Performance of different processing schemes in seismic noise cross-correlations

J.C. Groos,[1] S. Bussat[2] and J.R.R. Ritter[1]

[1]*Karlsruhe Institute of Technology, Geophysical Institute, Hertzstrasse 16, 76187 Karlsruhe, Germany. E-mail: joern.groos@kit.edu*
[2]*Statoil ASA, Postboks 7200, NO-5020 Bergen, Norway*

GJI Seismology

SUMMARY

The estimation of the Green's function between two points on the Earth's surface by the cross-correlation of seismic noise time-series became a widely used method in seismology. In general, very long time-series (months to years) as well as massive normalization and/or data selection are necessary to obtain useful cross-correlation functions. One task of this study is to evaluate the influence of different established normalization methods on the obtained cross-correlation functions. Furthermore, we evaluate two waveform preserving time domain normalizations as well as a new fully automated data selection approach. The cross-correlation functions analysed in this study are obtained from 12 months of seismic noise recorded in 2004 at five seismic stations in the United States with station distances on a continental scale. For practical reasons, the cross-correlation functions of such long time-series are calculated by stacking the cross-correlation functions obtained from shorter time windows. We use this stacking process for the implementation of the waveform preserving time domain normalizations. The time window length is in general an important parameter of the cross-correlation processing, as it influences the normalization and data selection. Therefore, we evaluate the cross-correlation functions obtained with 47 different time window lengths between one hr and 24 hr. The time domain normalizations intend to suppress the influence of transient signals like earthquake waves as well as long-term (e.g. seasonal) amplitude variations. We compare the proposed waveform preserving time domain normalizations with the established running absolute mean normalization and the one-bit normalization. We demonstrate that a waveform preserving time domain normalization can replace a non-linear time domain normalization, if a time window length similar to the duration of the typically occurring transient signals is used. Next to the time domain normalizations also the spectral whitening in the frequency domain is evaluated. Spectral whitening is a powerful normalization to improve the emergence of broad-band signals in seismic noise cross-correlations. Nevertheless, we observe spectral whitening to depend strongly on the time window length. An unwanted amplification of a persistent microseism signal is observed on the continental scale with time windows shorter than 12 hr. Our approach of automated data selection is based on a statistical time-series classification and reliably excludes time windows with transient signals occurring contemporaneously at both sites (e.g. earthquake waves). This data selection approach is capable to replace a non-linear time domain normalization, but no improvement of the waveform symmetry or the signal-to-noise ratio of the cross-correlation functions is observed in general.

Keywords: Time series analysis; Interferometry; Surface waves and free oscillations; Statistical seismology; Wave propagation.

1 INTRODUCTION

The estimation of Green's functions based on seismic noise cross-correlation functions (CCFs) evolved to an important and widely used technique in seismology (Weaver 2005; Curtis *et al.* 2006). It enables seismology to provide high-resolution tomography studies from local (e.g. Bussat & Kugler 2011) to continental (e.g. Shapiro *et al.* 2005) scale and independent from earthquake seismicity or active seismic sources. Nevertheless, practical experience shows that on a continental scale one has to use long time-series (months to years) and to apply extensive normalization to the seismic noise time-series to obtain CCFs which are suitable to reliably estimate Green's functions (Gouédard *et al.* 2008; Roux 2009; Yao & van der Hilst 2009). The symmetry of the seismic noise CCF around

lag time zero is the simplest criterion to evaluate its suitability to estimate the Green's function. From the theoretical point of view we expect symmetric, in terms of waveform and amplitude, CCFs, which are characterized by identical causal and acausal parts (Sabra *et al.* 2005). The cross-correlation of non-normalized 'raw' seismic noise time-series produces CCFs which are in general not suitable to estimate Green's functions due to disturbing dominant signals or instrumental irregularities. The mildest form of an unsuitable CCF is an asymmetrical CCF, but even the presence of any kind of clearly identifiable 'seismic' signal in raw seismic noise CCFs is not a matter of course. The data processing applied to obtain suitable seismic noise CCFs is therefore critical and underwent some kind of evolution during the last years. A current status of ambient seismic noise data processing was published by Bensen *et al.* (2007) and is referenced in more than 80 follow-up publications. Thus, the recommendations by Bensen *et al.* (2007) are established as some kind of processing standard. Nevertheless, differences in the actually applied data processing can be still found in the growing number of studies. This is especially true for the different types of normalization methods. The purpose of this paper is to investigate some critical aspects of seismic noise data processing like the time window length used to fragment long time-series (typically one day on a continental scale) and to study the performance of the most widespread time and frequency domain normalizations in dependence of this time window length. Furthermore, we propose an effective time domain normalization, which preserves the waveform of the seismic noise time-series as a further option. With this new time domain normalization no changes have to be applied to the original seismic noise data prior to the cross-correlation.

The established and wide spread non-linear time domain normalizations are the one-bit normalization (e.g. Shapiro *et al.* 2005) and the running absolute mean (RAM) normalization (Bensen *et al.* 2007). Bensen *et al.* (2007) state, that the RAM normalization is smoother and produces better results in terms of the signal-to-noise ratio (SNR) of the CCFs. In the frequency domain, the spectral whitening (SW) is the established method. The implementations of this method differ also between different authors (e.g. Bensen *et al.* 2007; Brenguier *et al.* 2008). A crucial point is that all established methods significantly distort the waveforms of the seismic noise time-series.

Another possibility to enhance seismic noise CCFs may be the selection of appropriate input data and the exclusion of problematic data pieces like (teleseismic) earthquakes and instrumental irregularities. Pedersen *et al.* (2007) applied a data selection approach based on global earthquake catalogues to exclude worldwide seismic events with magnitudes larger than five and a CCF amplitude threshold to exclude remaining small seismic events. The selection approach in general is also discussed by Bensen *et al.* (2007) but not further considered due to the difficulties by removing all time windows containing earthquake waves automatically based on earthquake catalogues or threshold-based methods.

We test a fully automated data selection approach independent of earthquake catalogues applying the noise classification method proposed by Groos & Ritter (2009). This method is capable to distinguish between normal distributed time-series and time-series dominated by transient signals. Time windows with contemporary transient signals at the two recording sites are problematic to obtain suitable seismic noise CCFs due to their dominance. This is also true, if both transient signals do not originate from the same source. Therefore, the CCFs obtained from such time windows may be better excluded from the waveform stacking. The seismic noise classification depends on the time window length in a selected frequency

band. Therefore, a main point of this study is the performance of the different normalization methods and the data selection approach with different time window lengths.

We use several parameters (e.g. SNR, Waveform Symmetry Coefficient (WSC) and amplitude percentiles) to evaluate the obtained CCFs. These parameters are introduced together with our data set and data processing in Section 2. In Section 3, we introduce and evaluate two waveform preserving time domain normalization methods. Our approach of automated data selection is discussed in Section 4. Then in Section 5 we analyse the effects of the time window length on the CCFs obtained with SW.

2 DATA SET, PROCESSING AND PARAMETERS

The data set used for this study are the vertical-component ground-motion velocity data of the year 2004 with a sampling rate of 1 Hz (LHZ channel) of the Global Seismographic Network (GSN) stations ANMO, CCM, DWPF, HRV and PFO in the United States (Fig. 1).

The waveforms and instrument characteristics were obtained as SEED volumes from the Incorporated Research Institutions for Seismology (IRIS) Data Management Centre. The CCFs on a continental scale were calculated and analysed for all 10-station pairs. The six station pairs ANMO–DWPF, ANMO–CCM, PFO–ANMO, DWPF–HRV, CCM–DWPF and PFO–HRV covering different station azimuths and distances are discussed in detail in this paper. CCFs of some of these station pairs were also discussed by Bensen *et al.* (2007).

2.1 Data processing

The waveform data of 2004 are obtained from the SEED volumes as fragmented time-series in fragments of different length (minutes to weeks). The mean value and the linear trend are removed from all time-series fragments before a cosine taper (4 per cent) and a zero-phase 0.003 Hz fourth-order high-pass filter are applied prior to the removal of the instrument response. Afterwards, all fragments are merged to a continuous 12-month time-series from 2004 January 1 to 2004 December 31. Missing data is zero padded to obtain complete time-series. The 12-month time-series are filtered again with the 0.003 Hz HP filter to remove low frequency artefacts due to the instrument response deconvolution. With this preparation we obtain the ground motion velocity in a broad frequency range. Prior to the cross-correlation the 12-month time-series are finally bandpass filtered with a fourth-order zero-phase filter in the period band of interest, which is 7–150 s in our case.
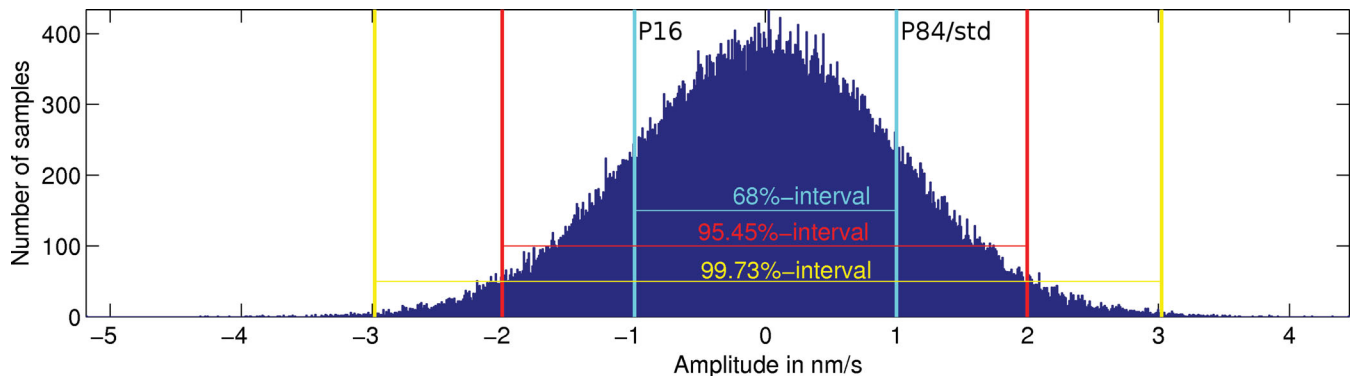
2.1.1 Noise classification

Prior to the cross-correlation each time-series is classified with the seismic noise time-series classification after Groos & Ritter (2009).

This classification uses ratios of time-series amplitude intervals (I) and percentiles (P) to identify and quantify the deviations of the time-series histogram from the Gaussian amplitude distribution. In the case of a zero mean Gaussian distribution (Fig. 2), 68 per cent of the measurements lie within an interval of one standard deviation away from zero (I68). 95.45 per cent of the measurements are within two times the standard deviation (I95) and 99.73 per cent of the measurements are within three times the standard deviation (I99). This is also known as the 2-σ and 3-σ, or the 'empirical', rule.
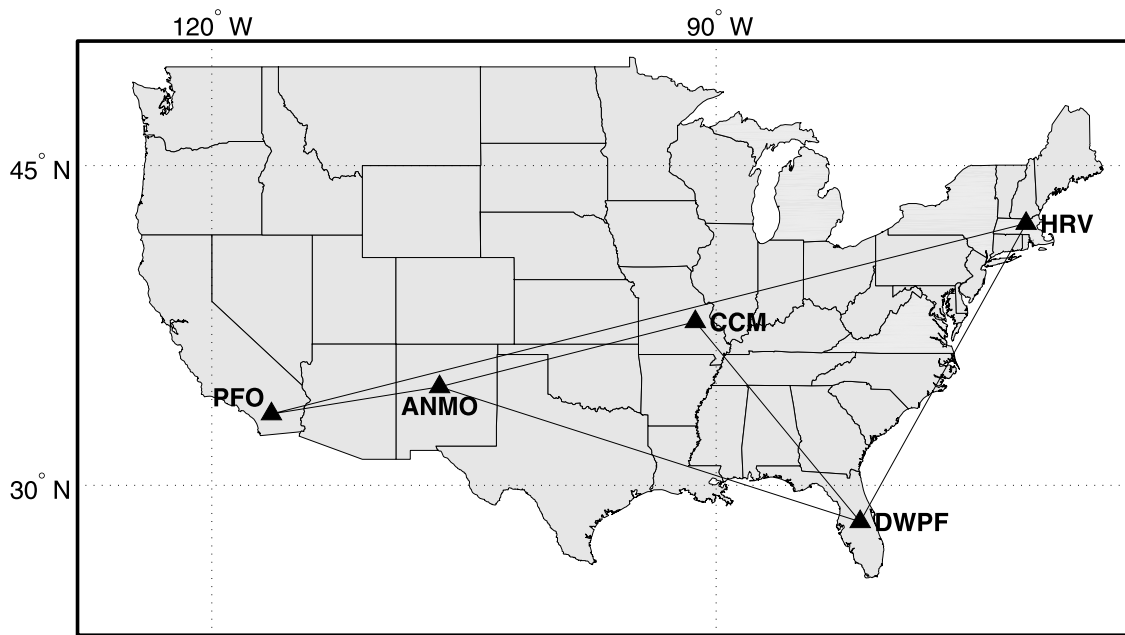
Figure 1. Map with used GSN stations ANMO (Albuquerque, New Mexico), CCM (Cathedral Cave, Missouri), DWPF (Disney Wilderness Preserve, Florida), HRV (Harvard, Massachusetts) and PFO (Pinon Flat Observatory, California). The IRIS Data Management System was used for access to waveform- and meta-data of these stations. The six station pairs analysed in detail are indicated by the connection lines.

Figure 2. Histogram of a Gaussian distributed time series with zero mean and a standard deviation of one. The amplitude intervals used for the seismic noise time-series classification after Groos & Ritter (2009) are annotated and marked by horizontal bars. The upper border of the 68 per cent amplitude interval (I68 in the text) is the 84-percentile (P84) and equals the standard deviation in the case of a Gaussian distribution.

Groos & Ritter (2009) introduce the ratio between I99 and I95 as the new quantity peakfactor (pf) to determine the 'peakedness' of the histogram. The pf of a Gaussian distributed time-series equals 1.5. Furthermore, the ratios of the lower and upper boundaries of the intervals (e.g. the 16-percentile (P16) and the 84-percentile (P84) for I68, Fig. 2) can reveal a possible asymmetry of the histogram (see Groos & Ritter (2009) for criteria). The ratios between the amplitude intervals (e.g. the pf) and the amplitude percentiles (e.g. P84/|P16|) of the time-series are used to classify the seismic noise time-series.

Six noise classes are introduced to classify the typically observed deviations of seismic noise time-series from the Gaussian distribution. Time-series are assumed to be Gaussian distributed (noise class 1, NC1) if the interval ratios exhibit only very small deviations (<3 per cent) from the empirical rule and the histograms are symmetric. Non-Gaussian but symmetric time-series are classified as NC2–NC5 depending on the observed deviation properties. Time-series with rather small and unspecific deviations from the Gaussian distribution (pf 1.5 ± 0.1) are classified as noise class 2 (NC2). Time-series with a gently peaked histogram in comparison to the Gaussian distribution ($1.6 < \text{pf} \leq 2$) due to few transient signals are classified as noise class 3 (NC3). A more pronounced peakedness of the histogram (pf > 2) results in a classification as NC4. Symmetric time-series with a flattened histogram in comparison to the Gaussian distribution (pf < 1.4) are classified as NC5. All time-series, which are not identified as symmetric time-series are classified as NC6. Furthermore, extreme absolute values of the amplitude intervals and/or amplitude interval ratios are used to identify reliably time-series of seismic data corrupted by technical problems such as data gaps, calibration pulses, offset changes or other aseismical signals. These corrupt time-series are classified as noise class 10 or higher and they are removed from the following cross-correlation processing.

In general, only the four noise classes NC1 to NC4 are relevant for analysis of the seismic noise for cross-correlations. Time-series of seismic noise classified as NC5 or NC6 are seldom observed (Groos

& Ritter 2009). Strong, short transient signals such as earthquake waves are identified mostly as NC4. This is used for the automated data selection introduced in Section 4.

2.1.2 Cross-correlation and stacking

The CCF is calculated as linear unbiased digital cross-correlation in the frequency domain (Bendat & Piersol 1994). We calculate the 12-month CCF by the stacking of several 'single-time window' CCFs obtained with a sliding overlapping time window. The sliding time window overlap equals the maximum analysed lag time of the calculated CCFs (2000 s in our case) to include all possible sample combinations for every lag time between 0 and 2000 s. We select a standard maximum lag time of 2000 s to include the signal time windows of all analysed station pairs and to simplify data handling. The CCFs obtained from time windows, which contain predominantly zeros (data gaps) are automatically excluded from the stacking.

2.1.3 Applied normalization methods

The established methods for the time domain normalization of the time-series (one-bit and RAM normalization) are applied prior to the cross-correlation. The RAM normalization weights every time-series value by the average of the absolute time-series values in a surrounding time window with the length of half of the largest analysed period of 150 s (Bensen *et al*. 2007). The one-bit (1B) normalization heavily distorts the original waveform by leaving only values of +1 for all positive and −1 for all negative amplitude values of the time-series. It is the extreme case of the RAM normalization with a time window length of one sample. The one-bit normalization is known to negatively influence the frequency content of broad-band time-series (Sabra *et al*. 2005; Pedersen *et al*. 2007). This effect is discussed in more detail together with the SW in Section 5. Notwithstanding the known shortcomings the simple 1B normalization is still widely used. Both time domain normalizations are used for the evaluation of the waveform preserving time domain normalizations discussed in Section 3 as well as the automated data selection discussed in Section 4.

For practical reasons the SW is applied after the cross-correlation and prior to the stacking of the single-time window CCFs. The numerical differences between the SW of the time-series prior to the cross-correlation and the SW of the CCFs after the cross-correlation can be neglected as discussed in Groos (2010). The significant influence of the time window length on the stacked CCFs obtained with SW are discussed in Section 5.

2.1.4 Time window length

The length of the sliding overlapping time window is varied for this study to analyse the influence of this technical processing parameter on the effect of the different normalization methods. Time window lengths between 1 hr (Li *et al*. 2010) and more often 24 hr (e.g. Bensen *et al*. 2007) were applied in published studies. The cross-correlation processing with daily time-series (24 hr time window) is the widely used standard for continental scale noise interferometry (e.g. Bensen *et al*. 2007; Yang & Ritzwoller 2008). Therefore, we focus our analysis on 47 different time window lengths between 1 hr and 24 hr by increasing the time window length step-wise by about 0.5 hr. We use the stacked 12-month CCF obtained with the 24 hr time window as reference CCF to analyse the variations of

the stacked CCFs caused by the different time window lengths. The application of the linear digital cross-correlation is mandatory for the calculation of continental scale seismic noise CCFs with a time window length shorter than 6 hr (Groos 2010).

### 2.1.5 Cross-correlation function processing

The broad-band (7–150 s) single-time window CCFs are stacked in the time domain to obtain the stacked 12-month CCF. The data selection is realized by the exclusion of the affected single-time window CCFs from the stacking. Prior to the stacking the mean value of the single-time window CCFs is removed. To analyse the broad-band CCF in a narrower frequency band a corresponding fourth-order zero-phase Butterworth bandpass filter can be applied to the stacked 12-month CCF. We analyse the broad-band stacked CCFs also in the narrower frequency bands 7–14 s, 20–50 s and 70–150 s to evaluate the frequency-dependent influence of the SW (Section 5).

2.2 Data parameters

We determine several parameters of the seismic noise time-series and the CCFs for evaluation. The determined parameters of the time-series are amplitude percentiles representing the 68 per cent, 95.45 per cent and 99.73 per cent amplitude intervals which are also used for the noise classification introduced in Groos & Ritter (2009). For the obtained CCFs the causal and acausal SNR are determined (Fig. 3). The SNR is calculated as the ratio between the peak signal in a signal time window and the root mean square (rms)-value of a noise time window (e.g. Bensen *et al*. 2007). The signal time window is selected individually for every station pair to contain Rayleigh waves with a propagation speed of 2.4–4.8 km s^{-1}. Furthermore, we are using the 'precursory' noise in a time window between lag time zero and the signal time window instead of the 'trailing' noise in a time window behind the signal time window (e.g. Bensen *et al*. 2007). The SNR calculated with the precursory noise time window depends more on the noise in the CCFs between lag time zero and the signal time window, which is generated by coherent transient signals (e.g. earthquake waves). For this reason, Bensen *et al*. (2007) suggest that the precursory noise SNR may be better to predict the quality of dispersion measurements from CCFs than the trailing noise SNR. Therefore, we use the precursory noise SNR for this study. The noise time window contains 60 per cent and neglects the first and the last 20 per cent of the CCF between lag time zero and the signal time window.

In addition to the SNR, the linear Pearson's correlation coefficient (Bendat & Piersol 1994) between the causal and acausal signal time window is calculated to evaluate the waveform symmetry of the cross correlation in the signal time windows (parameter WSC in Fig. 3). CCFs with high waveform symmetry but amplitude asymmetry may be interpreted to represent the different amount of seismic energy propagating in the two different directions of a station pair (Campillo 2006). CCFs with significant waveform and amplitude asymmetry may be unsuitable to estimate the Green's function at all (Yang & Ritzwoller 2008; Roux 2009). Therefore, an increase of waveform symmetry (or parameter WSC) may be a suitable criterion to evaluate the performance of a distinct processing scheme. The SNR and waveform symmetry measure WSC will be used to evaluate the quality of the CCFs obtained with the different processing schemes in Sections 3 and 4.

The variations of the 12-month stacked CCFs obtained with different time window lengths and/or normalization methods are
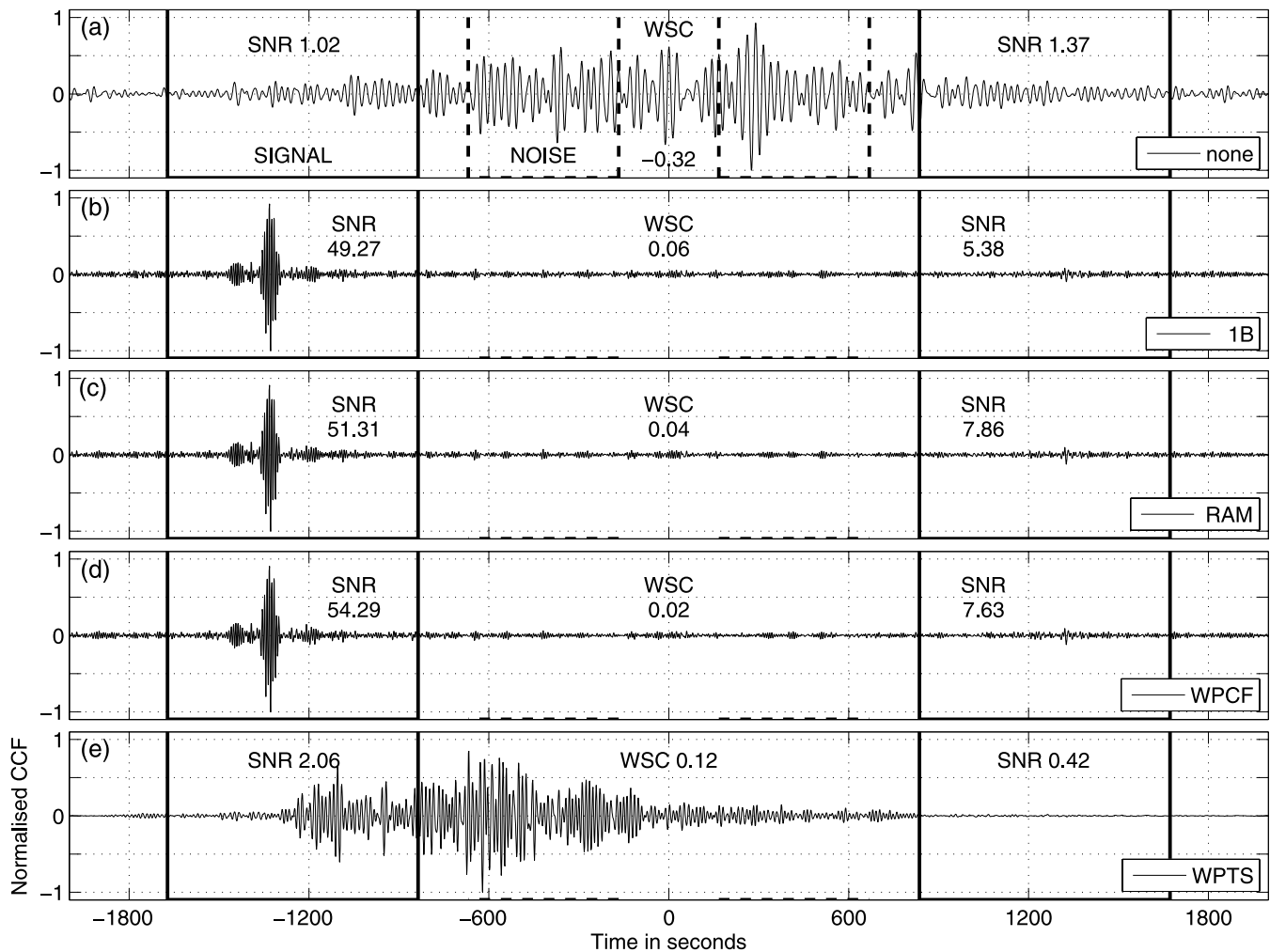
Figure 3. Stacked 12-month CCFs (PFO–HRV, 7–150 s, 2 hr time window) obtained with (a) none, (b) one-bit, (c) running absolute mean and, (d) waveform preserving normalization of the single-time window CCFs (WPCF normalization, see Section 3) and (e) waveform preserving normalization of the time-series (WPTS normalization). No spectral whitening is applied prior to the stacking. Only time windows without data are excluded from the stacking. The time windows used to calculate the SNR and the waveform symmetry coefficient (WSC) are displayed by solid (signal time windows) and dashed (noise time windows) boxes and bars. The SNR is given for the corresponding causal and acausal signal time window of the CCF. The WSC of the CCF is given at lag time zero.

also quantified with a correlation coefficient. The parameter CC (Fig. 4a,b) is the linear Pearson's correlation coefficient between the symmetric-component signal time windows of the compared CCFs. The symmetric-component CCF is the stack of the causal and acausal part of the CCF. The symmetric-component CCF is a widely used approach in continental scale seismic noise cross-correlation processing to deal with the typically asymmetric CCFs (Bensen *et al*. 2007). In general we compare the 12-month stacked CCFs obtained with different normalizations and/or time window length with our reference CCF obtained with the RAM normalization and a time window length of 24 hr. This corresponds to the processing scheme of Bensen *et al*. (2007).

3 WAVEFORM PRESERVING TIME DOMAIN NORMALIZATION

As mentioned above, all established normalization methods intend to equalize the signals contributing to seismic noise to enhance the

emergence of the Green's function and accept a distortion of the original seismic noise waveforms. The emergence of the Green's function in seismic noise CCFs is disturbed due to the impacts of dominating seismic signals such as transient earthquakes waves or monochromatic signals of ocean-generated microseism, even if very long time-series are used (Bensen *et al*. 2007).

Our approaches to improve the CCFs with a waveform preserving normalization benefit from the calculation of stacked CCFs with a sliding time window. The idea is to equalize the amplitude differences between the single-time window CCFs prior to the stacking to improve the emergence of the Green's function in the stacked CCF. To do so, we test two methods. The first one is to normalize both time-series before the cross-correlation and the second one is to normalize the single-time window CCFs after the cross-correlation prior to the stacking. The advantage of the second approach is that the original seismic noise time-series are not changed at all prior to the cross-correlation. Both normalizations are done by dividing the waveforms (CCFs or time-series, respectively) by an amplitude value and they are discussed in detail below.
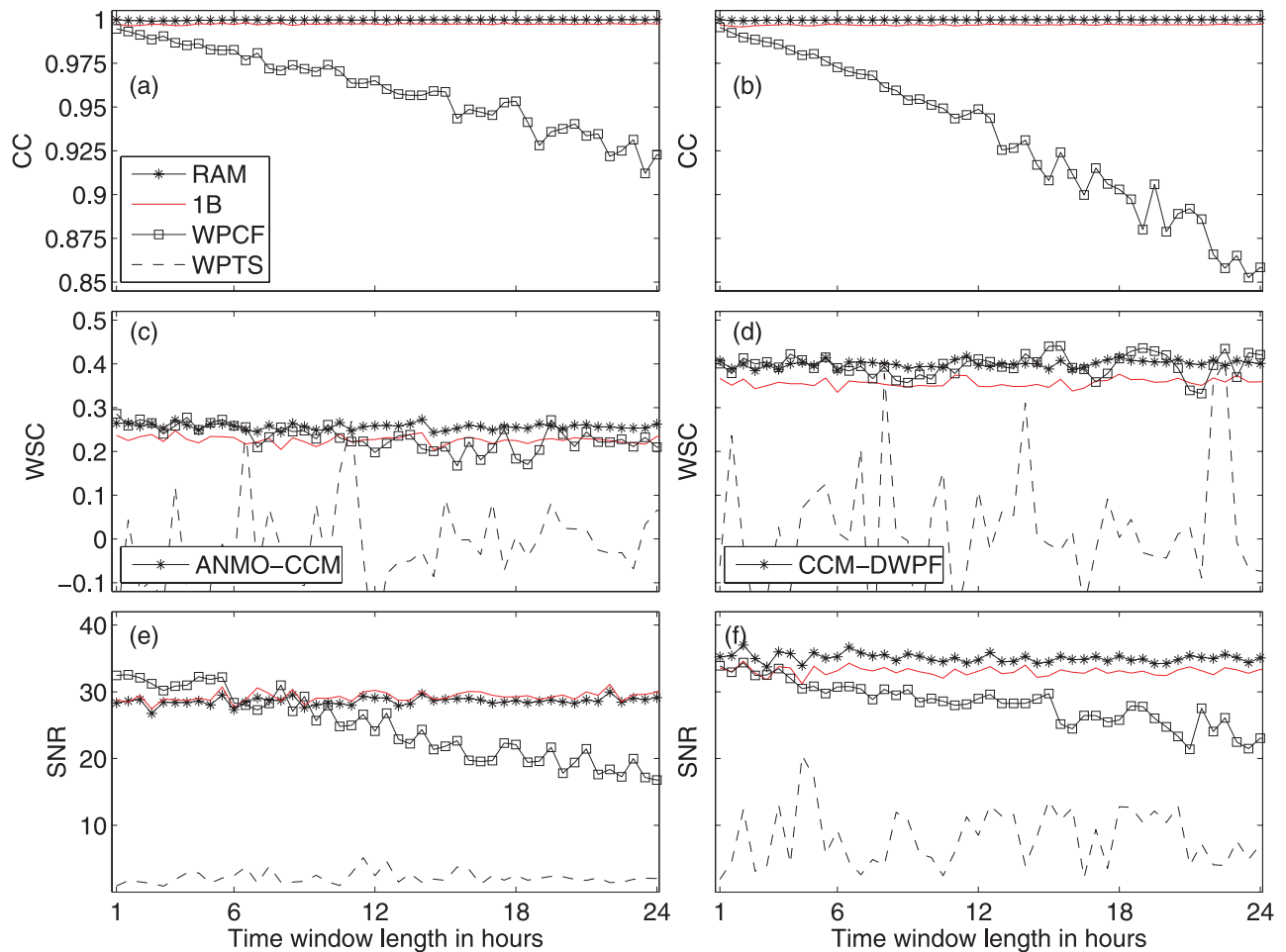
Figure 4. Detailed comparison of broad-band stacked 12-month CCFs (7–150 s) obtained with different normalization schemes and time window lengths for the station pairs ANMO–CCM (left-hand side) and CCM–DWPF (right-hand side) based on the parameters CC (a,b), WSC (c,d) and SNR (e,f). All parameters are introduced in detail in Section 2.2. The parameter CC quantifies the waveform similarity of the symmetric-component signal time window with the reference CFF (RAM CCF, 24 hr time window). The parameter WSC quantifies the waveform symmetry of the stacked CCFs and parameter SNR is the signal-to-noise ratio of the symmetric-component stacked CCFs. Only time windows without data are excluded from the stacking. The strong scattering of the in general very low WSC (c,d) and SNR (e,f) values of the WPTS CCFs (dashed lines) illustrates the failure of the WPTS normalization to produce clear signals in the CCFs. The CC values of the WPTS CCFs are below the range of values (0.85–1) shown in (a) and (b).

3.1 Normalization of the time-series before cross-correlation

The waveform preserving normalization of the two time-series (WPTS-normalization) before the cross-correlation is a simple linear normalization with an amplitude value. The time-series are divided by the value of their 68 per cent amplitude interval range which is used for the time domain classification (Groos & Ritter 2009) and therefore also for the automated data selection (see Section 4). By this approach the long-term amplitude differences between the single-time window CCFs (e.g. seasonal variations due to ocean-generated microseism) are balanced. Also the influence of CCFs from time windows with very large amplitudes is reduced. Nevertheless, single-time window CCFs originating from time windows with comparably very large transient signals at both sites, such as earthquake waves, can dominate the stacked CCF.

3.2 Normalization of the CCFs after cross-correlation

The waveform preserving normalization of the single CCFs (WPCF-normalization) after the cross-correlation and before the

stacking is similar to the method used by Campillo & Paul (2003) who normalize the amplitudes of each CCF with their absolute maximum. We extend this normalization method with an additional procedure to weight the single-time window CCFs prior to the stacking. We use the rms value of the entire waveform to normalize a CCF. To introduce a weighting scheme to the CCF normalization we assume that single-time window CCFs with large transient signals, especially outside the defined signal time window, are most likely dominated by strong transients such as teleseismic earthquakes or waveform irregularities (e.g. calibration pulses). With this assumption we normalize CCFs with an overall SNR (absolute maximum value divided by the rms value of the CCF) larger than 13 and/or a symmetric-component SNR smaller than two with their maximum instead of their rms value. By doing so, a weighting is introduced by reducing the contribution of CCFs to the stack, which are dominated by transient signals. We selected the upper SNR boundary of 13 based on our observations with the analysed station pairs. We did not find single-time window CCFs with a SNR larger than 13, which were not originating from time-series with obvious transient signals. The lower boundary of two for the symmetric-component SNR is chosen to identify CCFs with larger amplitudes near the

lag time zero than within the signal time window. The SNR of a symmetric-component CCF without any emerging signal is statistically expected between three and four with the assumption of a normal distribution.

3.3 Performance of the time domain normalizations

Both waveform-preserving normalizations are influenced by the time window length as they require the stacking of single-time window CCFs. The best results can be expected with a time window length in the same order of magnitude as the duration of the dominating disturbing signals (e.g. teleseismic surface waves). The performance of both normalizations in respect to the time window length is therefore discussed in detail later. In contrast, the performance of the 1B and the RAM normalization can be expected to be independent of the time window length as these normalizations are applied directly to the time-series.

In Fig. 3 stacked 12-month CCFs (7–150 s) are shown obtained with an overlapping sliding time window (2 hr) of the station pair PFO–HRV without any normalization (none, Fig. 3a), with the one-bit (1B, Fig. 3b), the RAM (Fig. 3c), the waveform preserving time domain normalization of the single-time window CCFs (WPCF, Fig. 3d) as well as waveform preserving time domain normalization of the time-series (WPTS, Fig. 3e). No SW is applied to the shown CCFs. Only time windows containing no data (data gaps) are excluded.

Without normalization no seismic signals emerge in the signal time windows of the stacked CCF (Fig. 3a). The largest amplitudes are observed near lag time zero within the 'noise time window'. The SNRs of the causal and acausal parts are therefore very small (<2). This is the typical effect of disturbing transient signals in the seismic noise with comparably large amplitudes such as earthquake waves or instrumental irregularities. The stacked CCF is furthermore strongly asymmetric as represented by the negative WSC value. The 1B and the RAM time domain normalization produce both clear signals (SNR ~ 50) in the acausal signal time window (Figs 3b and c). The signal in the causal part of the stacked CCFs is rather weak and the CCFs are not symmetric (WSC < 0.1) due to an unfavourable distribution of the noise sources for this station pair. The WPCF normalization produces nearly identical signals in the acausal and causal parts of the stacked CCF (Fig. 3d). The very simple WPTS normalization applied prior to the cross-correlation (WPTS, Fig. 3e) is not capable to produce clear signals in the signal time windows of the stacked CCF.

No obvious differences of the broad-band (7–150 s) stacked 12-month CCFs obtained with the 1B, the RAM and the WPCF normalization can be observed from the waveforms in Fig. 3 without SW. The most relevant differences at very long periods (>70 s) are not visible due to the very small relative amplitudes of the long-period signal content (Fig. 10e). The differences at long-periods are discussed in detail together with the SW in Section 5.

The performance of the time domain normalizations with different time window lengths is shown exemplary for the station pairs ANMO–CCM and CCM–DWPF in Fig. 4. The performance of the different time domain normalizations in comparison to each other is the same for all analysed station pairs. The comparison between the signal time windows of the symmetric-component stacked CCFs and the reference CCF (RAM normalization, 24 hr time window) is shown for station pair ANMO–CCM in Fig. 4(a) and station pair CCM–DWPF in Fig. 4(b). The stacked CCFs obtained with the RAM and the 1B normalizations are not changing with the time

window length. Nevertheless, small waveform differences occur between the CCFs obtained with these normalizations. The differences between the stacked CCFs obtained with the WPCF normalization and the reference CCF (RAM normalization, 24 hr time window) decrease systematically with decreasing time window length. This effect can be explained by the better suppression of earthquake signals and instrumental irregularities with the shorter time window lengths as these signals have a duration of less than 1 hr in general. The CC values of the stacked CCFs obtained with the failing WPTS normalization are outside the shown range of values (CC 0.85–1). The failing of the WPTS normalization is also illustrated by the strong scattering of the in general very low WSC (Fig. 4c,d) and SNR (Fig. 4e,f) values of the WPTS CCFs.

The waveform symmetry of the stacked CCFs obtained with the 1B and the RAM normalization is independent of the time window length (Fig. 4c,d). In general, the stacked CCFs obtained with the RAM normalization exhibit slightly higher WSC values than the stacked CCFs obtained with the 1B normalization at the analysed station pairs. The waveform symmetry of the stacked CCFs obtained with the WPCF normalization increases with decreasing time window length. In general it is comparable with the symmetry of the RAM CCFs for time window length shorter than 6 hr (Fig. 4c,d).

The SNR of the WPCF CCFs is increasing with decreasing time window length due to the better suppression of short transient signals (Fig. 4e,f). The SNR of the WPCF CCFs is also comparable with the SNR of the 1B and the RAM CCFs for time window lengths shorter than 6 hr. For most station pairs the SNR of the RAM CCFs is observed to be slightly higher than the SNR of the 1B CCFs as found by Bensen *et al.* (2007). Nevertheless, there are exceptions (pair ANMO–CCM, Fig. 4e).

Concluding, the waveform preserving normalization of the single-time window CCFs is capable to replace non-linear time domain normalizations if the time window length is comparable to the length of the disturbing signals (e.g. earthquake waves). Nevertheless, no significant improvement of the SNR or the waveform symmetry of the stacked 12-month CCFs is observed for the analysed station pairs in general. The advantage of the WPCF normalization is the fact, that the seismic noise time-series are not distorted prior to the cross-correlation.

4 AUTOMATED DATA SELECTION

Our proposed automated data selection considers the characteristics of the two broad-band time-series which are cross-correlated and is based on the noise classification introduced by Groos & Ritter (2009) which was briefly discussed in Section 2. The data selection is realised by the exclusion of single-time window CCFs from the stacking. We exclude a single-time window CCF if at least one of the two cross-correlated time-series is classified as corrupt (no data, extreme transients, step in the time-series) or if both time-series are dominated by transient signals (like contemporaneously arriving surface waves). Technically speaking, a single-time window CCF is excluded from the stacking if at least one of both cross-correlated time-series is classified as noise class 10 or larger (corrupt) or if both time-series are classified as noise class 3 or larger (transient signal(s); see Groos & Ritter (2009) for details).

In Fig. 5 the stacked 12-month CCFs (PFO–HRV; 7–150 s) obtained with the automated data selection are shown in comparison to the same stacked CCFs obtained without data selection (Fig. 3). Clear signals emerge now in the acausal signal time windows of
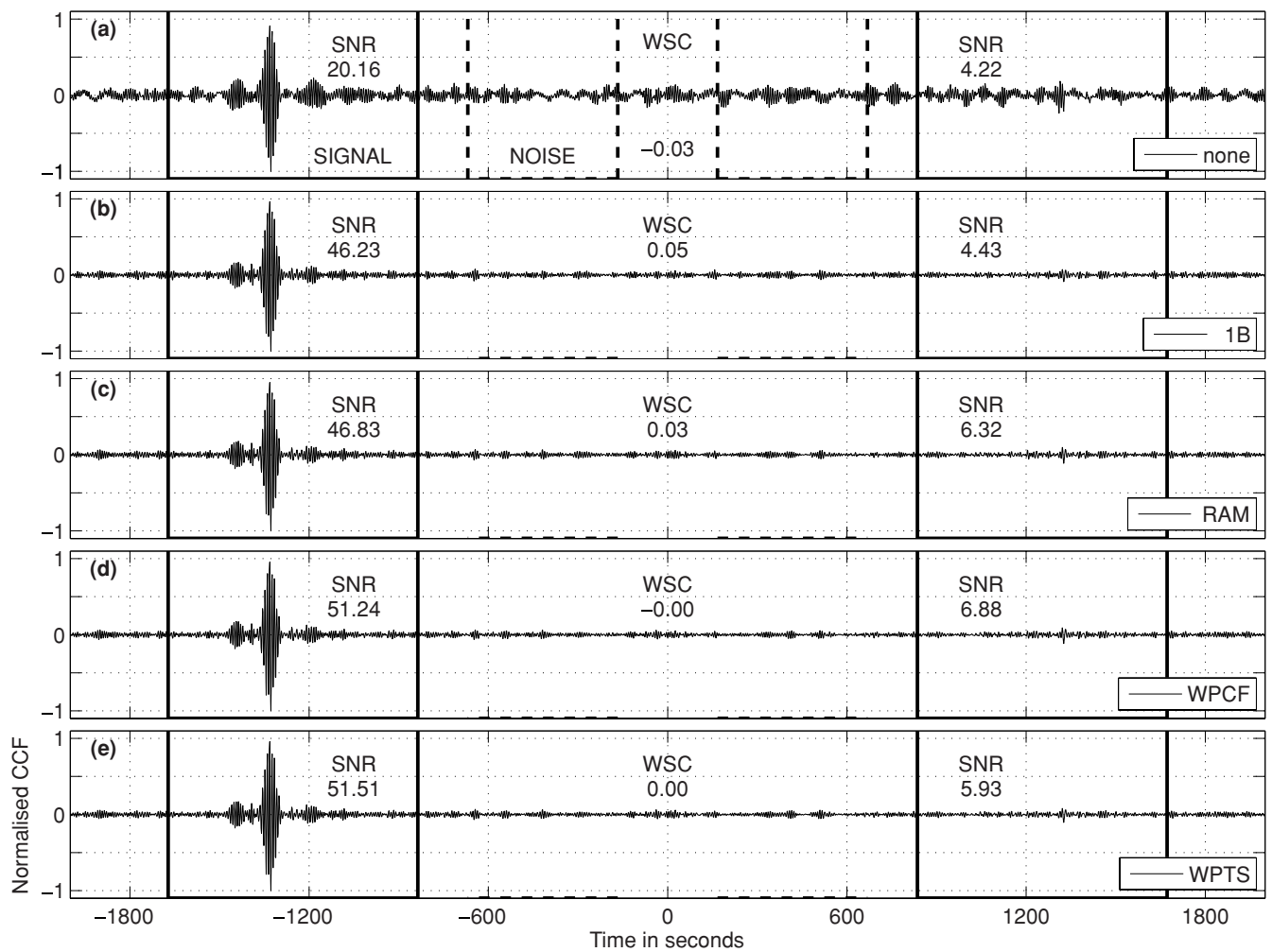
Figure 5. Stacked 12-month CCFs (PFO–HRV, 7–150 s, 2 hr time window) after data selection obtained with (a) none, (b) one-bit, (c) running absolute mean and, (d) WPCF (see Section 3) and (e) WPTS normalization. No spectral whitening is applied prior to the stacking. Time windows containing no data, corrupt data or with transient signals observed contemporaneous at both sites (see data selection in Section 4) are excluded from the stacking.

all stacked CCFs, even if no normalization is applied (Fig. 5a). This demonstrates that the removal of disturbing signals such as earthquake waves or instrumental irregularities by our automated data selection approach improves the obtained stacked CCFs. The stacked CCFs obtained with the different time domain normalizations are nearly identical with comparable SN ratios and waveform symmetry for all analysed time window lengths (Fig. 6c–f, station pairs ANMO–CCM and CCM–DWPF). The similarity between the obtained stacked CCFs and the reference CCF (RAM, 24 hr time window, no data selection) increases with decreasing time window length (Fig. 6a,b). The increase of the SNR, waveform symmetry and similarity with the reference CCF is caused by the increasing efficiency of the data selection approach with decreasing time window length. This is related to the better concurrence of a short time window length (<4 hr) with the length of the occurring transient signals (e.g. teleseismic surface waves). With a time window length considerably longer than the transient signals a significant amount of data is lost due to the exclusion of the unnecessary long time windows containing a strong but short transient signal. This is illustrated by the fact that in general all time windows longer than 48 hr would be excluded due to the continuous global seismicity. The total amount of used data increases from ∼20 per cent (24 hr time window) to approx. 65 per cent with a time window length of 1 hr.

The proposed fully automated exclusion of transient signals is operational and effective by replacing a non-linear time domain normalization. Nevertheless, no improvement in comparison to the stacked 12-month CCFs obtained without the exclusion of transient signals but an effective normalization of transient signals is observed. We conclude that an appropriate normalization of contemporaneous transient signals instead of excluding them from the cross-correlation processing should be preferred. Nevertheless, we implemented the noise classification in the cross-correlation processing to reliably exclude time windows containing data gaps or corrupt data automatically. Furthermore, time windows with problematic data can be easily identified for a further detailed analysis.

5 SPECTRAL WHITENING

SW of the time-series prior to the cross-correlation or the single-time window CCFs prior to the stacking is a massive intervention into the signal averaging. The intention is to improve the finally obtained stacked CCFs in terms of representing broad-band Green's function estimates (Bensen *et al.* 2007). The approach of SW assumes that the wanted signal is present in the data but buried by uncorrelated noise. Only in this case the SNR of the wanted signal is
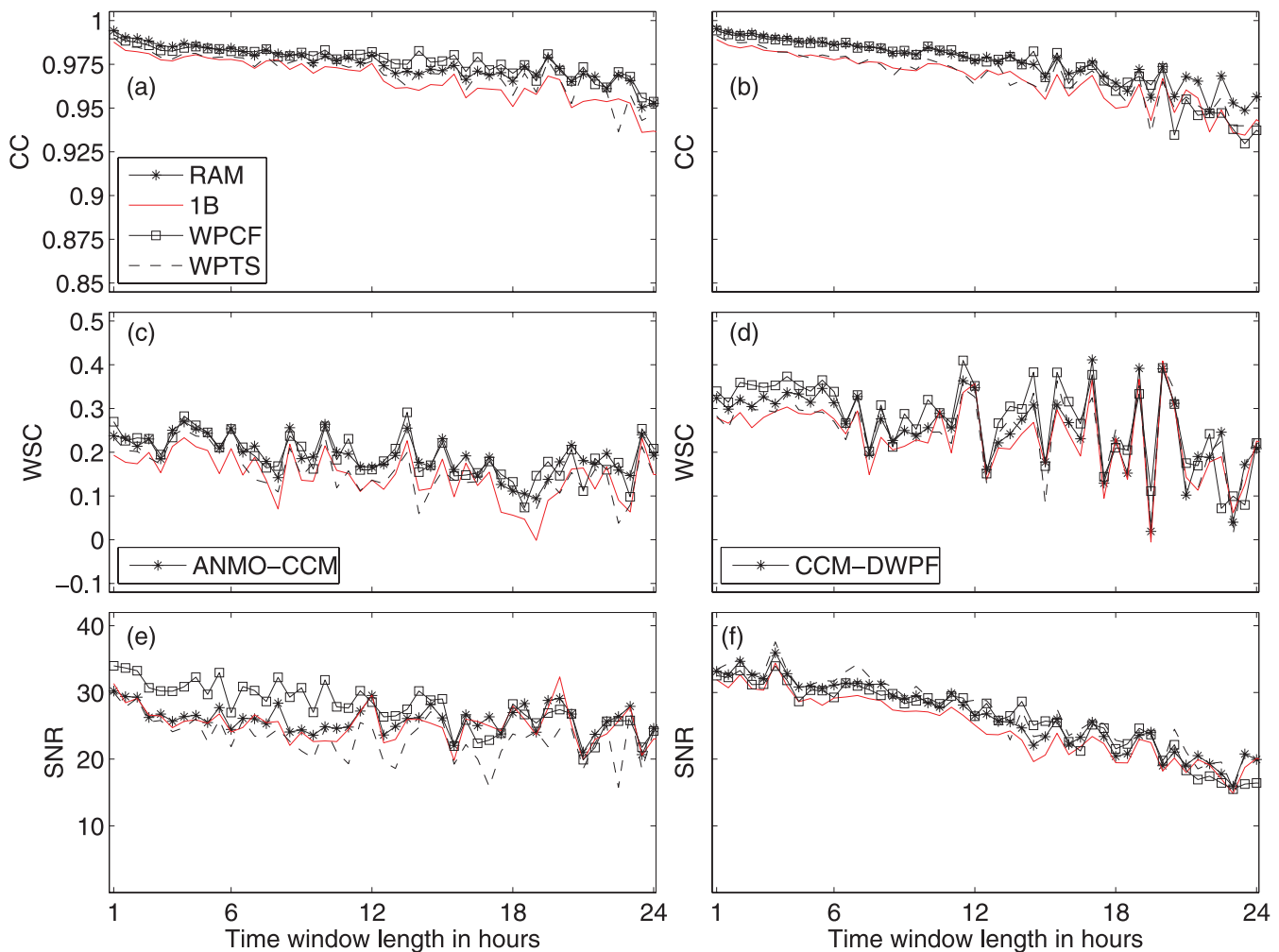
Figure 6. Detailed comparison of broad-band stacked 12-month CCFs (7–150 s) after data selection obtained with different normalization schemes and time window lengths for the station pairs ANMO–CCM (left-hand side) and CCM–DWPF (right-hand side). (a,b) CC, (c,d) WSC and (e,f) SNR. Time windows containing no data or with transient signals observed contemporaneous at both sites (see Section 4) were excluded from the stacking. Compare with Fig. 4.

increased by the equalization of the spectral amplitudes. In this section, we discuss the immense influence of the time window length on the stacked CCFs obtained with SW. A variety of different implementations of the SW into the cross-correlation processing can be found in published studies (Groos 2010). SW is applied to the time-series or CCFs, alone or in combination with a time domain normalization as well as before or after the time domain normalization. Also the implementation of the SW itself varies. We apply the SW itself as proposed by Brenguier *et al.* (2008). The SW is done by normalizing the complex spectrum of the CCF to an absolute value of one in the period range of interest (7–150 s) and to zero outside this period range. The SW is applied to the single-time window CCFs prior to the stacking and in general after the time domain normalization. A detailed analysis of the different SW strategies typically applied in cross-correlation processing can be found in Groos (2010). Here we focus on the evaluation of the SW alone and after the application of the 1B and the RAM normalization with different time window lengths.

The stacked 12-month CCFs (7–150 s, PFO-HRV, 24 hr time window) obtained with SW alone and in combination with the 1B and the RAM normalization are shown in Fig. 7. The differences between the stacked CCFs obtained with SW alone (Fig. 7a) and in combination with the RAM normalization (Fig. 7b) are rather

small. In both cases the long-period (>70 s) content of the signals is significantly amplified in comparison to the stacked CCFs obtained without SW (Fig. 3). The symmetry of the stacked CCF is significantly improved by the SW due to a more uniform and suitable source distribution at longer periods. An improvement of the SNR due to the RAM time domain normalization is observed but rather small. Significant differences are observed between the 1B-SW CCF (Fig. 7c) and the SW and RAM-SW CCFs. The 1B-SW CCF is significantly less symmetric (WSC 0.18 in comparison to 0.23 and 0.26). This effect is caused mainly by the differences in the long-period (>70 s) content of the CCFs, which can be observed directly from the waveforms especially at the beginning of the signal time window (around –1050 s in Fig. 7c). The long-period signals with comparably low spectral amplitudes in the original time-series are not adequately represented in the one-bit normalised time series. This kind of 'filter effect' was also observed by Pedersen *et al.* (2007). One solution proposed by Pedersen *et al.* (2007) is the narrow-band pre-filtering prior to the cross-correlation and to cross-correlate the seismic noise time-series separately in all frequency bands of interest. However, it is common practice for dispersion analyses to calculate broad-band CCFs and to apply narrow-band filters on the CCFs afterwards (e.g. Bensen *et al.* 2007; Pedersen *et al.* 2007). The intention is to reduce the computational costs
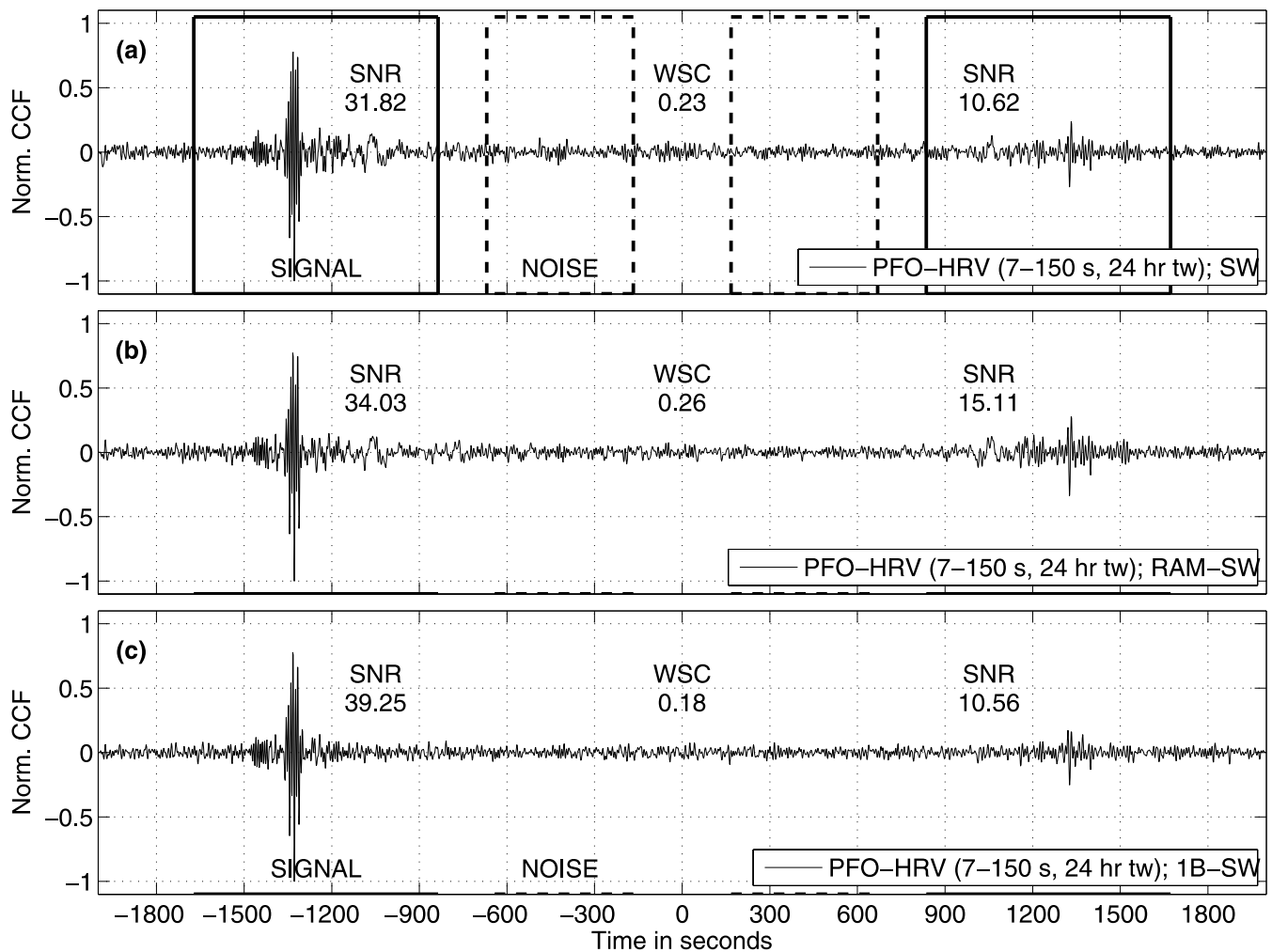
Figure 7. Three stacked 12-month CCFs (7–150 s) for the station pair PFO–HRV obtained with (a) none, (b) RAM and (c) one-bit normalization and a time window length of 24 hr. Spectral whitening (SW) is applied to the single-time window CCFs after the cross-correlation prior to the stacking. Significant is the lack of the long-period signal content in the causal and acausal signal time windows of the 1B-SW CCF (c) around lag time –1100 s.

and therefore processing time. In our opinion the 1B normalization should not be used for the cross-correlation of broad-band seismic noise time-series in general.

The evaluation of the stacked 12-month CCFs (7–150 s period) obtained with SW over time window length is shown in comparison to Fig. 4 in Fig. 8 (station pairs ANMO–CCM and CCM–DWPF). The direct comparison of the symmetric-component signal time windows (parameter CC, Fig. 8a,b) with the reference CCF (RAM-SW, 24 hr time window) reveals again the systematic waveform deviations of the 1B-SW CCFs in comparison to the SW and the RAM-SW CCFs due to the lack of the long-period signal content. Consequently, the waveform symmetry of the 1B-SW CCFs is lower than for the SW and the RAM-SW CCFs (Fig. 8c,d). The SNR of the 1B-SW CCFs is in general higher (Fig. 8e,f). Nevertheless, this indicates not a higher quality of the obtained CCFs as the high SNR is caused solely by the large amplitudes of the short-period (<20 s) content of the signal.

Next to this general differences between the stacked CCFs obtained with different normalization schemes we observe significant variations of the stacked CCFs with the time window length. For all station pairs the similarity of the SW and the RAM-SW CCFs increases systematically with decreasing time window length (indirectly seen by the identical CC values in Fig. 8a,b). The stacked

CCFs signal time windows are nearly identical for time window lengths shorter than 6 hr. This implies that the suppression of short transient signals by the SW increases with decreasing time window length. Here, no advantage can be taken from the time domain normalization if a time window length shorter than 6 hr is used.

Furthermore and more problematic, the deviations of the SW and RAM-SW CCFs in comparison to the reference CCF (RAM-SW, 24 hr time window) increase systematically with decreasing time window length (Fig. 8a,b). The waveform symmetry (Fig. 8c,d) and the SNR (Fig. 8e,f) are decreasing concurrently to the increasing waveform differences (Fig. 8a,b) with time window length. The obtained stacked CCFs are getting less suitable to estimate Green's functions. This effect is hardly observed with the 1B-SW CCFs. In the following we demonstrate that the cause of this effect is the amplification of a monochromatic persistent microseism signal by the SW. The magnitude of the amplification increases with decreasing time window length.

5.1 Amplification of persistent monochromatic signals by spectral whitening

The variations of the SW and the RAM-SW CCFs with decreasing time window lengths are a frequency dependent effect. Fig. 9 shows
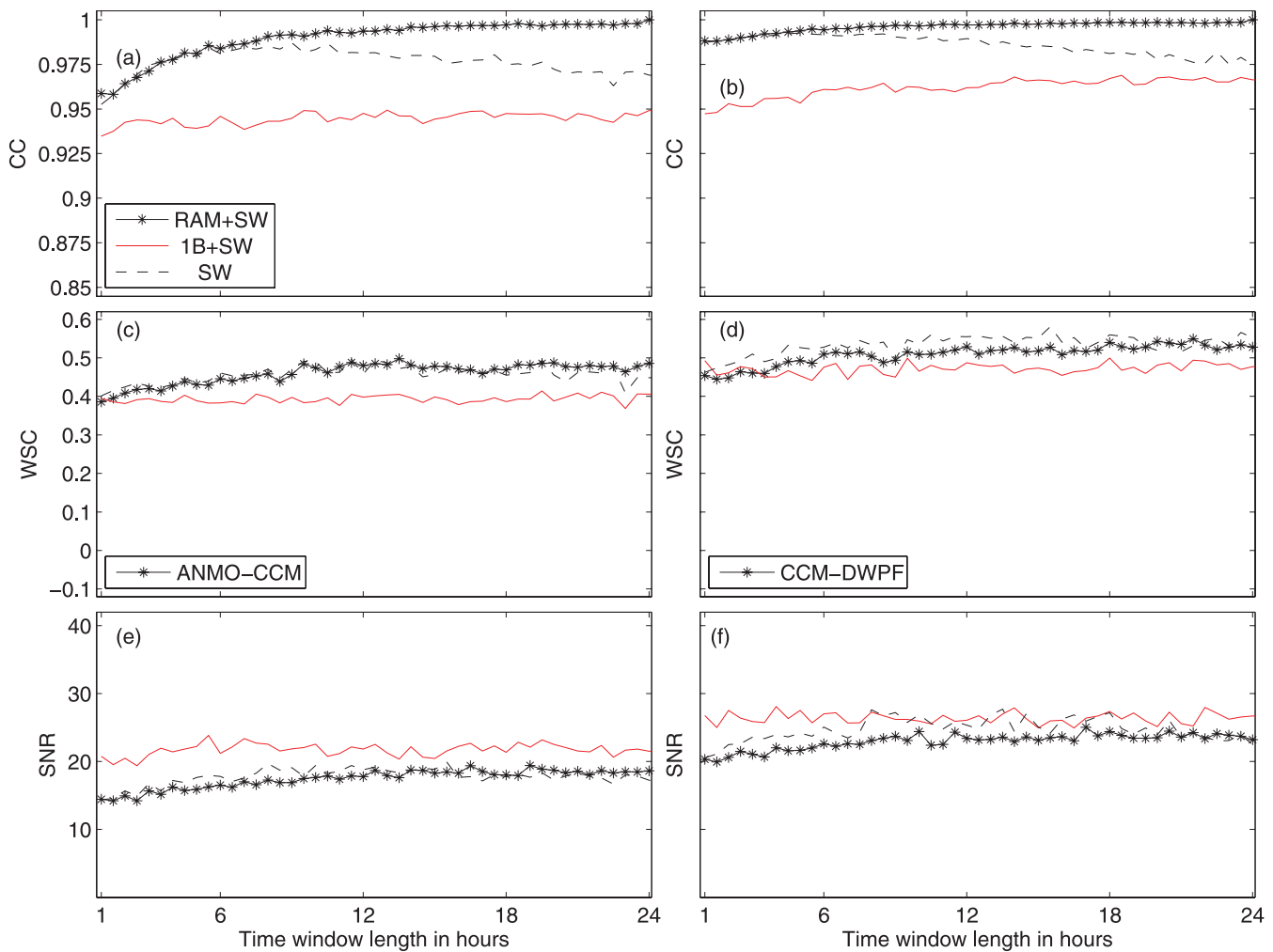
Figure 8. Detailed comparison of broad-band stacked 12-month CCFs (7–150 s) obtained with different normalization schemes and time window lengths for the station pairs ANMO–CCM (left-hand side) and CCM–DWPF (right-hand side) by the parameters CC (a,b), WSC (c,d) and SNR (e,f). Only time windows containing no data are excluded from the stacking. Compare with Fig. 4.

the CC values for the short period (7–14 s; asterisk), intermediate period (20–50 s; circle), long period (150–70 s; dots) and broadband (7–150 s, solid line) RAM-SW CCFs. Increasing waveform differences with decreasing time window length are not evident at short and long periods but pronounced in the intermediate period range 20–50 s. This effect is caused by a persistent narrow-band microseism signal at 26 s most probably originating from the Gulf of Guinea, West Africa (Shapiro *et al.* 2006), which is amplified by the SW as discussed below.

In Fig. 10(a) the 12-month RAM CCFs for station pair ANMO-CCM are shown which were obtained with a time window length of 24 hr (black solid line) as well as 2 hr (red dashed line). The corresponding RAM-SW CCFs are shown in Fig. 10(b). The waveform differences between the 2 hr-CCFs and the 24 hr-CCFs are given in Fig. 10(c) and Fig. 10(d), respectively. The amplitude spectra of the CCFs are shown in Figs 10(e) and (f).

No significant differences between the 2 hr-CCF and the 24 hr-CCF are observed in the waveforms or the amplitude spectra of the RAM CCFs (left side of Fig. 10). With SW a strong monochromatic signal with a period of approx. 26.33 s (0.038 Hz) emerges in the causal part of the 2 hr-RAM-SW CCF (right side of Fig. 10). A signal at the same period is also present in the other CCFs but with

significantly smaller relative amplitudes [compare the amplitude spectra in Figs 10(e) and (f)]. As aspired, the SW produces a more broad-band stacked CCF to estimate the Green's function. However, undesired is the amplification of the monochromatic signal not related to the Green's function.

The observed 26 s signal is in fact well known to emerge in continental scale CCFs especially in North America, Europe and Africa (Shapiro *et al.* 2006; Bensen *et al.* 2007). Shapiro *et al.* (2006) identified a monochromatic microseism signal at ~26 s which is propagating with ~3.5 km s^{-1} and seems to be excited by an unknown source located in the Gulf of Guinea. They observe this microseism signal to be very persistent in time with a seasonal amplitude variation with larger amplitudes during the northern hemisphere summer months. In Fig. 11 the differences between the 24 hr and the 2 hr RAM-SW CCFs (such as in Fig. 10d) in the period range 20–50 s are shown for the six station pairs indicated by the lines in Fig. 1). The difference CCFs are plotted dependent on the effective station distance in relation to the assumed origin of the microseism signal in the Gulf of Guinea. The red dashed line in Fig. 11 indicates the move-out of a signal travelling with a speed of 3.5 km s^{-1}. The causal part of the CCFs corresponds to a westward propagation of the signal. Concluding, the 26 s signal—amplified in the CCFs
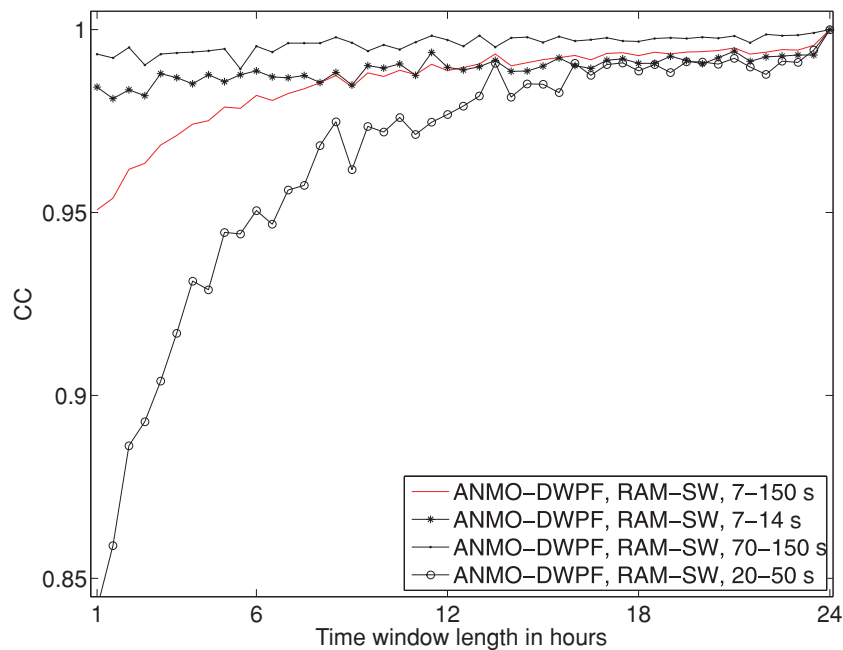
Figure 9. Comparison of RAM-SW CCFs obtained with different time window lengths with the reference CCF (RAM-SW, time window length 24 hr) for different period ranges (7–150 s, 7–14 s, 20–50 s and 70–150 s). Only time windows containing no data are excluded from the stacking. The effect of increasing deviations of the stacked CCFs from the reference CCF with decreasing time window length is observed only for the stacked CCFs in the period range 20–50 s due to a monochromatic persistent microseism signal around 26 s which is amplified by the spectral whitening.

obtained with SW and time window lengths shorter than 12 hours—shows the same behaviour as the persistent 26 s microseism signal observed by Shapiro *et al*. (2006).

The amplification of the originally very small signal in the single-time window CCFs is caused by the temporal and local stability of the source of the 26 s microseism signal in combination with the stacking. The amplitude information is discarded by the SW and only the phase information of the signal is retained. This leads to an amplification of temporally stable persistent signals even with small relative amplitudes if CCFs are stacked after SW. Although the amplitude spectra of the single-time window CCFs are flat after SW, the amplitude spectrum of the stacked 12-month CCF is not. The amplitude spectrum of the stacked CCF is related to the coherence of the single-time window CCFs and is shaped by the stacking process.

The temporally very persistent 26 s signal is enhanced by the higher summation order of the stacking with a shorter time window length. This is only possible because the signal emerges reliably in the single-time window CCFs, even if a short time window length of only 2 hr is used (Groos 2010). As the 26 s signal is excited continuously by a localized source, a short time window is sufficient to obtain a CCF containing the corresponding signal. In contrast, the emergence of signals representing the Green's function can be expected to be less effective in the same short time window due to the insufficient signal averaging. If two long time-series are cross-correlated, the signal averaging, which is necessary for the emergence of the Green's function, is obtained by the cross-correlation. By dividing the long time-series into shorter time windows, the signal averaging is increasingly transferred from the cross-correlation to the stacking process and therefore increasingly affected by the SW. This effect amplifies especially temporally persistent signals in the CCFs like the 26 s microseism which efficiently produce signals

in the single-time window CCFs (Groos 2010). Therefore, the time window length should be carefully selected.

Concluding, the time window length should be as long as possible to enhance the emergence of the Green's function in the single-time window CCFs independent of the stacking. Nevertheless, a stacking of CCFs after SW is recommendable. The stacking ensures the cancellation of incoherent noise in the single-time window CCFs, which is also amplified by the SW. The amplification of monochromatic persistent signals in CCFs can be a problem also for the calculation of seismic noise CCFs on a local scale in the frequency range between 1 and 60 Hz. Especially in urbanized environments numerous localized sources of persistent monochromatic signals exist (Groos & Ritter 2009). The far most of these monochromatic sources are industrial machines at all scales (e.g. ventilation, pumps, cement mills, . . .) driven by electrical motors.

5.2 Combination of spectral whitening with the waveform preserving time domain normalizations

A combination of SW with the proposed waveform preserving time domain normalizations is possible if a stacking process with two stages is applied. First, several single-time window CCFs obtained with a short time window length (e.g. 2 hr) are normalized in the time domain (WPCF or WPTS) and stacked to obtain CCFs for a longer time window (e.g. 1–2 d). These stacked CCFs are then whitened in the frequency domain and stacked again to obtain the complete (e.g. 12 months) stacked CCF.

For the analysed data set no systematic improvement of the CCFs is observed by the combination of SW with a time domain normalization in comparison to the SW CCFs. Nevertheless, this may be different with other data sets and data sets on other scales. The advantage of the WPCF normalization and the SW of the CCFs
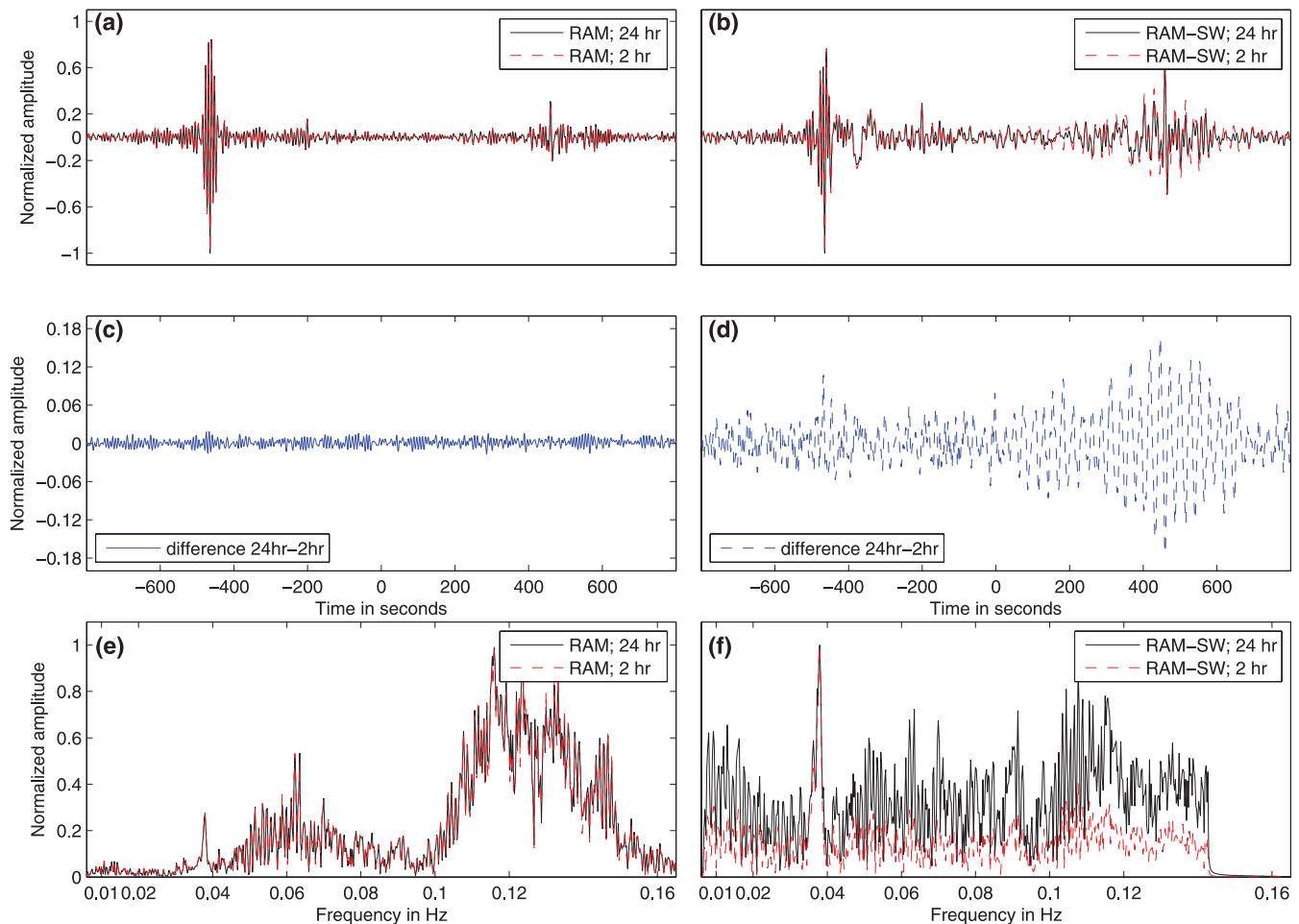
Figure 10. Comparison of stacked 12-month CCFs (ANMO–CCM, 7–150 s) obtained with different normalization schemes and time window lengths. (a) RAM–CCFs obtained with a time window length of 24 hr (black line) and 2 hr (red dashed line). (b) the same as in (a) for the RAM–SW–CCFs. (c) and (d) waveform differences between the 2-hr CCF and the 24 hr-CCF in (a) and (b), respectively. (e) and (f) corresponding amplitude spectra of the CCFs in (a) and (b), respectively.

instead of the time-series is the high flexibility. The seismic noise time-series are cross-correlated without any changes to their waveforms or amplitudes. The normalization(s) can be applied afterwards and the effect of different normalization schemes can be easily evaluated.

6 CONCLUSIONS

We analysed several topics of seismic noise cross-correlation processing. We compared established normalizations such as the RAM and the one bit normalization in the time domain as well as the SW normalization in the frequency domain. We tested approaches of waveform preserving time domain normalization before and after the cross-correlation. All normalization schemes were evaluated with different time window lengths between 1 hr and 24 hr. In addition to the normalizations we evaluated a new approach of fully automated data selection to improve the CCFs.

As already observed by other authors we also observe that the one-bit normalization is not suitable to calculate broad-band CCFs due to their negative influence on the frequency content of the broad-band seismic noise time-series. A narrow-band pre-filtering would be necessary to reduce this problem (Pedersen *et al.* 2007).

The RAM time domain normalization proposed by Bensen *et al.* (2007) is a highly effective non-linear time domain normalization to suppress transient signals. We use this normalization as reference time domain normalization as it is independent of the time window length. Our proposed waveform-preserving normalization of the CCFs prior to the stacking (WPCF normalization) is capable to replace a non-linear time domain normalization, if a suitable time window length is used. The time window length should be similar to the length of the typically occurring coherent transient signals (e.g. teleseismic earthquakes).

As addition to the normalization of the data we tested a fully automated data selection approach. The approach is based on a time-series classification method and excludes time windows containing contemporaneous transient signals at both sites. This approach effectively removes such time windows from the data and it is capable to replace a non-linear time domain normalization if a suitable time window length comparable to the length of the transient signals is used. Nevertheless, we observe no improvement of the obtained CCFs in comparison to the normalized CCFs obtained without data selection. In our opinion a suitable normalization of transient signals is preferable in contrast to a strict exclusion.

We observe the SW normalization to be significantly influenced by the time window length. The application of SW on very short time
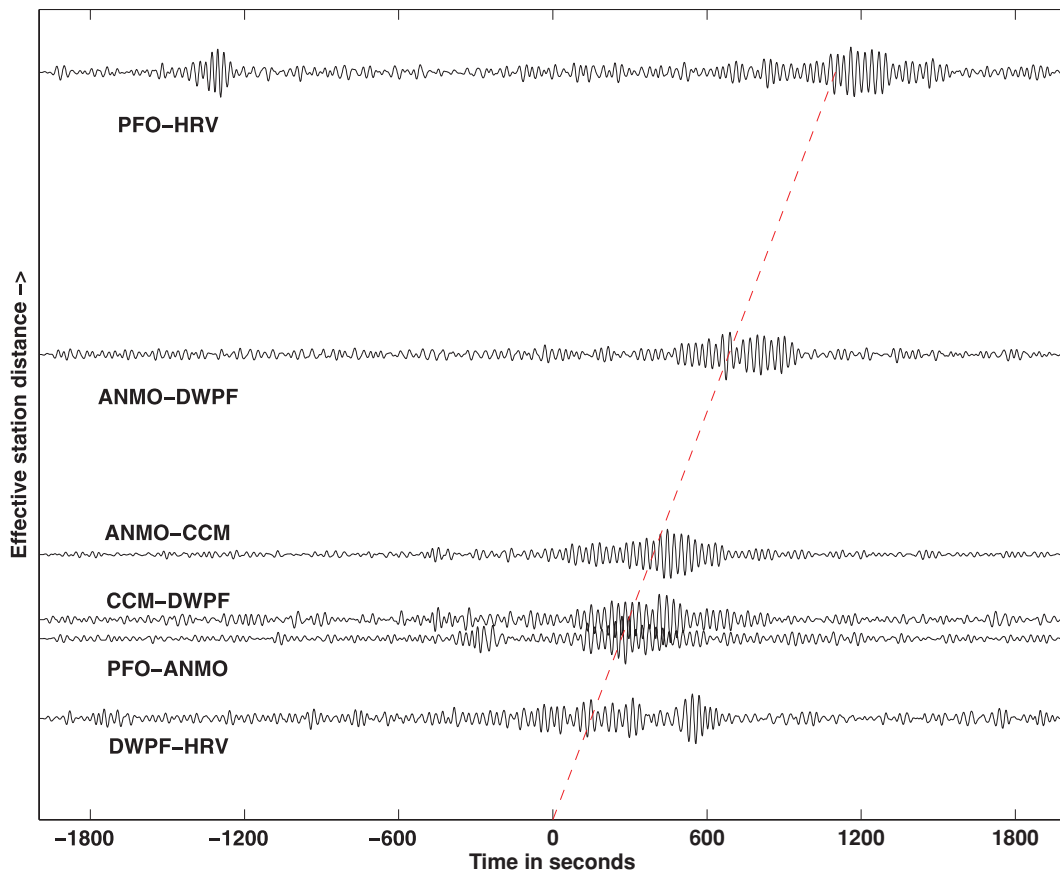
Figure 11. Illustration of the waveform differences between the 24 hr RAM-SW CCFs and the 2 hr RAM-SW CCFs in the period range 20–50 s. The waveform differences are plotted distant-dependent for the effective distances of the station pairs in relation to the source location of the 26 s microseism signal in the Gulf of Guinea. The red dashed line indicates the theoretical move-out of a wave travelling with 3.5 km s^{-1} as the 26 s microseism (Shapiro *et al.* 2006).

windows of a few hours undesirably amplifies temporally persistent coherent signals like the 26 s microseism signal most probably exited in the Gulf of Guinea. The time window length used with SW should be as long as possible to assure the emergence of signals of the Green's function in the single-time window CCFs. Nevertheless, a stacking of the single-time window CCFs after the SW is important to reduce the incoherent noise introduced to the CCFs by the SW. The SW normalization is capable to suppress also the influence of transient signals, if a suitable time window length is used which is comparable to the length of the transient signals. Unfortunately, such a short time window length leads to the discussed amplification of persistent signals.

Concluding, we recommend SW as the main normalization for the calculation of CCFs to estimate broad-band Green's functions on a continental scale. The SW is necessary to equalize the large differences in spectral amplitudes of the signals contributing to seismic noise in the analysed period range 7–150 s. The time window length used for the processing should be long enough (at least 24 hr on a continental scale) to avoid the unwanted amplification of monochromatic persistent signals by the SW. If large transient signals such as earthquake waves lead to biased CCFs despite the SW an additional time domain normalization should be used. In such a case we recommend to combine the proposed waveform preserving time domain normalization of the single-time window CCFs with SW by a stacking process with two stages. The waveform preserving time domain normalization without SW will be sufficient if the

spatially coherent part of the analysed seismic noise is characterized by a flat amplitude spectrum. This may be the case especially for applications on a local scale.

REFERENCES

Bendat, J.S. & Piersol, A.G., 1994. *Random Data: Analysis and Measurement Procedures,* 2nd edn, John Wiley & Sons, New York, NY.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.,* **169,** 1239–1260.

Brenguier, F., Shapiro, N.M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O. & Nercessian, A., 2008. Towards forecasting volcanic eruptions using seismic noise, *Nat. Geosci.,* **1,** 126–130.

Bussat, S. & Kugler, S., 2011. Offshore ambient-noise surface-wave tomography above 0.1 Hz and its applications, *Leading Edge,* **30,** 514–524.

Campillo, M., 2006. Phase and correlation in random seismic fields and the reconstruction of the Green Function, *Pure appl. Geophys.,* **163,** 475–502.

Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science,* **299,** 547–549.

Curtis, A., Gerstoft, P., Sato, H., Snieder, R. & Wapenaar, K., 2006. Seismic interferometry – turning noise into signal, *Leading Edge,* **25,** 1082–1092.

Gouédard, P., Cornou, C. & Roux, P., 2008. Phase-velocity dispersion curves and small-scale geophysics using noise correlation slantstack technique, *Geophys. J. Int.,* **172,** 971–981.

Groos, J.C. & Ritter, J.R.R., 2009. Time domain classification and quantification of seismic noise in an urban environment, *Geophys. J. Int.,* **179,** 1213–1231.

Groos, J., 2010. Broadband seismic noise: classification and Green's function estimation, *PhD thesis,* Karlsruhe Institute of Technology, 155 pp.

Pedersen, H.A., Krüger, F. & the SVEKALAPKO Seismic Tomography Working Group, 2007. Influence of the seismic noise characteristics on noise correlations in the Baltic shield, *Geophys. J. Int.,* **168,** 197–210.

Li, H., Bernardi, F. & Michelini, A., 2010. Surface wave dispersion measurements from ambient seismic noise analysis in Italy, *Geophys. J. Int.,* **180,** 1242–1252.

Roux, P., 2009. Passive seismic imaging with directive ambient noise: application to surface waves and the San Andreas Fault in Parkfield, CA, *Geophys. J. Int.,* **179,** 367–373.

Sabra, K.G., Gerstoft, P., Roux, P. & Kuperman, W.A., 2005. Extracting time-domain Green's function estimates from ambient seismic noise, *Geophys. Res. Lett.,* **32,** L03310, doi:10.1029/2004GL021862.

Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science,* **307,** 1615–1618.

Shapiro, N.M., Ritzwoller, M.H. & Bensen, G.D., 2006. Source location of the 26 sec microseism from cross-correlations of ambient seismic noise, *Geophys. Res. Lett.,* **33,** L18310, doi:10.1029/2006GL027010.

Weaver, R.L., 2005. Information from seismic noise, *Science,* **307,** 1568–1569.

Yao, H. & van der Hilst, R.D., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet, *Geophys. J. Int.,* **179,** 1113–1132.

Yang, Y. & Ritzwoller, M.H., 2008. Characteristics of ambient seismic noise as a source for surface wave tomography, *Geochem. Geophys. Geosyst.,* **9,** Q02008, doi:10.1029/2007GC001814.