F013

An Approach to Analyse Microseismic Event Similarity

M. Scholze* (Free University of Berlin), K. Stürmer (Free University of Berlin), J. Kummerow (Free University of Berlin) & S.A. Shapiro (Free University of Berlin)

SUMMARY

We introduce two parameters to analyse the similarity of close seismic events. The first is an extension of the normalised cross correlation coefficient (composite correlation measure, CM), and the second parameter depends on differential arrival times (composite differential arrival time measure, TM). TM quantifies the separation of pairs of seismic events. The relationship between TM and interevent distance Delta-r is derived, and theoretical TM-Delta-r-distributions are calculated for different source-receiver geometries. We find that a combined analysis of the three parameters CM, TM and Delta-r is practical for identification of time measurement errors and mislocations. The envelope of the CM-TM plot delineates event pairs with the same source mechanism (multiplets). We define an area around this envelope in order to account for statistical uncertainties. Therefore we develop two different approaches to obtain a proper measure of this area.
Introduction

The cross correlation coefficient is widely used in seismology to quantify the waveform similarity, e.g., between the seismograms of two events recorded at the same station. It is applied to improve the accuracy of differential arrival times (e.g., Rowe et al. (2002), Schaff et al. (2004)), to define clusters of seismic events (e.g., Aster and Scott (1993), Arrowsmith and Eisner (2006), Baisch et al. (2008)) or estimate interevent distance (Israelsson (1990), Menke (1999), Kummerow (2008)). Plots of correlation coefficients as a function of interevent distance are often contaminated by location errors. This is the motivation in the present study to define a robust measure, which characterises the separation between earthquakes. Our approach is performed in the time domain and therefore eliminates the dependency on an often inadequate velocity model.

Method

The normalised cross correlation function \( c_{ij}(t) \) between two seismogram traces \( u_i(t) \) and \( u_j(t) \) is defined as

\[
c_{ij}(t) = \frac{\int u_i(t') u_j(t' - t) dt'}{\left( \int u_i^2(t') dt' \times \int u_j^2(t') dt' \right)^{1/2}}
\]  

(1)

The maximum absolute value of \( c_{ij}(t) \) is generally referred to as the cross correlation coefficient \( C_{ij} = \max ||c_{ij}(t)|| \). It is a convenient measure of waveform similarity. The values range between 0 (no similarity) and 1 (identical waveforms). The cross correlation coefficient \( C_{ij} \) can be calculated for the complete seismograms or separately for specified time windows (e.g., P and S time windows). If we now consider \( k = 1, \ldots, M \) seismic stations, then a composite correlation measure, \( CM \), can be defined by

\[
CM_{ij} = \left[ \prod_{k=1}^{M} \left( P C_{ij}^k \right) \times \prod_{k=1}^{M} \left( S C_{ij}^k \right) \right]^{1/M}
\]

with the maximum correlation coefficient \( P|S C_{ij}^k \) for event pair \( ij \) at station \( k \) and \( P \) or \( S \) time window, respectively. The range of values for \( CM \) is between 0 and 1, where 1 expresses identical seismogram traces of one event pair \( ij \) for all receivers. To be independent from localisations we further define a differential time measure of arrival time differences, \( TM \), as follows

\[
TM_{ij} = \left[ \prod_{k=1}^{M} 1 - \frac{\left| (k t_S^i - k t_S^j) - (k t_P^i - k t_P^j) \right|}{t_{norm}} \right]^{1/M}
\]

(3)

with normalised time \( t_{norm} \), which we chose as the maximum possible travel time difference \( t_S - t_P \) between receivers and grid point in the model.

\[
t_{norm} = \max \left( t_S - t_P \right)
\]

(4)

\( S - P \) differential arrival times \( k t_S^i - k t_S^j \) can, e.g., be measured by cross correlation of the two seismograms. The values of \( TM \) range from 0, if the differential travel time is equal to \( t_{norm} \), and 1 in case that the differential travel times to every event pair \( ij \) are identical for each station.

Synthetic modelling

Hence \( TM \) is based on differential travel time measurement a relationship between \( TM \) and interevent distance \( \Delta r \) can be derived. Figure 1 left shows a schematic source receiver geometry. For such a case a simplified expression for \( TM \) is found under the assumption of constant velocity. This leads to:

\[
TM_{ij} = \left[ \prod_{k=1}^{M} 1 - \frac{v_S^{-1} (k r_i - k r_j) - v_P^{-1} (k r_i - k r_j)}{t_{norm}} \right]^{1/M}
\]

\[
= \left[ \prod_{k=1}^{M} 1 - f \cdot \Delta r_{ij} \cdot |\cos \beta| \right]^{1/M}
\]

(5)

Where \( f \) is a constant factor for one event pair \( ij \).
\[ f = \frac{(v_P - v_S)}{v_S \cdot v_P \cdot t_{norm}} \]  

Equation 5 shows that \( T_{Mij} \) depends on \( f \), \( \Delta r \) and the source-receiver geometry (\(|\cos\beta|\)). From this we can conclude that \( TM \) cannot be assigned an explicit value for \( \Delta r \). Instead we expect that \( TM \) varies against the cosine of the source receiver geometry controlled opening angle \( \beta \). We decided to model synthetic source receiver geometries in order to study the behaviour of \( TM \) on different geometries. It appears to be reasonable to assume constant velocities for the model and a spherical source geometry. Receivers are placed randomly on the surface. The normalisation term \( t_{norm} \) is set as the maximum possible travel time within the model, see equation 4. \( TM \) will be calculated for all event pairs \( N \).

Figure 1 Left: Schematic source receiver geometry for one receiver and two events (Event 1 und Event 2). Right: Schematic source receiver geometries for two events (stars) and one receiver (triangles).

Figure 2 shows the result of the modelled data in a \( TM - \Delta r \) density plot. As shown in equation 5 the \( TM \) values for a specific value of \( \Delta r \) depends on the cosine of the opening angle \( \beta \) which results from the source receiver geometry. In addition all \( TM \) values for a certain \( \Delta r \) are arranged within a linear upper and lower envelope. The upper envelope represents those \( TM \) values with the highest source receiver angle \( \beta \) the lower envelope match with the lowest values for \( \beta \) correspondingly. Apparently there is a clear density distribution between the envelopes. This arises from the shape of the event cloud as from its position regarding the receivers. Figure 1 right a) shows a source receiver geometry where the cosines of the angle \( \beta \) is equal to 0 and the travel time is nearly equal. Therefore the upper envelope of the \( TM - \Delta r \) plot represents those events with the highest opening angle. In figure 1 right b) the contrary case is shown, the differential travel time is maximal and the cosine of the opening angle is equal to 1. Those cases represent the lower envelope. As a result we obtain from modelling, that a plane receiver array is the most adequate receiver array design. In this case the array averages the opening angles, thus the effect of source-receiver-geometry is minimised and comparable dataset is recorded. Since the values for \( TM \) plot against the intervent distance \( \Delta r \) within linear bounds we conclude, that localisation dependent \( TM - \Delta r \) plots are helpful to detect time measurement or localisation errors.

Real data example

We calculate \( TM \) and \( CM \) for localised microseismic data recorded at the West Fissure Fault System in Northern Chile, see Kummerow et al. (2008). Waveform similarity is high for most events \( (CM > 0.7) \). First we compare the obtained real data with the results from the \( TM \) modelling. The \( TM - \Delta r \) plot...
(figure 3 left) reproduces essentially the same results from the synthetic modelling. Most data points plot within the upper and lower envelope. As we examine a preselected cluster with a maximum interevent distance of \( \sim 700\text{m} \) those events which plot outside the bounds seem to be error-prone.

![Figure 3](image)

**Figure 3** TM as function of interevent distance \( \Delta r \). The left plot features the unsorted data. The right plot shows the sorted data using the criterion of the upper and lower envelope.

**Error analysis**

The event pairs plotting below the lower bound feature a time measurement error. Thus the TM values tend to be too low. Furthermore some event pairs plot above the upper envelope. These events denote mislocations and poorly resolved locations because the value for the interevent distance \( \Delta r \) appears to be too high. The \( CM - \Delta r \) plot features an upper bound resulting from scattering. With increasing interevent distance the values for \( CM \) have to decrease (Kummerow (2008)). Those events plotting above the upper bound exhibit a localisation error as the cross-correlation works accurate therfor the interevent distance has to be too large. In order to obtain more accurate data we applied a sorting procedure. We analysed the frequency of events taking part in event pairs which plot outside the envelopes. If a event is exceptionally often part of a such event pair (more than 8 times), this event will be sorted out. After this sorting procedure the remaining event pairs accord with the upper and lower bound in the \( TM - \Delta r \) plot, see figure 3 right. It becomes apparent that the maximum interevent distance reduces to \( \sim 450\text{m} \) after applying the sorting procedure. A total amount of 11 events seem to be error-prone and were out sorted. The original data set contains 120 events.

**Identification of similar events using a statistical approach**

Assuming, the dataset is free from errors after sorting, we apply our approach to the dataset by plotting the resulting CM–values of event pairs against the resulting \( 1 - TM \)–values and define the upper envelope as “multiplet line”. Hence pairs of multiplets should plot along this line. To soften this criterion we introduce an interval around the multiplet line that we call \( \delta \)–interval. A doublet is then defined as an event pair having a CM–TM value located in this interval around the multiplet line (ML). The resulting multiplet interval (MI) can be described as

\[
MI = ML \pm \delta ML
\]

Due to the fact the transition between being a multiplet and not being a multiplet is quite continuous, it is not possible to define a general \( \delta \)–interval. We try to find a statistically based definition of this interval by having a closer look to this parameter. A larger \( \delta \)–value results in a larger number of multiplets. To find an optimal \( \delta \)–interval one need to account for the following trade–off: more events having a poor similarity or less events having a strong similarity, respectively. Therefore we analyse the statistical behaviour of the cumulative sum of CM–TM values at increasing \( \delta \)–values. Hence, we count the number of considered event pairs within a certain multiplet interval, determined by \( \delta \), until the mean increase of number of event pairs begins to fall. This point is reached where the gradient of the envelope function of the cumulative sum is maximum (see figure 4).
Figure 4 Results of our statistical approach. Left top: Cumulative sum of event pairs in dependence to $\delta$. Bottom: First derivation of the envelope function for the cumulative sum of event pairs in dependence to $\delta$. A maximum peak is clearly visible at $\delta = 0.035 = 3.5\%$. Right: Event pairs and their corresponding CM–TM values. Multiplets are events located within the multiplets interval (red circles).

Conclusion and Outlook

We define a composite correlation measure, $CM$, and a composite arrival time measure, $TM$. The relation between $TM$ and interevent distance $\Delta r$ is derived. Plots of $CM(\Delta r)$, $TM(\Delta r)$ and $CM(TM)$ are useful to identify errors in time measurements, poorly resolved locations and location errors. The envelope of $CM(TM)$ can be considered as a diagnostic tool to identify event pairs as doublets in the sense that they have the same source radiation pattern. This definition characterises event pairs as a doublet, if they have the same source mechanism, even if they originate at different positions of a fault segment. $\Delta r$ may be larger than a quarter of the dominant wavelength $\lambda$, which is often taken as the criterion for a doublet definition (e.g., Baisch et al. (2008) and references therein). We present a statistical approach to define multiplets as event pairs located in a certain multiplets interval determined by a heuristic parameter $\delta$. Our long term target is to develop a semi– or fully automated software for analysing microseismic datasets regarding the presence of multiplets and their physical meaning.

Acknowledgments

We thank the sponsors of the PHASE consortium for supporting the research presented in this paper.

References