Moment Tensor Inversion Based on the Principal Component Analysis of Waveforms: Method and Application to Microearthquakes in West Bohemia, Czech Republic

by Václav Vavryčuk, Petra Adamová, Jana Dubravová, and Hana Jakoubková

ABSTRACT

We develop and test a new hybrid approach of the amplitude and waveform moment tensor inversions, which utilizes the principal component analysis of seismograms. The proposed inversion is less sensitive to noise in data, being thus more accurate and more robust than the amplitude inversion. It also suppresses other unmodeled phenomena, like a directivity of the source, errors caused by local site effects at individual stations, and by time shifts in arrivals of observed and synthetic signals due to an inaccurate velocity model. This inversion is computationally less demanding than the full waveform inversion and thus applicable to large sets of earthquakes. The approach is numerically tested on synthetic data with various levels of noise. The applicability of the inversion is demonstrated on inverting more than 800 microearthquakes that occurred during the 2014 activity in West Bohemia, Czech Republic. The analysis revealed several distinct clusters of moment tensors. Focal mechanisms corresponding to moment tensors of three clusters are left-lateral strike slips associated with the most active fault in the focal zone. Another cluster is characterized by right-lateral strike slips associated with the fault conjugate to the main fault. Finally, we identified a cluster with pure reverse focal mechanisms that are anomalous and not expected to occur in the region. These mechanisms were not detected in previous seismic activity, and they have an unfavorable orientation with respect to regional tectonic stress. This might indicate a presence of local stress heterogeneities caused, for example, by an interaction of faults or fault segments in the focal zone.

INTRODUCTION

The seismic moment tensor (MT) describes equivalent body forces acting at a seismic point source. It comprises not only a double-couple (DC) component representing shear faulting on a planar fault in isotropic media but also non-double-couple (non-DC) components produced, for example, by shear faulting on nonplanar faults, by cavity collapses in mines, by tensile faulting induced by fluid injection in geothermal or volcanic areas, or by seismic anisotropy in the focal zone (Julian et al., 1998; Miller et al., 1998; Nakamichi et al., 2003; Templeton and Dreger, 2006; Ford et al., 2008; Minson and Dreger, 2008; Šílený and Milev, 2008; Vavryčuk et al., 2008; Guilhem et al., 2014). Hence, MTs carry information about the orientation of activated fractures, fracturing mode, and elastic properties of the material in the focal zone (Vavryčuk, 2006; Vavryčuk and Hrubcová, 2017).

The MT inversion is a delicate procedure that requires an accurate location of the source, a detailed velocity model, good azimuthal coverage of stations on the focal sphere equipped with sensors of rather broad and accurately defined frequency response, and low noise in data (Šílený, 2009; Ford et al., 2010; Stierle, Bohnhoff, et al., 2014; Stierle, Vavryčuk, et al., 2014). We can invert amplitudes of individual seismic phases (Vavryčuk et al., 2008; Fojtíková et al., 2010; Kwiatek et al., 2016), amplitude ratios (Miller et al., 1998; Hardebeck and Shearer, 2003; Jechumtálová and Šílený, 2005), or full waveforms (Dziewonski et al., 1981; Kikuchi and Kanamori, 1991; Šílený et al., 1992). The full waveforms are often inverted, assuming a point-source approximation and a time-independent focal mechanism. The inversion is performed in the time domain (Dreger and Woods, 2002; Sokos and Zahradník, 2008; Zahradník et al., 2008; Adamová et al., 2009), in the frequency domain using amplitude spectra (Cesca et al., 2006) or complex spectra (Cesca and Dahm, 2008), or using a combined time-frequency approach (Vavryčuk and Kühn, 2012). Each inversion method is suitable for a different range of epicentral distances and frequencies of analyzed waves. The waveform inversion is preferable when applied to broadband data of moderate or large earthquakes recorded at regional or global seismic networks. By contrast, the amplitude inversion is advantageous when applied to short-period data of small earthquakes or microearthquakes recorded at local networks. To extend the applicability of the MT inversion or to get more robust results, some hybrid approaches propose various combinations of joint...
inversion of polarities, amplitudes, waveforms, or amplitude ratios (Li et al., 2011a,b; Fojtíková and Zahradník, 2014; Vavryčuk and Kim, 2014).

The analysis of MTs of microearthquakes is challenging for several reasons. First, the waveforms are complex and of high frequency, reflecting small-scale inhomogeneities in the Earth’s crust. Second, the waveforms are often noisy because of low magnitude of events. Finally, the microseismic data comprise usually a large number of earthquakes that cannot be processed manually and need an automatic or semiautomatic processing. All these facts restrict the applicability of the waveform inversion and favor a much simpler amplitude inversion. The waveform inversion is rather laborious and time-consuming, and needs a detailed velocity model for computing accurate Green’s functions. In comparison, the amplitude inversion is simpler and less computationally demanding than the waveform inversion. However, it is usually less accurate; it needs observations at more stations, and it is more sensitive to noise in data than the waveform inversion.

In this article, we describe a new inversion technique that improves the performance of the amplitude MT inversion by incorporating some aspects of the waveform inversion. The idea is based on computing the source time function and effective amplitudes of direct waves needed in the inversion by applying the principal component analysis (PCA) to seismograms. The proposed hybrid approach of the amplitude and waveform inversions is less sensitive to noise, being thus more accurate and more robust than the amplitude inversion. Still, the inversion is simple and computationally less demanding than the waveform inversion and thus applicable to large sets of earthquakes. The approach is numerically tested on synthetic data with various levels of noise. The applicability of the inversion is demonstrated by inverting more than 800 microearthquakes that occurred during the 2014 activity in West Bohemia, Czech Republic.

**METHOD**

**Principal Component Analysis**

The PCA is a statistical method that uses an orthogonal decomposition to transform possibly correlated observations into a set of linearly uncorrelated variables called the principal components. The principal components are ordered in such a way that the first component has the highest possible variance retaining as much as possible of the variation present in the data. Each component has the highest variance under the constraint of its orthogonality to the preceding components. The resulting vectors form an uncorrelated orthogonal basis (Jackson, 1991; Jolliffe, 2002). The goal of the PCA is to condense the maximum amount of variance present in the data into the fewest number of principal components. The PCA is commonly used in natural as well as social sciences, typically, in problems related to analysis of large data sets and to reducing their multidimensionality. The PCA is widely used in geophysics and seismology (e.g., Hagen, 1982; Dong et al., 2006; Ocana et al., 2008; Kositsky and Avouac, 2010).

Earthquake-related applications of the PCA include the analysis of ionospheric and geomagnetic anomalies associated with earthquakes (Hattori et al., 2006; Lin, 2012a,b), spatiotemporal analysis of crustal deformations produced by large earthquakes (Kositsky and Avouac, 2010; Gualandi et al., 2014); and signal detection, filtering, and polarization analysis of seismic data (Bataille and Chiu, 1991; Wagner and Owens, 1996; Muti and Bourennane, 2007). In MT inversions, the PCA was first applied by Vasco (1989, 1990) for finding the common source time function from six MT rate functions. This approach appeared useful and effective, being widely applied by other researchers (Cesca and Dahm, 2008; Davi et al., 2010; Eyre et al., 2013; Zecevic et al., 2013).

**PCA Moment Tensor Inversion**

In the approach of Vasco (1989, 1990), the waveforms are first inverted using the generalized linear inversion for six MT rate functions. If the point-source approximation with a time-independent focal mechanism is assumed, the MT rate functions should be identical functions of time, except for scaling. However, the MT rate functions are frequently different because of noise in data, inaccurate Green’s functions, and/or numerical errors. For this reason, the PCA is applied to extract the common source time function from the six MT rate functions. Alternatively, the source time function can be approximated by averaging the MT rate functions (Ruff and Tichelaar, 1990) or by minimization techniques (Šílený, 1998; Wéber, 2009).

The presented approach applies the PCA to inverting MTs in a different way. If we assume a point-source approximation and invert for MTs using direct waves, the waveforms recorded at stations should be identical functions of time. The exceptions are waves distorted by overcritical reflections that might introduce phase shifts. If we exclude stations with such anomalous records, we can apply the PCA to waveforms recorded at stations for finding the common wavelet radiated by the source. Once we find the source time function, we can invert amplitudes represented by the calculated PCA coefficients at each station for a time-independent MT. Such an approach is advantageous for several reasons: (1) we eliminate or suppress all station-dependent effects in input data like locally generated noise, site effects, or source directivity; (2) the inversion is computationally more efficient because the waveform inversion is substituted by the amplitude inversion; and (3) the inversion is more stable and robust because the sensitivity of the inversion to noise and other unmodeled effects is minimized.

**Mathematical Description**

Assuming $X$ to be the data matrix composed of seismic waves recorded at the network of stations (i.e., each column of matrix $X$ contains a single station component) and preprocessed to have a zero mean, we compute the eigenvalues and eigenvectors of the covariance matrix $XX^T$. Because the covariance matrix is symmetric, it can be diagonalized in the following way (Jolliffe, 2002):

$$XX^T = WDW^T$$  \hfill (1)
in which $W$ is the matrix of orthonormal eigenvectors corresponding to eigenvalues ranked in a descending order which form diagonal matrix $D$. The eigenvectors are the sought-after principal components. The PCA is closely related to the singular value decomposition, because the square roots of the eigenvalues of $XX^T$ are the singular values of $X$. Subsequently, the PCA decomposition of $X$ reads

$$X_{\text{PCA}} = XW$$

and

$$X = X_{\text{PCA}}W^T.$$  \hspace{1cm} (3)

Hence, data matrix $X$ can be reconstructed by multiplying the matrix $W^T$ of the principal components by matrix $X_{\text{PCA}}$ of the corresponding PCA coefficients.

If matrix $X$ is formed by a set of similar traces, the first principal component is dominant and corresponds to the common wavelet. Therefore, the coefficients of the PCA decomposition of the individual seismic traces, which are multiplied by the first principal component in equation (3), are the sought-after effective wave amplitudes (the PCA amplitudes). In other words, the PCA amplitudes are the multiplication factors of the common wavelet that reproduce most accurately the individual traces.

Having computed the PCA amplitudes, we proceed similarly as in the standard MT inversion of amplitudes directly measured from waveforms. The inversion solves a set of linear equations expressed in the matrix form as follows:

$$Gm = u,$$  \hspace{1cm} (4)

in which $G$ is the $K \times 6$ matrix of the spatial derivatives of the Green’s function

$$G = \begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} & G_{15} & G_{16} \\ G_{21} & G_{22} & G_{23} & G_{24} & G_{25} & G_{26} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{K1} & G_{K2} & G_{K3} & G_{K4} & G_{K5} & G_{K6} \end{bmatrix},$$  \hspace{1cm} (5)

$0 0.2 0.4 0.6 0.8 1$

Time [s]

(a) 20

(b) 20

Figure 2. Station configurations used in the numerical tests. (a) 8 stations and (b) 20 stations. (Left) Map view and (right) focal spheres. Blue triangles mark positions of the stations. The red cross in the left plots shows the epicenter. The red cross and open circle in the right plots mark the T and P axes of the shear focal mechanism used in the tests. The strike, dip, and rake angles of the focal mechanism are 170°, 70°, and −45°, respectively.

Figure 3. The (a) displacement and (b) velocity records of the P wavelet used in the numerical tests.
\( \mathbf{m} \) is the vector of six components of relative MT \( \mathbf{M} \)

\[
\mathbf{m} = [M_{11} \ M_{22} \ M_{33} \ M_{23} \ M_{13} \ M_{12}]^T.
\]  

(6)

and \( \mathbf{u} \) is the vector of wave amplitudes at \( K \) one-component sensors. Quantities \( G_{k,i} \) are the components of the Green’s function matrix for the \( k \)th sensor:

\[
G_{k1} = G_{k1,1}, \quad G_{k2} = G_{k2,2}, \quad G_{k3} = G_{k3,3}, \quad G_{k4} = G_{k4,4},
\]

\[
= G_{k2,3} + G_{k3,2}, \quad G_{k5} = G_{k4,5} + G_{k3,4}, \quad G_{k6} = G_{k1,6},
\]

\[
= G_{k1,2} + G_{k2,1}.
\]

(7)

in which \( G_{k,i,m} = \partial G_{k,i} / \partial \xi_{m} \) is the spatial derivative of the Green’s function defined as the amplitude produced by the force couple at the source acting along the \( i \) axis with its arm in the \( m \) axis and recorded at the \( k \)th sensor along the sensor’s direction.

Because the MT inversion is linear, it is fast and computationally undemanding. In addition, the MT inversion produces the source time function, which is a time integral of the common wavelet (i.e., the first principal component function). The source time function can further be used for calculating the scalar moment and subsequently for proper scaling of the relative MT in equation (6).

**Inversion Steps and Code Structure**

The MT inversion of a single earthquake consists of data preprocessing and picking, alignment of traces, calculation of the PCA amplitudes, and the MT inversion using the PCA amplitudes (see Fig. 1). Individual steps of the inversion can be summarized as follows:

1. **Data preprocessing**
   - First, we oversample data to be able to perform an accurate alignment of waveforms by tiny shifts in the next processing of waveforms.
   - The oversampled data are band-pass filtered to enhance the signal-to-noise ratio (SNR).
   - A rough alignment of waveforms is done using manual picking, if available, or using an automatic picking algorithm. In the automatic processing, we use the Suspension Bridge Picking (SBPx) algorithm of Meier (see Data and Resources), based on optimizing a ratio of integrated weighted amplitudes from before and after the candidate pick found by the standard short-term average/long-term average algorithm.

2. **Two-step accurate alignment of waveforms**
   - The waveforms at all stations are shifted in time to maximize the cross correlation with the waveform of the highest SNR. Such aligned data are used for computing the principal components.
   - The waveforms at all stations are aligned using the cross correlation with the computed first principal component.
   - The aligned data are again decomposed using the PCA, and the principal components are refined.

3. **Calculation of the PCA amplitudes and the MT inversion**
   - We compute the ratio between the maximum of the first and second refined principal components to check whether the method is applicable. Because the waveforms at various stations should be similar (except for amplitudes), the first principal component must be dominant. If not, the point-source approximation assumed in the inversion is not valid, and the inversion might fail.
The PCA coefficients of the first principal component serve as the effective amplitudes (including their polarity) used for the standard-amplitude MT inversion. The correlation coefficients between individual traces and the first principal component serve as the weights in the linear MT inversion scheme. If a wavelet at a station is significantly different from the common wavelet found by the PCA, its correlation coefficient is low, and its weight is suppressed in the inversion.

The inversion produces a set of MTs because it is run repeatedly for several alternative band-pass filters. The optimum MT is that which produces the minimum root mean square (rms) of differences between synthetic and observed amplitudes. The optimum MT is relative because it is calculated from amplitudes but not from the low-frequency asymptote of displacement records.

- The scalar moment of the optimum MT is calculated by integrating the common (displacement) wavelet.

A crucial part of the PCA MT inversion, which calculates the PCA wavelet and the PCA amplitudes from input traces, is coded in MATLAB and freely available on the web (code PCA-DECOMPOSITION, see Data and Resources).

**NUMERICAL TESTS**

In this section, we perform numerical tests illustrating the effectiveness and robustness of the proposed inversion scheme. The velocity model, station configurations, event location, and focal mechanism are used to mimic observations of microearthquakes in the West Bohemia region used as an example in the Application to the 2014 Earthquake Sequence in West Bohemia section. We use synthetic data of one event recorded by two station configurations consisted of 8 and 20 stations, respectively. The input data are contaminated by various levels of noise. The depth of the event is 9 km, and the epicentral distance of stations is less than 20 km. We assume a shear earthquake with the focal mechanism defined by strike, dip, and rake of 170°, 70°, and −45°, respectively (see Fig. 2). This focal mechanism is typical for the activity in West Bohemia and represents the principal focal mechanism in the region (Vavryčuk, 2011).

The Green’s function needed for calculating synthetic amplitudes, and subsequently for inverting for the MT, was calculated for a 1D gradient isotropic velocity model using the ray method (Červený, 2001). The arrival times of the direct P waves are calculated by the kinematic two-point ray tracing. Geometrical spreading along the ray is obtained from the width of an elementary ray tube constructed from neighboring rays. The ray amplitudes include the conversion coefficients at the free surface. The source time rate function is formed by two one-sided pulses of different amplitudes (see Fig. 3). The total width of the pulse is about 0.18 s. The prevailing frequencies range from 10 to 16 Hz, corresponding to magnitudes of 2.5–1.5. The data are contaminated by random white noise with a uniform distribution. The noise level ranges from 0% to 200% of the maximum amplitudes.

A crucial part of the PCA MT inversion, which calculates the PCA wavelet and the PCA amplitudes from input traces, is coded in MATLAB and freely available on the web (code PCA-DECOMPOSITION, see Data and Resources).
Amplitude inversion

PCA inversion

Figure 6. The root-mean-square (rms) residuals of the amplitude and PCA inversions as a function of errors in time arrivals and noise in data. (Upper row) The amplitude inversion and (lower row) the PCA inversion. The rms is calculated using equation (9).

The vertical synthetic noisy P waveforms were generated 50 times for each combination of random noise in amplitudes and in time shifts and inverted repeatedly for the MT to obtain statistically significant results. We applied two MT inversions: (1) the standard amplitude inversion and (2) the PCA-based inversion, and compared their accuracy. The inversions differ in the way the P-wave amplitudes are measured at synthetic waveforms. In the amplitude inversion, the maximum amplitudes of the P waves are measured and inverted for the MT. In the PCA inversion, the effective amplitudes of the P waves are computed according to the Inversion Steps and Code Structure section and inverted for the MT.

The errors produced by the amplitude and PCA inversions are quantified by calculating the deviation of the DC component of the retrieved MTs from that of the synthetic MT. The DC deviation $\delta$ is computed as the average of deviations of the P/T axes of the retrieved focal mechanisms (defined by direction vectors $\mathbf{p}$ and $\mathbf{t}$) from the P/T axes of the true focal mechanism (defined by direction vectors $\mathbf{p}_{true}$ and $\mathbf{t}_{true}$):

$$
\delta = \frac{1}{2} (\cos(\mathbf{p} \cdot \mathbf{p}_{true}) + \cos(\mathbf{t} \cdot \mathbf{t}_{true})),
$$

in which the dot means the scalar product. The deviation $\delta$ calculated for individual retrieved MTs is then averaged over 50 realizations of random noise. Another option for quantifying the DC deviation could be, for example, calculating the Kagan angle (Kagan, 1991); see also a discussion in Tape and Tape (2012).

In addition, we quantify the inversion errors by calculating the non-DC components of the MTs, which are decomposed into the isotropic (ISO) and compensated linear vector dipole (CLVD) components. Advantages of this MT decomposition and properties of other possible MT decompositions are discussed in detail in Vavryčuk (2015). Because the true MT is pure DC, the true values of the ISO and CLVD components are zero. Consequently, the non-DC components of the retrieved MTs are spurious and directly measure the errors of the inversion. The percentages of the ISO and CLVD components were calculated using equations (6)–(10) of Vavryčuk (2015). The percentages are averaged over 50 realizations of random noise similarly as for the DC deviation.

Figure 5 indicates that both inversions are equally insensitive to errors in arrival times produced usually using a simplified velocity model in the Green’s function calculations. The insensitivity to inaccurate arrival times of waves is a great advantage of the amplitude inversion compared to the inversion applied to waveforms with no time-shift corrections. The PCA inversion also works well because the P waveforms at all stations are aligned by the two-step cross-correlation procedure; see Figure 1 and the Inversion Steps and Code Structure section. However, as regard the sensitivity of the MT inversions to noise in amplitudes, the performance of both approaches is different. Figure 5 demonstrates a clear superiority of the PCA inversion over the amplitude inversion. The PCA inversion produces lower errors in the orientation of the DC component as well as lower non-DC components of the MT. This applies to both station configurations, with 8 stations (upper panels) and 20 stations (lower panels). Because the true MT is pure DC, the true values of the ISO and CLVD are zero (blue color in panels in the middle and right columns of Fig. 5).

In addition, we observe that the spurious CLVD is about three times higher than the spurious ISO, indicating that the CLVD component is significantly more sensitive to numerical errors of the inversion than the ISO component.

A better performance of the PCA inversion compared to the amplitude inversion is also documented in Figure 6, showing the rms of the MT solutions defined as the normalized differences between the synthetic and observed amplitudes in the MT inversion:

$$
\text{rms} = \sqrt{\frac{\sum (A_i^{synth} - A_i^{obs})^2}{\sum (A_i^{synth})^2}},
$$

in which $i$ defines the station component, and the summation is over all stations and components which recorded the analyzed event. Obviously, the inversion of noise-free data produces the correct MT and the zero rms. When inverting noisy data, the rms is nonzero and increases with increasing noise. This increase is, however, different being lower for the PCA
inversion than for the amplitude inversion. This indicates that calculating the common wavelet by applying the PCA to waveforms before the inversion reduces noise in the input data efficiently.

**APPLICATION TO THE 2014 EARTHQUAKE SEQUENCE IN WEST BOHEMIA**

The West Bohemia region is situated in the western part of the Bohemian Massif at the contact of three Variscan structural units: (1) Saxothuringian, (2) Teplá-Barrandian, and (3) Moladanubian (Babuška et al., 2007). The triple junction of the units is characterized by thinning of the crust with depth of Moho ranging from 28 to 32 km (Hrubcová et al., 2013).

The tectonic structure of the area is characterized by two main fault systems: (a) the Mariánské Lázně northwest–southeast fault system and (b) the Ore-Mountain west-southwest–east-northeast fault system (Fig. 7b). However, the recently active faults in the area are the left-lateral strike-slip fault in the north–south direction with a strike of N169°E and the right-lateral strike-slip fault in the northwest–southeast direction with a strike of N304°E. The seismically active faults were identified at depth by foci clustering and by focal mechanisms (Vavryčuk et al., 2013), but they have also some geological evidence on the surface (Bankwitz et al., 2003). The maximum compressive stress determined from focal mechanisms has an azimuth of N146°E (Vavryčuk, 2011). The stress analysis indicates that the two active faults in the region with strikes of N169°E and of N304°E are optimally oriented for shearing with respect to the tectonic stress, thus forming a pair of conjugate principal faults in the area. The deviation of the principal faults from the $\sigma_1$ axis is about 32° and corresponds to fault friction of 0.5.

**Data**

The West Bohemia swarm area is characterized by persistent seismic activity with a periodic occurrence of earthquake swarms. Prominent earthquake swarms occurred in 1985/1986, 1994, 1997, 2000, 2008, 2011, and 2014 at the same epicentral area (Fig. 7), called the Nový Kostel seismic zone (Fischer et al., 2014; Čermáková and Horálek, 2015; Hainzl et al., 2016). The swarm duration was typically from two weeks to two months, and the activity was focused at depths of 6–11 km (Fig. 7c,d). The strongest instrumentally recorded earthquake was the $M_L$ 4.6 earthquake on 21 December 1985.

The most recent major seismic activity occurred in 2014 (Hainzl et al., 2016). It covered a period of almost three months and consisted of three distinct phases, with the strongest events on 24 May ($M_L$ 3.5), 31 May ($M_L$ 4.2), and 3 August ($M_L$ 3.6). The seismic activity was monitored by 22 short-period three-component seismic stations of the West Bohemia Network (WEBNET) consisting of 13 online and 9 offline stations. The stations have epicentral distances less than 25 km and cover the area uniformly with no major azimuthal gaps (Fig. 7b). Most of the stations are short period with a sampling frequency of 250 Hz and a flat frequency response between 1 and 60 Hz. In addition, the station with the nearest epicentral distance (NKC) was equipped by a broadband STS-2 seismometer. In total, about 5600 events with the local mag-
magnitude $M_L$ larger than $-0.5$ were detected during the 2014 activity.

We analyzed 833 microearthquakes with magnitude $M_L$ larger than 0.5 recorded at least at 16 WEBNET stations. The data set covers the whole period of the activity, including the three strongest events. The $P$ and $S$ arrival times were picked manually for the online stations and automatically using the SBPx algorithm of Meier (2015) for the offline stations. The locations of the events were calculated using the NonLin-Loc code (Lomax et al., 2000). The 1D-layered velocity model proposed by Málek et al. (2005) was used for locating the events. The Green’s functions were computed by the ray method (Červený, 2001) using a 1D-gradient velocity model obtained by smoothing the layered model of Málek et al. (2005). The velocity records were oversampled from 250 Hz to 1 kHz and band-pass filtered (see Fig. 8). We applied the two-sided Butterworth filters with the low corner frequency of 1 Hz and with the high corner frequency alternatively of 6, 8, 10, and 12 Hz. The PCA analysis was run on data filtered in several frequency bands because analyzed earthquakes cover a wide range of magnitudes from 0.5 up to 4.2, so the predominant frequency of the $P$ wave can vary. Based on the $P$-wave arrival times, we cut the window containing the direct $P$ wave (see Fig. 8b). Before calculating the effective $P$-wave amplitudes using the PCA, we applied the two-step alignment, as described in the Inversion Steps and Code Structure section and computed the common wavelet (see Fig. 8c,d). The computed PCA amplitudes corresponding to the individual frequency bands are used in the MT inversion. The optimum frequency band is that which produces an MT with the lowest rms.

Retrieved Moment Tensors
The retrieved MTs of the 833 analyzed earthquakes are differently accurate. This is indicated by a large scatter of the rms of the MT solutions (see Fig. 9). Therefore, we eliminated unreliable solutions and selected the 440 most accurate MTs by applying the two quality criteria: (1) the rms lower than 0.3 (see Fig. 9) and (2) the number of stations recording the earthquakes higher or equal to 18. As seen from Figure 10, the clusters of the $P$ and $T$ axes of the reliable solutions do not overlap and are well separated. However, both $P$ and $T$ clusters are rather large with a varying density (Fig. 10b) and can be divided into several smaller subclusters corresponding to several distinct types of MTs. For example, the $T$ axis is nearly horizontal for a majority of earthquakes, but some earthquakes are anomalous, having the $T$ axis nearly vertical (blue plus signs close to the center of the focal sphere in Figure 10a). By contrast, the non-DC components behave differently. They form just one cluster elongated along the CLVD axis (Fig. 10c,d). As observed in the synthetic tests, the higher variation of the CLVD can partly be caused by a higher sensitivity of the CLVD to numerical errors of the inversion. The CLVD and ISO components are negative for a majority of MTs.
their Euclidean norm: we apply the definition proposed by Willemann (1993):

\[ M^{(1)} = \sqrt{\frac{1}{2} M_{ij}^{(1)} M_{ij}^{(1)}} \]

and \[ M^{(2)} = \sqrt{\frac{1}{2} M_{ij}^{(2)} M_{ij}^{(2)}} \] (Silver and Jordan, 1982). The cosine distance in equation (10) measures the angle between two MTs. This angle is, however, different from the frequently used Kagan angle (Kagan, 1991). We do not use the Kagan angle because it can be applied to DC MTs only, and its physical interpretation is controversial (Willemann, 1993). A distance measure of MTs similar to equation (10) is also applied by other authors (Tape and Tape, 2012; Cesca et al., 2014).

Figure 11 and Tables 1 and 2 show the results of the classification of the MTs into five clusters. They present the number of earthquakes in individual clusters, the centroid moment tensors (CMTs) and their DC and non-DC components, and the mean rms of earthquakes in the clusters. The MTs are most frequently grouped in clusters 1 and 2, representing the left-lateral strike slips with a weak reverse component. The less frequent MTs are those in cluster 4, which are almost purely reverse, and those in cluster 5, which are the right-lateral strike slips with a normal component. The left-lateral strike slips (clusters 1–3) are associated with the most active fault in the focal zone having the strike between 155° and 170°. This fault was active during all prominent earthquake swarms in the last 40 years. The right-lateral strike slips (cluster 5) are associated with a less active fault with strike of 300°–310°. These mechanisms appeared, for example, in the 2008 swarm activity. Both faults are symmetrically oriented with respect to the maximum compression of the regional tectonic stress and are close to the principal faults in the region (Vavryčuk, 2011).

Interestingly, the almost pure reverse focal mechanism of cluster 4 is anomalous. It is not well oriented for shearing under the regional tectonic stress, and it has not been detected in the previous seismic activity in the region. Because the three strongest microearthquakes of the 2014 activity display this type of the focal mechanism, the 2014 activity seems to be exceptional and distinctly different from the previous seismic activities. This observation is also supported by anomalies in the magnitude–frequency relations found by Hainzl et al. (2016).

The CMTs differ not only in the orientation of the DC component but also in the percentage of their DC and non-DC components. Clusters 1 and 3 are formed by almost pure DC earthquakes (DC > 99%), with the mean rms of 0.22 and 0.27, respectively. By contrast, cluster 2 consists of earthquakes...
with the lowest DC percentage (DC ≈ 75%), with the mean rms of 0.13. No positive correlation between the percentage of the non-DC components and the rms indicates that the non-DC components are not errors produced by the inversion but that they are of physical origin. Except for cluster 3, which has negligible non-DC components, the ISO and CLVD are negative. The negative values of the ISO and CLVD have also been observed in other earthquake swarms in the studied area being interpreted in terms of fault weakening due to erosion by fluids manifested by shear-compressive faulting (Vavryčuk and Hrubcová, 2017).

**DISCUSSION AND CONCLUSIONS**

The numerical tests demonstrated that the PCA-based MT inversion is an efficient and robust tool for determining MTs of microearthquakes. The method is less sensitive to seismic noise in data than the amplitude inversion. It also suppresses other unmodeled phenomena, such as a directivity of the source, errors caused by local site effects at individual stations, and time shifts in arrivals of observed and synthetic signals due to an inaccurate velocity model. The method is fast and less computationally demanding than the waveform inversion. It can work in automatic or semiautomated regime, and hence it is suitable for processing large sets of earthquakes.

The analysis of the 2014 earthquake activity in West Bohemia using the PCA-based MT inversion revealed several distinct clusters of the MTs. The CMTs differ by the orientation of their DC component as well as by the amount of the non-DC components. Three CMTs are left-lateral strike slips associated with the most active fault in the focal zone. Interestingly, two of them are almost shear, with the DC higher than 94%, and the third type is slightly non-DC (DC = 75%), with a negative CLVD and ISO. We identified two additional clusters: (1) one cluster with right-lateral strike slips and (2) one cluster with reverse focal mechanisms. Both CMTs have DC about 80%, with a negative CLVD and ISO. The right-lateral strike slips are associated with another fault previously active in the focal zone. By contrast, the reverse focal mechanisms in the area are anomalous and unexpected. They have not been detected in the previous seismic activity in the region, and they also have an unfavorable orientation with respect to regional tectonic stress. This might indicate a presence of local stress heterogeneities caused, for example, by an interaction of faults or fault segments in the focal zone.

![Figure 11. Classification of the 440 reliably determined MTs into five clusters. (a) The focal sphere with the P/T axes. (b) Histogram of the number of events in individual clusters. (c) Focal mechanism plots corresponding to the centroid moment tensors of individual clusters. The individual MT clusters are color coded. The black open circles and plus signs in (c) mark the P and T axes, respectively.](image-url)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of Events</th>
<th>Mean rms</th>
<th>M11</th>
<th>M22</th>
<th>M33</th>
<th>M23</th>
<th>M13</th>
<th>M12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120</td>
<td>0.221</td>
<td>0.4100</td>
<td>-0.1750</td>
<td>-0.2747</td>
<td>-0.3915</td>
<td>0.0445</td>
<td>0.4557</td>
</tr>
<tr>
<td>2</td>
<td>112</td>
<td>0.130</td>
<td>0.2224</td>
<td>0.2572</td>
<td>-0.5911</td>
<td>-0.2583</td>
<td>0.2314</td>
<td>0.3836</td>
</tr>
<tr>
<td>3</td>
<td>106</td>
<td>0.274</td>
<td>0.3688</td>
<td>-0.4224</td>
<td>0.0661</td>
<td>-0.2056</td>
<td>-0.1597</td>
<td>0.5223</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>0.220</td>
<td>-0.3519</td>
<td>-0.1596</td>
<td>0.4556</td>
<td>-0.3868</td>
<td>0.2784</td>
<td>0.3073</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>0.184</td>
<td>-0.2355</td>
<td>0.4940</td>
<td>-0.3239</td>
<td>-0.3125</td>
<td>0.3250</td>
<td>0.3073</td>
</tr>
</tbody>
</table>

Number of events is the number of earthquakes in individual clusters; mean rms is the root mean square averaged over earthquakes in individual clusters. Components of the CMTs are expressed in the Cartesian coordinate system with \( x_1 = \text{N}, x_2 = \text{E}, x_3 = \text{down} \).
DATA AND RESOURCES

The data used in this study are available from the authors upon request. The MATLAB codes for the principal component analysis (PCA) analysis of waveforms are available at http://www.ig.cas.cz/pca-decomposition (last accessed June 2017). The MT decomposition and visualization (Fig. 10) were performed using the open-access MATLAB code MT-Decomposition available at http://www.ig.cas.cz/mt-decomposition (last accessed June 2017). The Suspension Bridge Picking (SBPxs) algorithm of Meier available at https://www.mathworks.com/matlabcentral/fileexchange/51996-suspension-bridge-picking-algorithm–sbpx– (last accessed June 2017) was used for the automatic picking of data.

ACKNOWLEDGMENTS

The authors thank Carl Tape and one anonymous reviewer for their helpful comments. This study was supported by the Grant Agency of the Czech Republic, project number 16-19751J.

REFERENCES


Bankwitz, P., G. Schneider, H. Kämpf, and E. Bankwitz (2003). Structural characteristics of epicentral areas in central Europe: Study case Cheb basin (Czech Republic), J. Geodyn. 35, 5–32.


Table 2

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Strike1 (°)</th>
<th>Dip1 (°)</th>
<th>Rake1 (°)</th>
<th>Strike2 (°)</th>
<th>Dip2 (°)</th>
<th>Rake2 (°)</th>
<th>DC (%)</th>
<th>CLVD (%)</th>
<th>ISO (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>156.7</td>
<td>70.8</td>
<td>−37.5</td>
<td>260.9</td>
<td>54.9</td>
<td>−156.3</td>
<td>93.7</td>
<td>−4.5</td>
<td>−1.8</td>
</tr>
<tr>
<td>2</td>
<td>162.0</td>
<td>52.9</td>
<td>−53.2</td>
<td>290.9</td>
<td>50.3</td>
<td>−128.4</td>
<td>74.5</td>
<td>−20.8</td>
<td>−4.8</td>
</tr>
<tr>
<td>3</td>
<td>342.9</td>
<td>82.3</td>
<td>20.8</td>
<td>249.9</td>
<td>69.4</td>
<td>171.7</td>
<td>97.6</td>
<td>1.8</td>
<td>0.6</td>
</tr>
<tr>
<td>4</td>
<td>40.8</td>
<td>66.8</td>
<td>78.1</td>
<td>248.9</td>
<td>25.9</td>
<td>115.7</td>
<td>82.6</td>
<td>−15.0</td>
<td>−2.5</td>
</tr>
<tr>
<td>5</td>
<td>305.0</td>
<td>49.1</td>
<td>−153.1</td>
<td>196.7</td>
<td>70.0</td>
<td>−44.1</td>
<td>78.9</td>
<td>−18.2</td>
<td>−2.9</td>
</tr>
</tbody>
</table>

The DC, isotropic (ISO), and compensated linear vector dipole (CLVD) percentages were calculated using equations (6)–(10) of Vavryčuk (2015).


