summation and identifies the components with the correct polarity (see Fig. 10a; the EW component is the correct one). In the case of the reversed polarity on the Z component, or if both horizontal components are reversed, the differences between the Rayleigh-wave polarizations and the theoretical back azimuths attain values around 180° for all back azimuths (Fig. 10b). Moreover, we have also identified an interchange of both horizontal components in combination with their polarity reversals. This complicated case can be solved by the combination of both methods mentioned above and by a careful analysis of the results. Similarly to the component interchange, the component reversal can be corrected either in the metadata or preferably in the miniSEED data.

4.5 Gain imperfection

Anomalous signal amplitudes due to imperfectly set gains on one or more components are not very frequent in comparison with the sensor misorientations, but the danger of imperfect gains is similarly large for data analysis procedures. We can recognize anomalously large or low recorded amplitudes in two ways: first, by means of technical devices, such as control and calibration units (see Sect. 3), and, second, by means of software methods applied on recorded seismic signals.

One possible software inspection of the amplitude size can be based on ambient noise, which is evaluated regardless. Moreover, ambient noise is the only continuous signal in seismic data. We have implemented a new ambient noise gain method which compares ratios of normalized power spectra between the three components in a range of 4–8 s. In this range, the secondary microseisms are substantially larger than noise from local sources. The directionality of the microseisms due to different sources is eliminated by normalizing the spectrum of each trace via an average spectrum calculated over the traces of surrounding stations. The spectra are calculated within different time intervals, e.g. weeks, months, or a complete time range. The resulting ratios of the spectra provide a running record of individual channel sensitivity and allow us to follow potential changes in the amplitudes in a course of time. In combination with sporadic in situ gain controls by the Gain and calibration box (Sect. 3.5), we have reliable control of the potential anomalous size of recorded amplitudes and thus we can determine when a detected change in the gain occurred. We estimate the precision of the gain determination by the ambient noise gain method at 1–2 dB depending on the length of the time period analysed. Tiny variations of the curves in Fig. 11 are within this limit, but the differences between the curves are stable.

We document a successful use of hardware and software methods on data from the two seismic experiments. During the data processing, we have found that the power spectra of the EW components at stations AAE14 (AlpArray EASI) and A087A (AASN) are lower by approximately 11 dB (Fig. 11a). The NS/Z component ratio is close to zero, while the EW/Z and EW/NS ratios, where the EW component is involved, are 10 dB lower. Station documentations identi-
fied that stations AAE14 and A087A were equipped with an identical sensor and data acquisition system. Therefore, afterwards we tested the gain of each component of the sensor–DAS pair with the calibration boxes as described in Sect. 3. The test confirmed the amplitudes recorded in the EW component were 3.6 times smaller ($20 \times \log_{10}3.6 = 11 \text{ dB}$) than they should be. The technical error in the acquisition system was identified and repaired. If such error is identified by an in situ measurement, then it can be immediately eliminated (DAS can be repaired or changed, as it was possible in the case of running station A087A). Metadata of A087A for a previous period, as well as the metadata of the AAE14 station active in the finished AlpArray-EASI measurements, were corrected subsequently. In another case we have found that either the amplitudes on EW components are about 2 times larger or the gains of the NS and Z components are lower by $\sim 6 \text{ dB}$ at stations AAE15 (AlpArray EASI) and A088A (AASN; Fig. 11b). The results of the normalized PPSD ratios are only relative ones. The absolute value – the half-size gain compared with the declared one – was identified by an in situ measurement with the use of the STS-2 control and calibration unit (see Sect. 3.4). The source of the low gain was localized in a defect cable of the seismometer. The double-checked gain levels of each component (by the hardware units and by the software calculating the normalized PPSD ratios) enabled us to reliably correct the gains in the station metadata files and thus to correct anomalous amplitudes.

### 4.6 Drift of sensor mass position

One of artefacts seen in the PPSD reflects a failure of the automatic mass recentring of the sensor (McNamara and Bu- land, 2004). If a seismometer is not able to correct a drift of the mass position itself, the amplitudes of seismic signals become saturated. The signal corresponding to such a time period has a characteristic “flat” spectrum shape (Fig. 12a). The flat course in an interval of $\sim 0.3–50 \text{ s}$ differs clearly from the shape of the noise distribution modulated by secondary microseisms. The large undesirable drift of the mass position from its central position limits the dynamic range of the sensor and, therefore, needs to be identified as soon as possible. Running information about a sensor mass “drift” comes from the MAX/MIN amplitude extremes reported by GAIA DAS in daily SMS reports (see also Fig. 13). In addition to this metric check, daily means of recorded amplitudes (Fig. 12b) serve as an independent fast and easy tool for ex-post identification of the mass centring problem. Better assessment of the state of health of each station can be achieved if we complement the daily amplitude means with their standard deviations and absolute values of daily amplitude extremes (maxima or minima) (Fig. 12c).

To summarize the application of different methods of seismometer–GAIA DAS pair operation and recorded data quality checking, either by software or hardware tools presented above, we plot an optimal workflow in Fig. 13. The scheme comes from our experience with data from several previous passive experiments. Some methods give indications about an error, which requires further verification. Some of the methods are repeated in time in attempts to detect changes which can occur during station operation and thus could not be revealed by the Huddle pre-installation test.
5 Conclusions

We have developed both the hardware and software tools to contribute reliable high-quality waveform data to passive seismic experiments. At present, 20 broadband stations of the Czech MOBNET pool of temporary stations are incorporated in the AlpArray Seismic Network. The stations were also deployed in the previous AlpArray-EASI complementary experiment. To assure a high-degree of reliability of the STS-2/CMG-seismometer–DAS pairs’ performance, we have developed four special control devices for seismometers of different types and one for the GAIA DAS. The devices calibrate both the sensors and data acquisition systems in situ and allow us to check the gain and the polarity of all three components. We emphasise the importance of precise sensor orientation by using a gyrocompass both during station installations and during its regular checks during the field measurements. The information extracted from probabilistic power spectral density, spectra ratios, and averages of daily amplitudes and other parameters, followed by the designed procedures in routine data processing, allow us to identify several other problems, e.g. imperfectly set gains, interchange of components and polarity reversals, insufficient sensor mass centring, and, last but not the least, time issues. The hardware control in situ and the ex-post software data checking represent the double check of data quality. The former removes problems immediately in field, and the latter allows restoring data back in time, until the moment when a problem occurred. The fully automated software methods could be used for the entire AlpArray data set. We believe that the newly developed control and calibration units for setting sensor–DAS systems and the documentation of the significance of careful data quality control with the use of the software tools could be helpful for other groups participating in collaborative passive seismic experiments.

Data availability. Data from the MOBNET pool as a part of the AlpArray project is stored in EIDA (www.orfeus-eu.org/eida/), currently with restricted access (http://www.alparray.ethz.ch/research/complementary-experiments/easi/data-access-citation/ and http://www.alparray.ethz.ch/seismic_network/backbone/data-access/).

The Supplement related to this article is available online at https://doi.org/10.5194/gi-6-505-2017-supplement.

Team list. The complete member list of the AlpArray working group can be found at http://www.alparray.ethz.ch.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Cooperation with participants of the AlpArray projects is greatly appreciated as well as the suggestions and recommendations of both anonymous reviewers, which substantially improved the manuscript. Research of the Czech team was supported by grant no. M100121201 of the Czech Academy of Sciences and by the project CzechGeo/EPOS-Sci (CZ.02.1.01/0.0/0.0/16_013/0001800, OP RDE) financed from the Operational Programme Research, Development and Education. Data acquisition from permanent observatories and the development of the devices was supported by the large research infrastructure project CzechGeo/EPOS, grants nos. LM2010008 and LM2015079. The development of the software for data quality check was supported by grant no. COST LD15029 of the Ministry of Education, Youth and Sports. Several figures and calculations have been prepared with the use of the Generic Mapping Tools (Wessel and Smith, 1998) and ObsPy (Krischer et al., 2015).

Edited by: Luis Vazquez
Reviewed by: three anonymous referees

References


Institute of Geophysics: Academy of Sciences of the Czech Republic: Czech Regional Seismic Network, International Feder-
L. Vecsey et al.: Data quality control and tools in passive seismic experiments


