

Broadband Seismic Noise: Classification and Green's Function Estimation

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Diplom-Geophysiker Jörn Christoffer Groos

aus

Mannheim

Tag der mündlichen Prüfung:

3. Dezember 2010

Referent:

PD Dr. Joachim Ritter

Korreferent:

Prof. Dr. Friedemann Wenzel

'Noise is that part of the data that we choose not to explain.'

Scales & Snieder (1998)

Abstract

Several efforts are undertaken in seismology to turn ambient seismic noise into signal. A key motivation for this approach is to overcome obstacles hampering established methods of active source and passive earthquake seismology. Techniques based on seismic noise are independent from earthquake activity or active seismic sources. This is a significant advantage in areas where the natural seismicity is low and/or where active seismic sources (explosions, large vibrators) can't be used. Passive seismic noise measurements can be conducted also in sensitive areas such as city centres and nature reserves due to their low environmental impact. Seismic noise is a low-cost and easy-to-measure signal which is available everywhere and at every time.

At the same time the social and economical importance of seismic hazard assessment and mitigation in (mega)cities is rapidly increasing due to the exploding urbanisation, especially close to major fault systems. Site effect analyses, wave propagation scenarios and early warning concepts are high-priority issues for such urban regions. Therefore, the number of passive seismic measurements in urban environments is permanently increasing to provide the required information about the underground by utilising seismic noise. Urban seismic noise is evolving to one of the most important signals of modern seismology.

This thesis aims at a better determination and understanding of the spatial and temporal variations of the amplitudes as well as the statistical properties of the (urban) seismic noise wave field. A good knowledge of these spatial and temporal variations of the seismic noise is crucial to identify noise sources on the one hand and to be able to consider the actual noise conditions by the utilisation of seismic noise on the other hand.

A new statistical time series classification is presented which is capable to distinguish between corrupt and non-corrupt time series as well as to classify non-corrupt time series in six meaningful noise classes. The time series classification is used to conduct a comprehensive analysis of the spatial and temporal variations of the seismic noise between 8 mHz and 45 Hz in the metropolitan area of Bucharest, Romania. This analysis improves the understanding of the statistical properties of the urban seismic noise due to temporally and spatially varying noise sources. The combination of the time series classification with an unsupervised neural network technique, the Self-Organizing Map method, is demonstrated to be a promising approach to enhance the analysis of complex urban seismic noise data sets.

The time series classification is furthermore used to realise a data selection approach for the estimation of Green's functions from seismic noise cross-correlation functions. The implementation of this data selection approach involves a comprehensive evaluation of the common seismic noise cross-correlation processing. Based on this evaluation a more flexible processing scheme is realised and critical parameters of the processing such as the time window length are identified. Furthermore, a wave form preserving time domain normalisation and a second data selection approach are presented and evaluated in this context to improve the calculation of seismic noise cross-correlation functions.

Concluding, an effective time series classification for seismic noise time series is proposed in this thesis. It is demonstrated that the time series classification can be used to obtain new insights into the temporal and spatial variations of (urban) seismic noise. The time series classification provides furthermore valuable data selection capabilities for all methods utilising seismic noise.

Contents

Abstract	i
Contents	iii
1 Introduction.....	1
2 Seismic noise	5
2.1 'Seismic noise', an ambiguous term.....	5
2.2 The seismic noise wave field	6
2.2.1 The sources of the seismic noise wave field	6
2.2.2 The composition of the seismic noise wave field	7
2.3 Utilisation of seismic noise.....	8
2.3.1 Spectral H/V ratio	8
2.3.2 Array methods	9
2.3.3 Seismic interferometry.....	10
2.3.4 Other applications using seismic noise	14
3 Data.....	17
3.1 URS Bucharest.....	17
3.2 GSN data.....	19
3.3 Inconsistencies in SEED instrument response metadata	20
3.3.1 The SEED format	20
3.3.2 The seismological measuring system	21
3.3.3 The transfer function: Description of the seismometer	23
3.3.4 The digital low-pass decimation filter	25
3.3.5 The inconsistencies	27
3.3.6 Some remarks about <i>evalresp</i>	30
3.3.7 Occurrence of the inconsistencies	30
3.3.8 Summary	32
4 Time series classification	33
4.1 Time series properties used for quantification and classification	33
4.2 Preparation of the time series prior to the classification	35
4.3 Observed deviations from the Gaussian distribution	35
4.4 Noise classification scheme.....	37
4.4.1 Corrupt time series	37
4.4.2 Non-corrupt time series	37
4.5 Classification of synthetic data.....	39
4.6 Summary of chapter 4	41
5 Analysis of urban seismic noise.....	43
5.1 Time-Frequency analysis.....	44
5.1.1 Urban seismic noise above 1 Hz ('microtremor').....	45
5.1.2 Urban seismic noise 0.6-1 Hz (natural and man-made sources).....	45

5.1.3	Urban seismic noise below 0.6 Hz ('microseism')	47
5.2	Analysed frequency bands and time windows	48
5.3	Analysis of selected working days	50
5.3.1	Microtremor (1-45 Hz)	51
5.3.2	Transitional range (0.6-1 Hz)	52
5.3.3	Microseism (0.04-0.6 Hz)	54
5.3.4	Frequency range 0.008-0.04 Hz	58
5.4	Analysis of the complete URS data set	60
5.4.1	General discussion of the seismic noise amplitudes	60
5.4.2	Comparison of all / selected working days (low wind velocities)	60
5.4.3	Comparison of working days / Sundays (low wind velocities)	62
5.4.4	Comparison of working days with high / low wind velocities	64
5.4.5	Comparison of the vertical and horizontal components	65
5.4.6	Time-domain H/V ratio	67
5.5	Analysis with Self-Organizing Maps (SOMs)	70
5.5.1	The Self-Organizing Map	70
5.5.2	Application to the URS data set	71
5.5.3	Results of the SOM analysis	72
5.6	Summary of chapter 5	78
6	Improved calculation of seismic noise cross-correlation functions	81
6.1	Data processing	82
6.1.1	Time series preprocessing	82
6.1.2	Data processing scheme	82
6.1.3	The digital cross-correlation	86
6.1.4	Data parameters	86
6.1.5	Time window length	89
6.1.6	Further processing of the CCFs	90
6.2	Spectral whitening	91
6.2.1	Strategies	91
6.2.2	Amplification of persistent monochromatic signals	93
6.2.3	Discussion	95
6.3	Data selection	96
6.4	Waveform preserving time domain normalisation	98
6.4.1	Normalisation of the time series before cross-correlation	99
6.4.2	Normalisation of the CCFs after cross-correlation	99
6.4.3	Discussion	100
6.5	Evaluation of the different normalisation schemes	101
6.5.1	Observations	102
6.5.2	Discussion	104
6.6	Evaluation of the data selection approaches	105

6.6.1 Time series approach (TSA).....	106
6.6.2 Waveform symmetry approach (WSA).....	107
6.7 Summary of chapter 6	111
7 Summary and Conclusions	113
Appendix A: Karlsruhe Seismology Processing Toolbox	117
Appendix B: Metadata of seismic stations	119
Appendix C: Noise classification of the URS data set	123
List of Figures	137
List of Tables	139
List of Abbreviations	141
References	143
Used hard- and software.....	151
Danksagung / Acknowledgment.....	153
Curriculum Vitae	155

1 Introduction

Several efforts are undertaken in seismology to turn ambient seismic noise into signal. Since the 1950s several techniques based on array measurements as well as single station measurements of seismic noise are developed and applied to derive information about the subsurface. In the last decade the estimation of Green's functions of the underground based on seismic noise cross-correlation functions, called *seismic interferometry*, evolved to an important and widely used technique in seismology (Weaver, 2005; Curtis et al., 2006). Seismic interferometry enables seismology nowadays to provide high-resolution tomography studies from local to continental scale from seismic noise.

A key motivation to turn seismic noise into signal is to overcome obstacles hampering established methods of active source and passive earthquake seismology. Techniques based on seismic noise are independent from earthquake activity or active seismic sources. This is a significant advantage in areas where the natural seismicity is low and/or where active seismic sources (explosions, large vibrators) can't be used. Passive seismic noise measurements can be conducted also in sensitive areas such as city centres and nature reserves due to their low environmental impact. The comparably low financial effort necessary for the passive seismic noise measurements in comparison to active source seismology is a further advantage. Concluding, seismic noise is a low-cost and easy-to-measure signal which is available everywhere and at every time.

At the same time the social and economical importance of seismic hazard assessment and mitigation in (mega)cities is rapidly increasing due to the exploding urbanisation, especially close to major fault systems (United Nations, 2006; Montgomery, 2008). Site effect analyses, wave propagation scenarios and early warning concepts are high-priority issues for such urban regions which require reliable information about the underground. Therefore, the number of passive seismic measurements in urban environments is permanently increasing to provide the required information about the underground by utilising seismic noise (e.g. Milana et al., 1996; Scherbaum et al., 2003; Ritter et al., 2005; Fäh et al., 2008). Concluding, urban seismic noise is evolving to one of the most important signals of modern seismology.

Seismic noise in cities is driven by numerous processes such as our cultural life, individual and public commuter traffic and the production as well as transportation of all kinds of goods. Time-dependent and reoccurring seismic signals are generated and emitted by the man-made physical processes. These emissions, or man-made ground motions, are superimposed with natural ground motions (wind-, ocean wave- or earthquake-induced tremor) to make up the temporal and spatial highly variable urban seismic noise (Groos & Ritter, 2009). The high variability of the seismic noise wave field is a fact which must be addressed by all methods utilising seismic noise. Bonnefoy-Claudet et al. (2006a) criticise that the authors of most studies, utilising seismic noise, assume that the underlying assumptions about the seismic noise wave field are fulfilled without further discussion. They state also, that the majority (~95%) of publications about seismic noise deals with the utilisation of seismic noise and not with the nature of the seismic noise wave field. Concluding, seismological research must significantly improve the understanding of seismic noise to successfully and reliably utilise seismic noise in general and especially in urban environments (Bonnefoy-Claudet et al., 2006a; Campillo, 2006).

This thesis aims at a better determination and understanding of the spatial and temporal variations of the amplitudes as well as the statistical properties of the (urban) seismic noise wave field. A good knowledge of these spatial and temporal variations of the seismic noise is crucial to identify noise sources on the one hand and to be able to consider the actual noise conditions by the utilisation of seismic noise on the other hand. This thesis suggests a statistical time series classification for the determination of these

seismic noise conditions. The proposed time series classification is used for a comprehensive analysis of the statistical properties of the seismic noise between 8 mHz and 45 Hz in the metropolitan area of Bucharest, Romania, to improve the understanding of urban seismic noise. Furthermore, the time series classification is used to realise and evaluate a data selection approach based on the noise conditions for seismic interferometry.

Chapter 2 provides a review of the fundamentals of and the current knowledge about seismic noise. The ambiguous definition of the term 'seismic noise' is discussed prior to the state of knowledge about the sources and the composition of the seismic noise wave field. The most important methods utilising seismic noise including single station and array techniques are discussed in the third section of chapter 2 with a focus on seismic interferometry.

The two different data sets of ground motion recordings which are used for this thesis are introduced in chapter 3. The data set of the URban Seismology (URS) project (Ritter et al., 2005) was measured with 31 stations of the mobile KARlsruhe Broad Band Array (KABBA) in the metropolitan area of Bucharest between October 2003 and August 2004. This data set is used for a comprehensive analysis of the urban seismic noise in chapter 5. The second data set consists of one year (2004) time series of several Global Seismographic Network (GSN) stations in the USA and is used to evaluate and improve the data processing for seismic interferometry in chapter 6. While working with the GSN data set several inconsistencies in the instrument response metadata distributed together with the waveform data became apparent to the author. It turned out, that the metadata of dozens of seismic stations world-wide were (and partly still are) affected by these inconsistencies. The observed inconsistencies in the SEED metadata and their practical relevance are discussed in the third section of chapter 3.

Chapter 4 introduces the time series classification which is used to determine the spatial and temporal variations of the amplitudes as well as the statistical properties of the (urban) seismic noise. The classification combines amplitude information with few distinct noise classes (Gaussian distribution, presence of large transient or periodic signals) characterising the seismic noise. The time series classification is also capable to identify common technical artefacts occurring during (mobile) passive seismic measurements (e.g. data gaps, direct mechanical impacts to the sensor). The classification allows next to the analysis of the seismic noise conditions also an automated data selection from large data sets for a consecutive seismic noise processing by identifying corrupt and/or inappropriate (e.g. dominating periodic or transient signals) time windows of seismic noise. The classification is based on time series percentiles and their ratios. The time series properties used for the classification are introduced in the first section of chapter 4. In the second section the classification scheme based on the time series properties is discussed prior to the evaluation of the classification with a synthetic data set in the third section of chapter 4. The last section summarises chapter 4.

Chapter 5 provides the detailed discussion of the broad-band Urban Seismic Noise (USN) in the metropolitan area of Bucharest. The data set of the URS project is a rare opportunity to analyse the urban seismic noise continuously in a broad frequency range (0.008-45 Hz) and over a long time (9 months). The temporal variability and the typical sources of the USN in Bucharest are discussed in the first section with the help of a time-frequency analysis. Purpose of the time-frequency analysis is the identification of suitable frequency bands and time windows for the consecutive analysis with the time series classification. The frequency bands and time windows as well as the selection of 11 working days for the further analysis are discussed in the second section of chapter 5. The temporal and spatial variability of the vertical-component urban seismic noise in Bucharest is discussed in detail with the help of the selected 11 working days in the third section. The purpose of the selection is to analyse predominantly the USN caused by

sources inside the metropolitan area with a focus on man-made sources. Therefore working days are selected which are not affected by strong natural sources outside the city area such as earthquakes and ocean-generated microseism. The analysis of the complete URS data set with the time series classification including the horizontal component USN is discussed afterwards in the fourth section. The analysis of such a large and complex noise classification data set by an analyst involves a considerable effort. The next following step is therefore the analysis of an obtained seismic noise classification data set with machine learning and pattern recognition techniques to support the human analyst. A feasibility study with the Self-Organizing Map (SOM) technique, which is a neural network technique, is presented for the URS data set in the fifth section of this chapter. The chapter is concluded with a short summary in the last section.

Several methods of automated data selection and normalisation to improve the calculation of seismic noise cross-correlation functions (CCFs) for seismic interferometry are proposed and evaluated in chapter 6. A detailed overview about the applied data processing and the specific aspects of the data processing addressed by this thesis is given in the first section of chapter 6. The significant influence of the time window length on the important frequency domain normalisation is discussed in the second section of chapter 6 prior to the introduction of the two fully automated data selection approaches in the third section. The first approach is based on the time series classification introduced in chapter 4. The second approach uses properties of cross-correlation functions for data selection. A new waveform preserving normalisation to improve the cross-correlation functions is proposed in the fourth section. The considered normalisation schemes are evaluated without data selection in the fifth section and with data selection in the sixth section. The chapter is completed by a short summary of the most important conclusions in the last section.

The seventh and last chapter summarises the main results and conclusions of this thesis.

2 Seismic noise

The purpose of this chapter is to give a review of the fundamentals and the current knowledge about seismic noise. The first problem one encounters with seismic noise is the term 'seismic noise' itself. Therefore, the definition of seismic noise is discussed in the first section of this chapter. The second section gives a summary of the state of knowledge about the sources and the composition of the seismic noise wave field. The chapter is completed by a third section with a brief review of the most important methods utilising seismic noise. A main focus is laid on a technique called 'seismic interferometry' which is directly addressed by this thesis later on.

2.1 'Seismic noise', an ambiguous term

The term 'noise' is widely and naturally used in science. This is especially the case for scientists working with measured data such as seismologists. Consequently, 'seismic noise' is a common term in seismology. Even though the term 'noise' is commonly used, its meaning is quite ambiguous. This led Scales & Snieder (1998) to ask the interesting question: '*What is noise?*' Commonly, noise is regarded to be somehow 'random' and is supposed to bias the 'deterministic signal'. This is also a common point of view in seismology looking at earthquakes or seismic waves excited by active sources as deterministic signals. Nevertheless, as stated by Scales & Snieder (1998), it is hardly possible to give a precise (mathematical) definition of randomness which satisfies everyone. Furthermore, many signals appear random to us only because we do not know all underlying (deterministic) processes. It is obvious that there are conceptual problems with the term 'noise', especially if we remind the old dictum '*One man's noise is another man's signal*'. Scales & Snieder (1998) consider the transcription '*noise is that part of the data that we choose not to explain*' as the best way to describe what most scientists really mean, if they talk about 'noise'. Therefore, the task of this section is to clarify what is and was meant by the term 'seismic noise' and how this term is used in this thesis.

The beginning of seismology was the interest to understand earthquakes. Quite early, seismologists were able to use the observed earthquakes to learn about the inner structure of the Earth (e.g. Rebeur-Paschwitz, 1889). The seismic waves excited by an earthquake are the traditional signals of seismology. The exploration of the inner structure of the Earth using earthquake waves is still a key task and competence of seismology. It is therefore not surprising, that the permanent movements of the Earth's surface which could not be related to earthquakes were identified as 'seismic noise'. The first comprehensive treatise of this subject was published by Beno Gutenberg (1911) in his dissertation at the University of Göttingen. It is interesting to mention, that his dissertation written in German is entitled with '*Die seismische Bodenunruhe*' which is in fact not well translated by '*seismic noise*'. A better translation would be something like '*The seismic restlessness of the ground*'. Like the original German title this translation carries no implications about the character or origin of the permanent movements causing the 'restlessness' of the ground. In contrast, the term 'seismic noise' implies randomness and unpredictability. In fact, the main general sources of seismic waves next to earthquakes such as industry, traffic, ocean-generated microseism and meteorological causes are known to seismologists for a long time and far from the beginning (Bertelli, 1872; Gutenberg, 1911; Gutenberg, 1924). The unpredictability of the permanent movements of the Earth is caused mainly by the fact that we do not know all individual sources and that our knowledge of the Earth's structure is also not sufficient. The term 'seismic noise' evolved basically as an abbreviation of the transcription 'the permanent movement of the Earth's surface with exception of earthquakes'.

Nowadays, the usage of the term 'seismic noise' is getting more and more complicated. Several sources of seismic waves such as the ocean-generated microseism are well understood and often used as signals themselves (e.g. Grevemeyer, 2000;

Zhang et al., 2010). Furthermore, several techniques to utilise 'seismic noise' to learn about the structure of the underground are developed since the 1950s (see section 2.3). All of these techniques make extensive assumptions about the character and sometimes the sources of that part of the seismic noise wave field that is actually used. Therefore, the meaning of the term 'seismic noise' lies nowadays in the eye of the beholder. The sole commonality is that earthquake waves are in general not considered to be noise. Nevertheless, the exclusion of one well-known type of source such as earthquakes from the definition is in fact purely arbitrary and provokes misunderstandings as earthquake waves are not the exclusively used signals in seismology anymore. Therefore, the term 'seismic noise' is used in this thesis simply as an abbreviation to describe the 'permanent movements of the Earth's surface'.

2.2 The seismic noise wave field

A comprehensive and complete literature review of the current knowledge about the seismic noise wave field is given by Bonnefoy-Claudet et al. (2006a). They state also, that the majority (~95%) of publications about seismic noise deals with the utilisation of seismic noise and not with the nature of the seismic noise wave field. This is still the case today. The most significant progress since 2006 is achieved concerning the utilisation of seismic noise by estimating the Green's function of the Earth with seismic noise cross-correlations (see also chapter 6 of this thesis). This new and very fast evolving method is introduced with other applications in section 2.3 about the utilisation of seismic noise. Here a brief summary is given about our current knowledge about the seismic noise wave field.

2.2.1 The sources of the seismic noise wave field

In general, the seismic noise wave field is a superposition of a large amount of (deterministic) seismic signals excited by numerous natural and man-made physical processes. The seismic noise wave field altogether is unpredictable for us as we don't know all individual sources and the structure of the underground. Natural sources of seismic signals in general are tides, water-waves striking the coast, standing water waves in the open seas due to storm systems, air pressure changes, turbulent wind or wind-induced vibrations of trees or tall buildings. Just to name a few and the most important. The man-made sources are also numerous such as car and train traffic, industrial machines, explosions or the exploitation of underground reservoirs (e.g. hydrocarbons, hot water). The most important natural and man-made sources were already identified by Gutenberg at the beginning of the 20th century (Gutenberg, 1911; Gutenberg, 1924). The typically occurring sources of the seismic noise wave field are discussed in detail in chapter 5 for the Bucharest metropolitan area in Romania.

As a rule of thumb, the seismic noise wave field is dominated by natural sources at low frequencies (<0.5 Hz), by man-made sources at high frequencies (>5 Hz) and by both in between (Bonnefoy-Claudet et al., 2006a). An exact separation of both parts of the wave field by a 'border frequency' is not possible and strongly site dependent as it is significantly influenced by the dominant noise sources and the local geological conditions. It is known that man-made sources can dominate the seismic noise wave field down to 0.5-1 Hz in a setting with soft subsoil such as the deep sedimentary basin of the Bucharest area (Groos & Ritter, 2009). The dominance of man-made signals ends around 1 Hz or at even higher frequencies in a hard rock setting (Bonnefoy-Claudet et al., 2006a).

Following these observations about the sources of seismic noise, two terms evolved in literature. The low frequency part of the seismic noise wave field dominated by natural sources is often called 'microseism'. The high frequency part with dominating man-made sources is called 'microtremor'. Unfortunately, the usage and separation of both terms in literature is not consequent which causes discrepancies (Bonnefoy-Claudet et al., 2006a). The term microseism is furthermore often used as an abbreviation of the term 'ocean-

generated microseism' which specifies only one of the numerous natural sources. Therefore, both terms are avoided as far as possible in this thesis.

2.2.2 The composition of the seismic noise wave field

The extensive review of Bonnefoy-Claudet et al. (2006a) demonstrates very impressive that a somehow 'traditional' assumption about the composition of the seismic noise wave field is not supported by the available data. The seismic noise wave field was and sometimes still is commonly assumed to consist mainly of fundamental mode Rayleigh waves. The assumption of dominating fundamental mode Rayleigh waves holds in far most cases for the double-frequency ocean-generated microseism in the period band 5-20 s (Brooks et al., 2009). The double-frequency ocean-generated microseism is a long known (Gutenberg, 1911) and relative well understood (Longuet-Higgins, 1950) source of seismic energy contributing to the seismic noise wave field. It seems so, that the observation of dominant fundamental mode Rayleigh waves for this dominant and important source was simply adopted for the remaining sources of the seismic noise wave field for a long time.

In fact, the presence of body waves (P, SV, SH) as well as surface waves (Rayleigh and Love waves, fundamental and higher modes) in the seismic noise wave field is observed for a long time. This is recalled by the careful literature review of Bonnefoy-Claudet et al. (2006a) and confirmed by new analyses. The presence of these wave types in the seismic noise wave field is not a matter of course. The seismic noise wave field observed in the direct vicinity of local sources cannot be split up in these wave types observed in the layered earth and in the far-field of a seismic source. The same is true for a wave field which consists of signals with wave lengths of the same order of magnitude as the spatial heterogeneities of the medium. The current state of knowledge about the seismic waves excited by distant sources typically present in the seismic noise wave field is summarised in the following.

Nowadays, the presence of body and surface waves in the seismic wave field can be observed in a broad frequency range due to the capability and large amount of modern 3-component seismometers and modern computers. Kurrle & Widmer-Schmidrig (2008) demonstrate the existence of Rayleigh and Love waves in the Earth's hum at very low frequencies (<7 mHz). Lacoss et al. (1969) already observed fundamental mode Rayleigh and Love waves to be the dominant components of the seismic wave field in the frequency band 0.04-0.15 Hz. In general, the ocean-generated microseism is observed to consist not only of narrow-band fundamental mode Rayleigh waves but also higher mode Rayleigh waves (Brooks et al., 2009; Koper et al., 2010), Love waves (Saito, 2010) as well as P, PP and PKP body waves (Gerstoft et al., 2009; Zhang et al., 2009; Koper et al., 2010) in a broad frequency band (0.05-2 Hz). Koper et al. (2010) observe the fundamental mode Rayleigh wave to be in fact the least significant component of the vertical wave field in the frequency range 0.4-4 Hz. The knowledge about the composition of the seismic noise wave field at higher frequencies larger than 1 Hz is improved by the increasing amount of three-component array measurements especially in urban areas and accompanying synthetic seismic wave field simulations (e.g. Bonnefoy-Claudet et al., 2006b; Bonnefoy-Claudet et al., 2008). The three-component array measurements for the frequency-wavenumber, or FK, technique (e.g. Fäh et al., 2008) as well as the SPatial AutoCorrelation (SPAC) technique (e.g. Köhler et al., 2007; Endrun et al., 2010) are actually conducted to provide S wave velocity profiles mainly for geotechnical engineering and seismic hazard assessment purposes (see also 2.3.2). The composition of the seismic noise wave field is determined as some kind of 'by-product' in these cases. The array measurements demonstrate us, that the proportion of the different wave types in the seismic noise wave field above 1 Hz is strongly site dependent. This affects the proportion between body and surface waves, Rayleigh and Love waves as well as fundamental and higher modes (Bonnefoy-Claudet et al., 2006a). The coexistence of fundamental and higher mode Rayleigh and Love waves is commonly observed at frequencies above 1 Hz

(e.g. Köhler et al., 2007; Fäh et al., 2008; Endrun et al., 2010). The proportion of Rayleigh waves in comparison to Love waves is observed to be highly variable between 10% and 90% (Bonney-Claudet et al., 2006a) but a dominance of Love waves is observed in far most cases (Bonney-Claudet et al., 2006a; Köhler et al., 2007; Endrun et al., 2010). Quantifications of the proportion between body and surface waves are rare in literature and allow up to now no clear conclusions (Bonney-Claudet et al., 2006a).

The wave field composition is observed to depend massively on the type and orientation of the sources (vertical, horizontal), the location of the sources (near, far; at the surface, below the surface) and the underground structure (hard, soft; strong, weak impedance contrasts). These observations are supported by wave field simulations (e.g. Bonney-Claudet et al., 2006b; Bonney-Claudet et al., 2008). Consequently, there is no general and simple assumption about the composition of the seismic noise wave field at the moment! And it seems hardly possible that a general assumption can be found. The composition of the seismic noise wave field is highly variable with space, time and frequency as most of the affecting boundary conditions and noise sources exhibit a distinct spatial and temporal variability.

The high variability of the seismic noise wave field is a fact which must be addressed by all methods utilising seismic noise described in the following section 2.3 and is therefore a key motivation of this thesis. A good knowledge of the utilised seismic noise wave field is necessary in every individual case. Bonney-Claudet et al. (2006a) criticise that the authors of most studies, utilising seismic noise, assume that the underlying assumptions about the seismic noise wave field are fulfilled without further discussion. This is especially the case if methods based on one station (e.g. H/V, section 2.3.1) or two stations (e.g. seismic interferometry, section 2.3.3) are applied. The missing discussions can be easily explained by the difficulties to determine the composition of the seismic noise wave field without suitable array measurements.

This is the crucial point where this thesis contributes to the current research effort. The automated time series classification introduced in chapter 4 improves the knowledge about the spatial and temporal variations of the amplitudes as well as the statistical properties of the seismic noise wave field with a seismic network only. In chapter 6 new normalisation and data selection approaches are introduced and evaluated to improve the applicability of seismic interferometry (section 2.3.3) under inappropriate noise conditions.

2.3 Utilisation of seismic noise

The most important methods and applications in seismology based on seismic noise are summarised in this section. Most methods are introduced very briefly and for completeness only, as they are not directly addressed by this thesis. The estimation of the Green's function by seismic noise cross-correlations, also called seismic interferometry, is introduced in more detail as the seismic noise cross-correlation processing is treated comprehensively and improved in chapter 6.

2.3.1 Spectral H/V ratio

The spectral H/V ratio is calculated as the ratio of the Fourier amplitude spectra of the horizontal component and the vertical component of the seismic wave field and is mainly used as a tool for the microzonation of urban areas for seismic hazard assessment (Bonney-Claudet et al., 2006a). Nogoshi & Igarashi (1971) introduced this measure and observed a correlation between the peak frequencies of the so called 'H/V spectra' and the resonance frequencies of sites in sedimentary settings. The technique was not widespread in the first decades as the original work of Nogoshi & Igarashi (1971) was published in Japanese. This changed dramatically with the promotion of the then called 'Nakamura's technique' by Nakamura (1989, 1996, 2000). Experiments (e.g. Fäh, 1997) and 1-D simulations (e.g. Lachet & Bard, 1994) showed that the H/V ratio technique was

capable to identify the resonance frequency of a soft sedimentary deposit separated by a sharp impedance contrast from the underlying stiffer basement. Since then the H/V-ratio technique evolved to a widespread standard technique in urban microzonation. As proposed by Nakamura (1989, 2000) the H/V spectrum is commonly used to estimate the fundamental frequency by the H/V peak frequency and the site amplification factor by the H/V peak amplitude for sites located on sedimentary deposits. The technique is commonly applied although the theoretical basis of the method is still unclear (Bonney-Claudet et al., 2006a; Bonney-Claudet et al., 2006b; van der Baan, 2010).

Nakamura (1989, 2000) assumes that the H/V spectrum is shaped by the resonance of vertically incident SH waves in the sediments which lead to increased amplitudes on the horizontal components. In this case the amplitude of the H/V peak should be directly related to the amplification factor of the site. Other authors (e.g. Lachet & Bard, 1994) assume that the H/V spectrum is shaped by the ellipticity of the fundamental mode Rayleigh waves. The peak in the H/V spectrum would be caused then by the vanishing of the vertical component of the Rayleigh waves and therefore no relation of the peak amplitude to the site amplification factor would be given. Based on wave field simulations Bonney-Claudet et al. (2008) argue that the origin of the H/V peak can be explained by the Rayleigh wave ellipticity, the Airy phase of Love waves and the S wave resonance. A combination of several of these causes may also occur. They assume that the H/V peak frequency can be used anyway to estimate the sediment resonance frequency independent of the actual cause. Nevertheless, the amplitude of the H/V peak is significantly influenced by the wave field composition, especially by the proportion of Love waves (Bonney-Claudet et al., 2008; van der Baan, 2010). Based on their simulations Bonney-Claudet et al. (2008) and van der Baan (2010) argue that the fundamental frequency can be reliably estimated from the H/V spectrum but that an in-depth knowledge of the seismic noise wave field composition is required for a reliable interpretation of the H/V peak amplitude to estimate the site amplification factor. Unfortunately, the knowledge about the wave field composition is not available for most H/V studies as the measurements are rarely accompanied by suitable array measurements due to the high personal and organisational effort.

The seismic noise recorded in the metropolitan area of Bucharest (see chapter 3.1) was analysed also with the H/V technique by Ziehm (2006). Her results regarding the resonance frequency of the sediments in the Bucharest area are compared at some points with the results obtained by the time series classification in chapter 5.

2.3.2 Array methods

In the late 1950s seismologists began to develop and adapt array techniques to derive 1D shear wave velocity profiles of the underground from seismic noise. The array measurements are used to obtain the dispersion curve of surface waves (Rayleigh or Love waves) as first step. The 1D shear wave velocity profile is derived then by inversion of the obtained dispersion curve as second step. This array approach utilising seismic noise and first proposed by Aki (1957) is nowadays widespread to estimate shear wave velocity profiles especially in urban areas (Bonney-Claudet et al., 2006a). The shear wave velocity profiles are crucial background information for seismic hazard assessment. The passive seismic noise array measurements can be conducted with low costs as well as low impact on structures and environment. The approach can be applied in general with all array techniques capable to provide at least Rayleigh wave dispersion curves. The most important and widespread techniques are the spatial autocorrelation technique (SPAC) introduced by Aki in 1957 and the frequency-wavenumber technique (FK) introduced by Capon et al. in 1967. The current standard is the deployment of three-component seismometers to be able to conduct a combined inversion of Rayleigh and Love wave dispersion curves. The dispersion curves are commonly obtained with advanced versions of the SPAC (e.g. Köhler et al., 2007) or FK (e.g. Fäh et al., 2008)

technique. Also the joint inversion of dispersion curves with H/V spectra is applied (Bonney-Claudet et al., 2006a).

The SPAC method was already introduced by Aki (1957) as a three-component technique capable to provide the dispersion curves of Rayleigh and Love waves from seismic noise without prior knowledge of the source locations or directions of wave propagation. A disadvantage of the original SPAC technique is the need of an array with circular geometry and a centre station. It is very often difficult to realise such a specific array layout in an urban area. A result of the consequent improvement of the SPAC technique in the last decades is the modified three-component SPAC method (3c-MSPAC) which is capable to provide Rayleigh and Love wave dispersion curves also with arbitrary array geometries (Köhler et al., 2007). The SPAC method is based on the assumption of a stochastic wave field stationary in space and time. The cross-correlation of recordings of this stationary stochastic wave field at different locations in space is considered by Aki (1957) as a spatial auto-correlation of the wave field. Tsai & Moschetti (2010) demonstrated recently, that the SPAC theory is equivalent to the theory of seismic interferometry (section 2.3.3).

The FK technique was introduced by Capon et al. (1967) to improve the capability to monitor nuclear weapon tests and small earthquakes with the Large Aperture Seismic Array (LASA). The FK method was used early by Lacoss et al. (1969) also for a detailed general analysis of the composition of the seismic noise wave field at the LASA array. The seismic noise wave field is assumed to be a homogeneous random field. In this case a spectral representation exists which consists of a superposition of propagating waves (Capon, 1969). The homogeneous random field can then be characterised by a frequency-wavenumber spectral density function which allows the determination of the direction and apparent velocity of propagating waves. The main advantage of the FK technique is the compatibility with arbitrary array geometries. A disadvantage is the necessary rotation of the horizontal components North and East in radial and transversal for all azimuths to be able to distinguish between Love and Rayleigh waves (e.g. Lacoss et al., 1969; Fäh et al., 2008). Many FK studies were therefore limited to the determination of Rayleigh wave dispersion curves with the vertical seismic noise wave field due to the simpler interpretation and the deployment of cheaper vertical one-component seismometers (Fäh et al., 2008). The application of the FK technique to the vertical seismic wave field in the Bucharest area is discussed in chapter 5. The FK technique was used in this case to reveal ocean-generated microseism from the Black Sea and the Mediterranean Sea consisting of fundamental mode Rayleigh waves propagating across the Bucharest area.

2.3.3 Seismic interferometry

The term seismic interferometry is referred by Wapenaar & Fokkema (2006) to the 'principle of generating new seismic responses by crosscorrelating seismic observations at different receiver locations'. First, the representation of the response of the Earth by the elastodynamic Green's function is introduced in this section. Afterwards a summary is given about the estimation of the Green's function between two points *a* and *b* (specified in space by the vectors *a* and *b*) on the Earth's surface from seismic noise cross-correlation functions. The section ends with an overview about the most important and forward-looking applications of seismic interferometry.

2.3.3.1 The Green's function

The elastodynamic Green's function is a special solution of the *equation of motion* for an anisotropic inhomogeneous linearly elastic medium (Aki & Richards, 2002). The equation of motion connects displacements of a particle in a continuum to forces acting within the continuum and on its surface. Therefore the equation of motion is fundamental for seismology as the displacements of particles at the Earth's surface can be measured by seismometers.

The *displacement* vector $\mathbf{u}(\mathbf{x}, t)$ denotes the vector distance of such a particle at time t from the position \mathbf{x} in Cartesian coordinates $x_1, x_2,$ and x_3 that the particle occupied at some reference time t_0 . Position \mathbf{x} is regarded as a position of equilibrium and \mathbf{x} and t_0 are assumed to equal zero in the following.

The distortion of the medium is described for infinitesimal small deformations with relative length scales in the order of 10^{-6} by the *strain* tensor (Aki & Richards, 2002)

$$\mathbf{e}_{ij} \equiv \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right). \quad (2.1)$$

The contact forces acting on an internal surface S of the continuum are described by the *traction* vector $\mathbf{T}(\mathbf{n})$ and the *stress* tensor σ_{ij} (Aki & Richards, 2002). The traction vector $\mathbf{T}(\mathbf{n})$ describes the force per unit area and quantifies the contact force at the surface. The vector \mathbf{n} is the normal unit vector at a considered point on S . The traction vector and the stress tensor are connected by the relation

$$T_i = \sigma_{ij} n_j. \quad (2.2)$$

The stress tensor is symmetric due to the conservation of angular momentum and the relation between the strain tensor \mathbf{e}_{ij} and the stress tensor σ_{ij} is given by *Hook's law* for the considered linear elastic medium (Aki & Richards, 2002) as

$$\sigma_{ij} = c_{ijkl} \mathbf{e}_{kl}. \quad (2.3)$$

The fourth-order tensor c_{ijkl} is called the *stiffness* tensor and contains 81 elastic moduli. Only 21 of the elastic moduli are independent in the considered case of an anisotropic medium (Aki & Richards, 2002).

Finally, a particle originally at position \mathbf{x} at time t_0 is furthermore subject of noncontact *body forces* $\mathbf{f}(\mathbf{x}, t)$ acting per unit volume. Due to the conservation of momentum within the considered medium with volume V and surface S the equation of motion is now found to provide the relation between accelerations, body forces and surface tractions (Aki & Richards, 2002). With the assumption that volume V and mass density ρ are constant with time the equation of motion can be given in its differential form as

$$\rho \frac{\partial^2 u_i}{\partial t^2} = f_i + \frac{\partial \sigma_{ij}}{\partial x_j}, \quad (2.4)$$

where u_i is the i -th component of the three dimensional displacement vector \mathbf{u} .

The concept of the elastodynamic Green's function is now introduced to develop 'representations for the displacements that typically occur in seismology' (Aki & Richards, 2002). Basic idea is to construct the displacement at point \mathbf{b} caused by a complicated (realistic) seismic source at point \mathbf{a} from the displacement produced by a very simple source. The simplest source which is well defined in space and time is the unidirectional unit impulse. The elastodynamic Green's function is therefore defined as the displacement at point \mathbf{b} caused by a unit impulse at point \mathbf{a} . The i -th component of displacement \mathbf{u} at point \mathbf{b} and time t caused by a unit impulse at point \mathbf{a} and time τ in the n -th direction is described by $G_{in}(\mathbf{b}, t; \mathbf{a}, \tau)$. The complete Green's function is a second-order tensor with nine components and depends on both receiver and source coordinates.

The Green's function in a volume V satisfies

$$\rho \frac{\partial^2}{\partial t^2} \mathbf{G}_{in} = \delta_{in} \delta(\mathbf{b}-\mathbf{a}) \delta(t-\tau) + \frac{\partial}{\partial x_j} (c_{ijkl} \frac{\partial}{\partial x_l} \mathbf{G}_{kn}). \quad (2.5)$$

It is defined that $\mathbf{G}_{in}(\mathbf{b}, t; \mathbf{a}, \tau)$ and $\partial \mathbf{G}_{in}(\mathbf{b}, t; \mathbf{a}, \tau) / \partial t$ are zero for $t \leq \tau$ and that $\mathbf{a} \neq \mathbf{b}$ as the initial conditions. If it is assumed that the boundary conditions are independent of time and homogeneous on surface S the space-time reciprocity $\mathbf{G}_{nm}(\mathbf{b}, t; \mathbf{a}, \tau) = \mathbf{G}_{mn}(\mathbf{b}, -\tau; \mathbf{a}, -t)$ of the Green's function can be derived as shown by Aki & Richards (2002). The Green's function is specified uniquely by the specification of the boundary conditions on S . Boundary conditions for different applications as well as the computation of Green's functions for different settings are discussed also by Aki & Richards (2002).

2.3.3.2 Estimation of the Green's function

In the last decade many authors showed theoretically (e.g. Campillo, 2006; Wapenaar et al., 2006; Gouédard et al., 2008a) and experimentally (e.g. Weaver & Lobkis, 2001) that the time domain Green's function of a medium between two points \mathbf{a} and \mathbf{b} can be derived by the cross-correlation of recordings of a coherent random wave field at those points. The convergence of the cross-correlation function to the Green's function is based on the equipartition of the wave field (Gouédard et al., 2008a). A suitable wave field can be caused by a random uniform distribution of uncorrelated noise sources, reverberations in an enclosure with an irregular bounding surface, multiple scattering between heterogeneities in a disordered medium or a combination of all of these causes (Wapenaar & Fokkema, 2006).

Recently, Tsai & Moschetti (2010) approved the assumption that the SPAC theory of Aki (1957) to retrieve the dispersion properties of surface waves from seismic noise is just another formulation of the interferometry approach discussed here. The first application of the interferometry approach is nevertheless commonly related to helioseismology notwithstanding the long assumed physical accordance between the seismic interferometry and the SPAC theory. The interferometry approach was used in helioseismology to retrieve time-distance information on the solar surface from recordings of the random motions of the Sun's surface (Duvall et al., 1993; Gilles et al., 1997). Later, the interferometry approach was used successfully on a laboratory scale to retrieve the Green's function of an aluminium sample by the cross-correlation of the thermal noise recorded at the sample's surface (Weaver & Lobkis, 2001). The breakthrough of seismic interferometry was the successful application of the interferometry approach to recordings of scattered coda waves (Campillo & Paul, 2003) as well as ambient seismic noise (Shapiro & Campillo, 2004) to retrieve the Rayleigh wave part of the Green's function. The estimation of Green's functions based on seismic noise cross-correlation functions (CCFs) evolved quickly to an important and widely used technique in seismology (e.g. Weaver, 2005; Curtis et al., 2006). It enables seismology to provide high-resolution tomography studies from local (e.g. Bussat & Kugler, 2009) to continental (e.g. Shapiro et al., 2005) scale and independent from earthquake seismicity or active seismic sources. This application of seismic interferometry is commonly called *ambient noise tomography*. The fundamentals and main difficulties of the Green's function estimation from seismic noise for ambient noise tomography and other applications are discussed in the following.

First step to estimate the time domain Green's function of Earth between two points \mathbf{a} and \mathbf{b} on the Earth's surface from seismic noise is the calculation of the cross-correlation function C_{ij} . The CCF C_{ij} is calculated from the recorded displacements $u_i(\mathbf{a}, t)$ at time t and location \mathbf{a} in direction i and $u_j(\mathbf{b}, t)$ at time t and location \mathbf{b} in direction j by integration over the whole observation period T with

$$C_{ij}(\mathbf{a}, \mathbf{b}, t) = \int_0^T u_i(\mathbf{a}, \tau) u_j(\mathbf{b}, t + \tau) d\tau. \quad (2.6)$$

The relationship between the cross-correlation function C and the time domain Green's function G is given by

$$\frac{\partial C_{ij}(\mathbf{a}, \mathbf{b}, t)}{\partial t} \approx -G_{ij}(\mathbf{a}, \mathbf{b}, t) + G_{ij}(\mathbf{a}, \mathbf{b}, -t), \quad (2.7)$$

as discussed by Gerstoft et al. (2006). Signals propagating from \mathbf{a} to \mathbf{b} yield a positive delay time t and signals propagating from \mathbf{b} to \mathbf{a} yield a negative delay time. In the case of a coherent random wave field with a uniform spatial distribution of seismic sources the CCF will be a symmetric function with respect to the delay time. The causal part represents the Green's function and the acausal part the time-reversed Green's function (Gerstoft et al., 2006). In general, long time series have to be considered to ensure a random spatial distribution of the sources and therefore sufficient averaging of signals propagating from \mathbf{a} to \mathbf{b} and vice versa (Gouédard et al., 2008a). All nine elements of the Green's function tensor can be estimated by the cross-correlation of all combinations of the components of motion (vertical, transversal, radial) recorded at both sites.

Practical experience with seismic noise shows, that one has to use not only long time series (months to years) but also to apply extensive normalisation and/or data selection to obtain CCFs which are suitable to reliably estimate Green's functions (Roux, 2009; Yao & van der Hilst, 2009). This is caused on the one hand by the dominance of single transient and coherent signals with comparably large amplitudes in the seismic noise wave field which disturb the averaging process. Most prominent examples of such transient signals are earthquake waves which are recorded coherently over long distances. On the other hand, a non-uniform distribution of noise sources also disturbs the averaging process and causes 'one-sided' CCFs (Gerstoft et al., 2006; Stehly et al., 2006). This effect is commonly observed due to ocean-generated microseism (e.g. Roux, 2009) which dominates the seismic noise wave field in a broad frequency range (see 2.2.2) and is often excited in a limited source area (e.g. at the near coast line). A detailed discussion of the sophisticated seismic noise cross-correlation processing which intends to reduce the impact of these problems by normalisation and/or data selection is given in chapter 6. Two new normalisation methods as well as two new fully automated approaches of data selection are introduced also in chapter 6.

Additional to the problems with dominating transient signals and non-uniform source distributions far most applications of ambient noise tomography are limited to the cross-correlation of the vertical component of the seismic noise wave field. This limitation is caused by the high effort necessary to cross-correlate all components of motion and to rotate the horizontal components for every station pair on the one hand as well as the properties of the seismic noise wave field itself on the other hand. The coherence of the horizontal component wave field is in far most cases significantly lower than that of the vertical component wave field (Gerstoft et al., 2006). This causes a lower signal to noise ratio of these components of the Green's function tensor which contain horizontal components of motion. This mainly explains the fact that far most applications of ambient noise tomography are based on the Rayleigh wave part of the Green's function. Nevertheless, reconstructions and applications of the Love wave part of the Green's function also exist (e.g. Lin et al., 2008; Roux, 2009).

An ongoing challenge is the reconstruction of the body wave part of the Green's function from seismic noise. The presence of P waves in the vertical component part of the Green's function was demonstrated by Roux et al. (2005) for the very dense and small (11x11 km²) Parkfield array at the San Andreas fault in California. The reconstruction of the complete Green's function including direct and reflected body waves was also demonstrated on a seismic-prospecting scale of less than 1000 meters (e.g. Gouédard et al., 2008c). Nevertheless, the reconstruction of the body wave part of the Green's function over distances larger than several kilometres is difficult due to the very small part of coherent body waves in the seismic noise wave field in this case. This explains the dominance of the slowly-attenuated surface (Rayleigh) waves in the estimated (vertical component) Green's functions on regional, continental and global scales (Gouédard et al., 2008a).

2.3.3.3 Applications of seismic interferometry

The ambient noise tomography discussed above is the most widespread application of seismic interferometry which is used from laboratory to continental scales as well as on land and on the ocean bottom (e.g. Bussat & Kugler, 2009). Nowadays, ambient noise tomography can be denoted as a standard tool of seismology. An out-standing example is the three-dimensional S wave velocity model of the Piton de la Fournaise volcano which was derived from seismic noise cross-correlation functions (Brennguier et al., 2007).

Another important application of seismic interferometry is the monitoring of temporal changes of the seismic velocities in the subsurface. Velocity variations of 0.1% or even less are derived from the temporal changes of the estimated Green's functions with a temporal resolution between one and a few days (Sens-Schönfelder & Wegler, 2006; Brennguier et al., 2008a). This application, sometimes called 'passive image interferometry' (Sens-Schönfelder & Wegler, 2006), is based on the analysis of coherent phases in the Green's functions caused by multiple scattered or reflected waves. This approach was successfully used to observe changes in seismic velocities at volcanoes (Sens-Schönfelder & Wegler, 2006; Brennguier et al., 2008a), fault zones (Wegler & Sens-Schönfelder, 2007; Brennguier et al., 2008b) and sedimentary basins (Meier et al., 2010). Some of these velocity variations can be explained by changes of the ground water table (Sens-Schönfelder & Wegler, 2006; Meier et al., 2010). Nevertheless, also changes of the structure inside volcanoes or on fault planes as well as changes of the stress state can be observed by such velocity variations. Therefore, seismic interferometry is a promising tool to improve the capabilities to forecast eruptions of volcanoes (Brennguier et al., 2008a) or even earthquakes. The same technique can be used furthermore to identify unwanted instrumental time shifts between the clocks of seismic stations (Stehly et al., 2007; Sens-Schönfelder, 2008).

Less widespread but anyhow noteworthy is the application of seismic noise cross-correlation functions to locate 'diffuse' sources of seismic energy such as ocean-generated microseism (e.g. Shapiro et al., 2006).

2.3.4 Other applications using seismic noise

A variety of other, less widespread, applications using seismic noise exist next to the well established and predominantly used array techniques and seismic interferometry. Most of these applications are based on the analysis of the spectral properties of seismic noise and how seismic noise varies with time and/or space. The new statistical time domain classification of seismic noise introduced in chapter 4 is capable to provide not only such amplitude information but also statistical information about the seismic noise dependent on time, space and frequency. All applications based on a spectral time-frequency analysis of seismic noise are therefore also prospective applications for the new time domain classification and may potentially take profit from the additional statistical information. Therefore, some of the applications based on temporal and/or spatial variations of the seismic noise wave field are introduced in this section.

Important is the application of seismic noise for environmental monitoring. One example is the approach to analyse temporal variations of the global climate with changing seismic noise (e.g. Grevemeyer, 2000; Aster et al., 2008). This approach is based on the observation of ocean-generated microseism and the assumption that variations of the global climate induce variations of the wave climate of the oceans. Duration, amplitude and frequency of events of ocean-generated microseism are analysed to infer information about the large-storm intensity. This is a promising approach to study the highly discussed changes of the global climate in the last century as continuous seismic recordings over decades are available at some seismic observatories.

Similar is the approach to obtain information about local or regional meteorological conditions from seismic noise. Several meteorological phenomena are known to induce seismic signals such as air pressure changes (Zürn et al., 2007) or wind (Withers et al., 1996; Groos & Ritter, 2009). An experiment with contemporaneous seismic and meteorological measurements, called METSEIS, was conducted by the Karlsruhe Institute of Technology in the last years (Ritter et al., 2009). The comprehensive data set is now analysed to retrieve site-dependent relationships between the observed meteorological parameters (e.g. air pressure, wind speed) and the seismic noise. The analysis of the METSEIS data set combines a spectral time-frequency analysis with the new statistical time domain classification.

Another interesting application of seismic noise for environmental monitoring was proposed by Burtin et al. (2008). They demonstrated that a spectral analysis of seismic noise can be used to monitor spatiotemporal changes in stream hydrodynamics. They recorded seismic noise in vicinity of the river Trisuli in the Himalaya and were able to explain a significant part of the observed seismic noise to be caused by ground vibrations generated by the bed load transport in the river.

The exploration of hydrocarbon reservoirs is a highly debated further application of seismic noise (e.g. Mohammed et al., 2007; Walker, 2008). This application, often denoted as the 'direct hydrocarbon indicator (DHI) method', is based on spatial variations of the spectral amplitudes and the H/V ratio of the seismic noise wave field in the frequency range between 1 Hz and 4 Hz. It is assumed, that the spectral amplitudes and the H/V ratio of the seismic noise wave field are changed significantly by a hydrocarbon reservoir in the subsurface. This assumption originates from few observations of seismic noise above or in the vicinity of known and producing hydrocarbon reservoirs. The DHI method is promoted to be able to explore reservoirs with low-cost passive seismic measurements. Several theories to explain the assumed effect are proposed and discussed but neither is proven so far. The debate about the approach is dominated by the question if the observed signal is actually caused by the natural seismic noise wave field and the reservoir alone. Hanssen & Bussat (2008) analysed the seismic noise recorded above and around a reservoir in the Sahara desert to test the DHI approach. They observe that regarded variations of the seismic noise wave field over the reservoir are caused by anthropogenic signals (technical processes related to oil production, building activity) and influenced by the topography. Several companies in hydrocarbon exploration and production industry are working on the DHI approach with significant expense although the underlying physics are debated and unclear. Therefore further developments and new insights regarding the DHI method can be expected in the future. With an improved understanding of the underlying physics the DHI approach may be a potential application for the statistical time domain classification.

3 Data

Two different data sets of ground motion recordings are used for this thesis and introduced in this chapter. The first data set (URS Bucharest, see 3.1) was measured in the metropolitan area of Bucharest during the URban Seismology (URS) experiment and is used to analyse the urban seismic noise wave field (chapter 5). The second data set (GSN data, see 3.2) consists of one year of recordings of several stations of the Global Seismographic Network (GSN) in the USA and is used for the analysis of the seismic noise cross-correlation processing (chapter 6). The specific data processing of the two data sets with the Karlsruhe Processing Toolbox for MATLAB (see Appendix A) is discussed in the respective chapters. The detailed metadata of both data sets are provided in Appendix B.

It is crucial for both applications to work with the ‘true’ ground motion without distortions of the phase or the amplitude of the signal due to the seismological measuring system. It is therefore necessary to remove carefully the corresponding instrument responses from the digital recordings of both data sets. While working with the GSN data set several inconsistencies in the instrument response metadata distributed together with the waveform data became apparent to the author (see 3.3). It turned out, that the metadata of dozens of seismic stations world-wide were (and partly still are) affected by these inconsistencies (see 3.3.7).

3.1 URS Bucharest

The URS project (Ritter et al., 2005) was conducted in Bucharest (Figure 3.1), the capital of Romania, whose 2.5 million inhabitants are endangered by devastating earthquakes from the nearby Vrancea subduction zone (for a comprehensive review see Wenzel et al., 1999). The measurements within the URS project were conducted with the KARlsruhe BroadBand Array (KABBA) owned by the Karlsruhe Institute of Technology. The URS data set was recorded with 32 24-bit data-loggers (EarthData) and 22 Streckeisen STS-2 (fundamental period $T_0=120$ s), five Geotech KS-2000 ($T_0=100$ s), two Güralp CMG40T ($T_0=30$ s), one Güralp CMG3ESP ($T_0=30$ s) and two Lennartz LE-3D5s ($T_0=5$ s) seismometers. These stations were deployed at 34 different sites within the metropolitan area of Bucharest from October 2003 until August 2004. Detailed information about the station locations and further metadata such as the instrument responses are given in Appendix B. The instruments were located mainly in cellars of public and private buildings. The network recorded continuously the ground motion velocity with a sampling rate of 100 Hz during 10 months. This data set provides an opportunity to analyse Urban Seismic Noise (USN) in a wide frequency range of 0.008-45 Hz with a spatial resolution of a few kilometres.

A careful removal of the instrument response is necessary to obtain comparable amplitudes of the seismic noise wave field from the different instruments in the frequency range above 1 Hz and below 90% of the Nyquist frequency. Above 1 Hz the frequency responses of the different used sensor types differ and are not perfectly flat. The amplitude response deviations from unity increase with frequency up to ± 3 dB at 50 Hz. For the results of the time domain analysis at frequencies above 1 Hz amplitude deviations from the true ground motion velocity of ± 0.6 dB (1-25 Hz) and ± 3 dB (25-45 Hz) would occur without the removal of the instrument responses from the data.

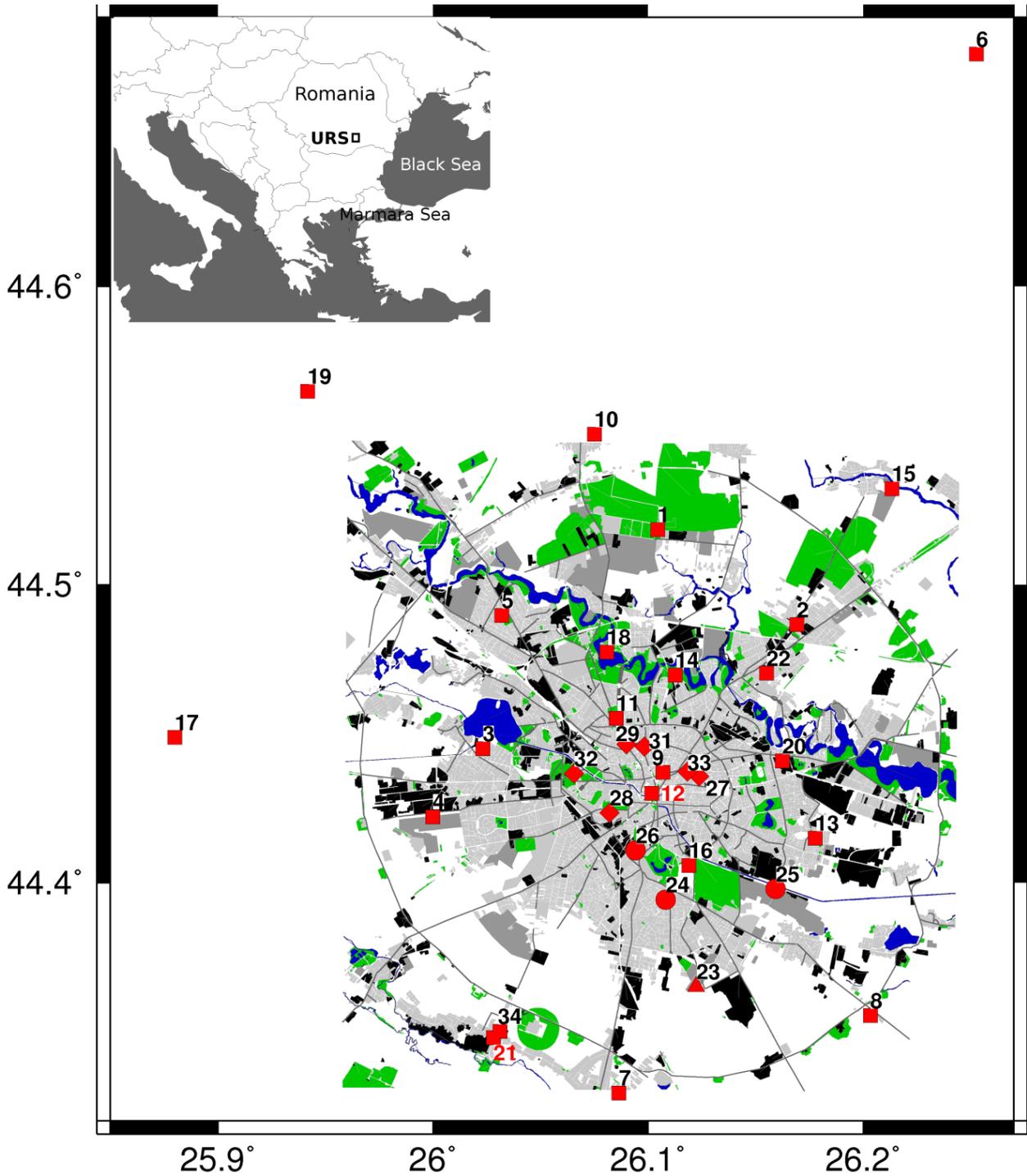


Figure 3.1: Station network of the URS project.

The map shows Bucharest and its surroundings with land usage. Different symbols indicate different sensor types, □: 22 Streckeisen STS-2, ◇: 5 Geotech KS2000, ○: 3 Güralp CMG40T/CMG3ESP and △: Lennartz LE3D/5s. The land usage is shown by colours, green: parks and woods, blue: lakes, rivers and waterways, black: heavy industry areas, light grey: residential areas, dark grey: commercial areas. The inset displays the regional context. The diameter of the characteristic ring road is about 20 km.

3.2 GSN data

The data set used for the study of the seismic noise cross-correlation processing consists of vertical component ground motion velocity recordings of the year 2004 with a sampling rate of 1 Hz (LHZ channel) of the Global Seismographic Network (GSN) stations ANMO (KS-54000; $T_0=333$ s), CCM (STS-1/VBB; $T_0=360$ s), DWPF (KS-54000; $T_0=333$ s), HRV (STS-1/VBB; $T_0=360$ s) and PFO (STS-1/VBB; $T_0=360$ s) in the United States of America (Figure 3.2). The waveforms and instrument responses were obtained as SEED (Standard for the Exchange of Earthquake Data) volumes from the Incorporated Research Institutions for Seismology (IRIS) Data Management Centre (DMC). Detailed station metadata and the instrument responses are also given in Appendix B.

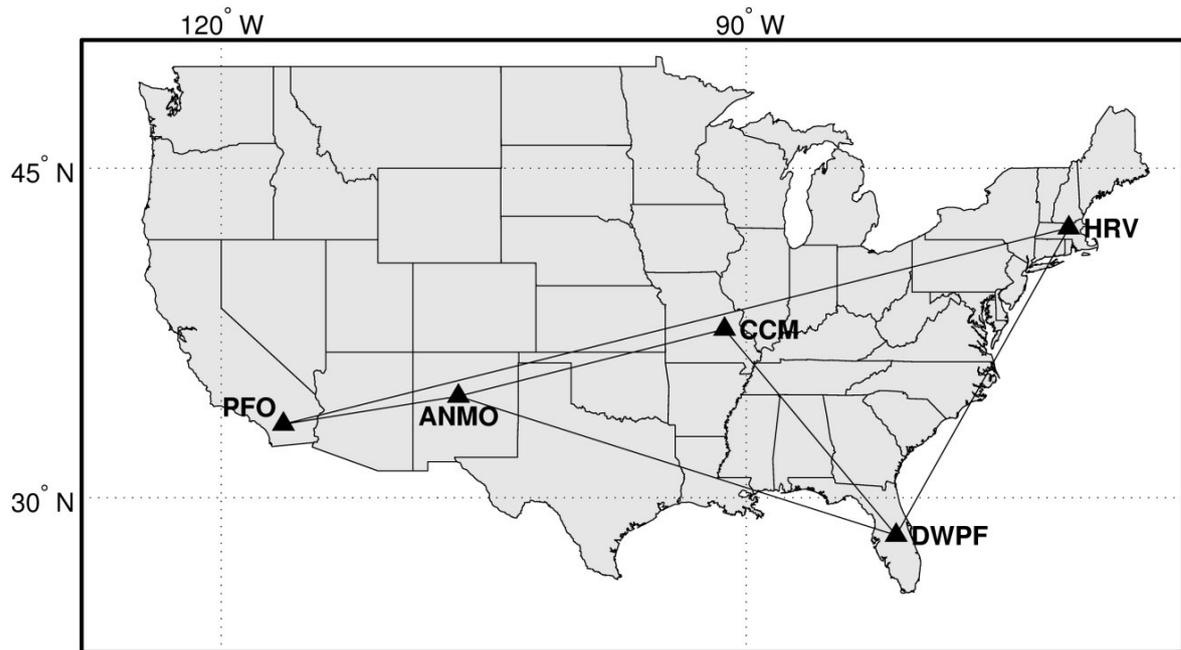


Figure 3.2: Map with used GSN stations in the USA.

The map shows the 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 DMC Data Management System was used for access to waveform and metadata of these stations. Six of the ten station pairs are highlighted in the discussion (chapter 6) and are therefore indicated by the connection lines.

The Cross-Correlation Functions (CCFs) are calculated and analysed for all ten station pairs (chapter 6). The six station pairs ANMO-DWPF, ANMO-CCM, PFO-ANMO, DWPF-HRV, CCM-DWPF and PFO-HRV indicated by the connection lines in Figure 3.2 are highlighted in the discussion in chapter 6. They represent all available azimuths as well as small, intermediate and large distances. The CCFs of some of these station pairs were also discussed by Bensen et al. (2007) in the context of seismic noise cross-correlation processing which was the motive for the selection of this data set. The fragmented raw time series obtained from the SEED volumes are merged to obtain complete 12-months time series of the ground motion velocity for each station. Missing data is zero padded. The alternative to the zero padding is the cutting of individual time windows with differing lengths and for every station pair with the exclusion of every time window for which only data of one station is available. This approach causes high organisational and computational effort due to the individual time window selection for every station pair as well as the resulting varying time window length. It is therefore usual to use a fixed time window length and accept zero padding (e.g. Bensen et al., 2007). Working with the GSN data set the author observed also no negative influence of the zero padding on the obtained CCFs in comparison to the strict exclusion of time windows for which only data at one station is available. Time windows which consist mainly of padded zeros are reliably

identified by the time series classification. The single time window CCFs obtained from such time windows are excluded from the stacking. It is necessary to remove the instrument responses from the data to cross-correlate the ground motion recordings without phase distortions due to the different sensor types. It was recognised during the preparation of the GSN data set that the instrument response metadata of station HRV deviate from the conventions of the SEED format. The recognised metadata inconsistencies were observed to affect several stations world-wide and are therefore discussed in detail in the following section.

3.3 Inconsistencies in SEED instrument response metadata

All main seismological data centres such as the IRIS DMC (www.iris.edu) distribute seismological data in the SEED format (see 3.3.1). The SEED format provides the possibility to distribute comprehensive metadata about the instrument responses together with the corresponding seismic recordings. All components of the seismological measuring system such as the sensor, the analogue-to-digital converter (ADC) or a digital anti-alias filter (DAA) can be described in the SEED format by their transfer functions. All of the observed inconsistencies discussed in this section concern the description of the DAA filters, which are digital low-pass decimation filters, used by modern data acquisition systems.

In the first subsection the SEED format is introduced together with the software which is provided by IRIS to work with SEED data. Section 3.3.2 introduces the seismological measuring system. A short summary about the general description of an instrument by its transfer function is given exemplary for the analogue part of the system, the seismometer, in section 3.3.3. In section 3.3.4 the description of the digital low-pass decimation filters of the measuring system is introduced in relation to the SEED format and the observed data format inconsistencies. The discussion of the measuring system and its description with transfer functions is strictly limited to the fundamentals necessary for the discussion of the observed data format inconsistencies in section 3.3.5. A comprehensive treatment of the fundamentals of digital seismology can be found in Scherbaum (2001). The impact of the observed inconsistencies on frequency response functions calculated with the widespread software *evalresp* is discussed in section 3.3.6. In the last section an overview is given about the occurrence of the data format inconsistencies in distributed SEED metadata.

The software written by the author to conduct the SEED metadata analysis and to calculate frequency response functions from SEED metadata is part of the Karlsruhe Seismology Toolbox (Appendix A).

3.3.1 The SEED format

The SEED format was developed mainly in the 1980s to provide a Standard for the Exchange of Earthquake Data especially between institutions of earthquake research (SEED Reference Manual, 2010). Main task of such a standard is to provide digital seismic recordings together with all necessary metadata (e.g. station coordinates, instrument responses, data quality reports, etc.). The development was promoted by the International Federation of Digital Seismograph Networks (FDSN, www.fdsn.org) which took over the responsibility to develop such a standard exchange format right after its foundation in 1985 (SEED Reference Manual, 2010). The FDSN holds commission status within the International Association for Seismology and Physics of the Earth's Interior (IASPEI) today. The first officially released version of the SEED format was version 2.0 in 1988 by Halbert, Buland and Hutt. The SEED manual, in recent version from May 2010, is maintained and provided by the FDSN. All important international seismological data centres such as IRIS or ORFEUS (Observatories and Research Facilities for European Seismology) use the SEED format to exchange data between each other and to distribute data to the seismological research community.

Naturally, a comprehensive data exchange format such as SEED is felt to be rather complicated by far most 'end-users', here seismologists. It is therefore common to obtain the desired data from a data centre as a SEED volume and to convert the data to a less comprehensive data format for further work. The most popular software for this conversion is called *rdseed* and is provided by IRIS (www.iris.edu/manuals/rdseed.html). The software *rdseed* is capable to convert the seismic recordings in a SEED volume to less comprehensive data formats such as SACbinary, gse, miniSEED or others which can be directly processed with wide-spread seismic analysis programs such as the *Seismic Analysis Code (SAC)* or *Seismic Handler (SH)*. The instrument response information contained in SEED volumes, which is not supported by far most of the other data formats, is written by *rdseed* to a human readable ASCII file, often denoted as the 'SEED RESP file'.

These SEED RESP files can be used furthermore as input for the software *evalresp* (<http://www.iris.edu/manuals/evalresp.htm>). This software is also provided by IRIS and calculates the discrete frequency response function (see 3.3.2) of a seismic channel from the instrument response information in the SEED RESP file. The wide-spread software packages SAC and *gsac* use frequency response functions calculated with *evalresp* to remove the instrument response from a seismic time series if a corresponding SEED RESP file is available.

Concluding, the process to obtain data from a data centre and afterwards to obtain the ground motion from these raw recordings is standardised and simplified for the end-user by wide-spread software. Drawback of this standardisation is the fact that inconsistencies in the metadata of the SEED data can affect the obtained ground motion time series without becoming apparent to the user.

3.3.2 The seismological measuring system

All time series used for this thesis were recorded with seismometers which consist of one (vertical component) or three (vertical and horizontal components) pendulums each with an electro-dynamic feedback system. Such instruments measure the ground motion velocity and correspond to the state-of-the art for modern broadband seismometers (Aki & Richards, 2002; Scherbaum, 2001). The output signal of this type of instrument is a voltage (one for each component of motion) which is digitised and stored as a discrete time series by a data acquisition system ('data-logger'). The analogue signal coming from the seismometer is often digitised with several sampling rates for different seismological applications. The different discrete data streams obtained from a single analogue sensor (e.g. components Z, N, E; several sampling rates) are denoted as 'channels'. The complete measuring system (sensor and data-logger) is considered to be a linear, time invariant (LTI) system which transfers the input signal $g(t)$ to the output signal $r[t]$. The different components of the measuring system are denoted as the 'stages' of the system. In the case of digital seismology the ground motion (displacement/ velocity/ acceleration) is the analogue input signal $g(t)$ and the stored recording of the ground motion (displacement/ velocity/ acceleration) is the discrete output signal $r[t]$. A typical measuring system is visualised in Figure 3.3 with the help of the BHZ (BH: seismic data sampled with 20 vertical component) channel of GSN station HRV.

The ADC of the data-logger provides the important conversion of the analogue signal $x(t)$ in volt coming from the seismometer to a discrete time series $x[t]$. This conversion introduces unavoidable quantisation noise to the signal (Scherbaum, 2001). An important technique to reduce the influence of quantisation noise and to improve the amplitude resolution is *oversampling*. The sampling of the analogue data stream by the ADC is conducted with a sampling rate several times higher than the finally desired sampling rate. The original discrete data stream $x[t]$ is then low-pass filtered and decimated (downsampled) to obtain finally the discrete time series $r[t]$ with the desired sampling rate. As a further advantage, no analogue anti-alias filter is necessary with this technique. A

comprehensive treatment of the oversampling technique can be found in Scherbaum (2001).

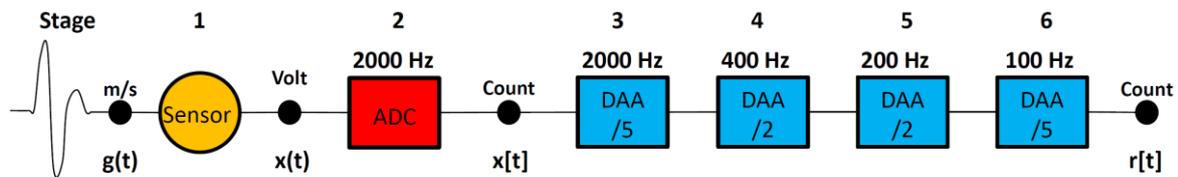


Figure 3.3: Block diagram of the BHZ channel (vertical component, 20 Hz) of GSN station HRV. The block diagram includes the sensor, the ADC and the low-pass decimation filter stages which are denoted also as Digital Anti-Alias (DAA) filter stages. The decimation factor of the DAA stages is given in the corresponding blocks. The input sampling rate is given above the corresponding DAA box. The original sampling rate of the ADC is 2000 Hz which is reduced by four DAA stages to the desired 20 Hz.

Important for this discussion is that common modern data acquisition systems contain such digital low-pass decimation or Digital Anti-Alias (DAA) filters (Scherbaum, 2001). In general, the downsampling from the original to the desired sampling rate is not obtained by one step but obtained by a cascade of several decimators with DAA filters as in the example presented in Figure 3.3.

It is obvious that the removal of the instrument response from the output signal $r[t]$ is capable to recover only $g[t]$ but not $g(t)$. Nevertheless, the concept of the transfer function is introduced in the next section by a discussion of the seismometer in the ‘analogue domain’ ignoring the analogue-to-digital conversion for now. In practice all operations are applied to discrete time series which is possible by replacing the analogue transforms by their discrete counterparts (Scherbaum, 2001). The digital DAA filters discussed in section 3.3.4 are developed and defined already in the ‘discrete domain’.

At this point it is necessary to clarify that only the instrument responses of the seismometers are removed from the time series of the URS and the GSN data set used for this thesis. The data-logger parts of the measuring systems are ignored!

This is justified as the digital stages of a seismological measuring system are not relevant for most seismological applications. Modern data-loggers are designed to avoid influences on the measured data as far as possible. The remaining influence on the signal is illustrated in Figure 3.4 again with the help of the frequency response function of the discussed BHZ channel of station HRV. The complete amplitude response (top) and phase response (bottom) spectra of the BHZ channel are plotted in black. The corresponding spectra of the seismometer frequency response are plotted in light blue. Most important and dominating part of the channel frequency response function is in fact the seismometer frequency response. A significant influence of the data-logger can be observed only near the Nyquist frequency of the channel (here 10 Hz) and only in the amplitude response spectrum. It is a rule of thumb, that the influence of a well-designed modern data-logger can be ignored if the upper boundary of the analysed signal’s frequency content is below 90% of the Nyquist frequency. It is therefore justifiable in far most cases as well as for the URS and the GSN data set to remove only the influence of the seismometers from the digital recordings to reconstruct the ‘true’ ground motion velocity in the analysed frequency range.

Nevertheless, the SEED format is capable to provide the instrument response information of all analogue and digital stages and most of the station operators include also the transfer functions of the DAA filter stages. The *evalresp* software of IRIS uses by default all instrument response information given in the SEED metadata. Therefore, inconsistencies in the SEED metadata of the DAA filter stages are relevant, especially as they are causing serious but hardly recognisable distortions of the obtained ground motion time series as is demonstrated in the following.

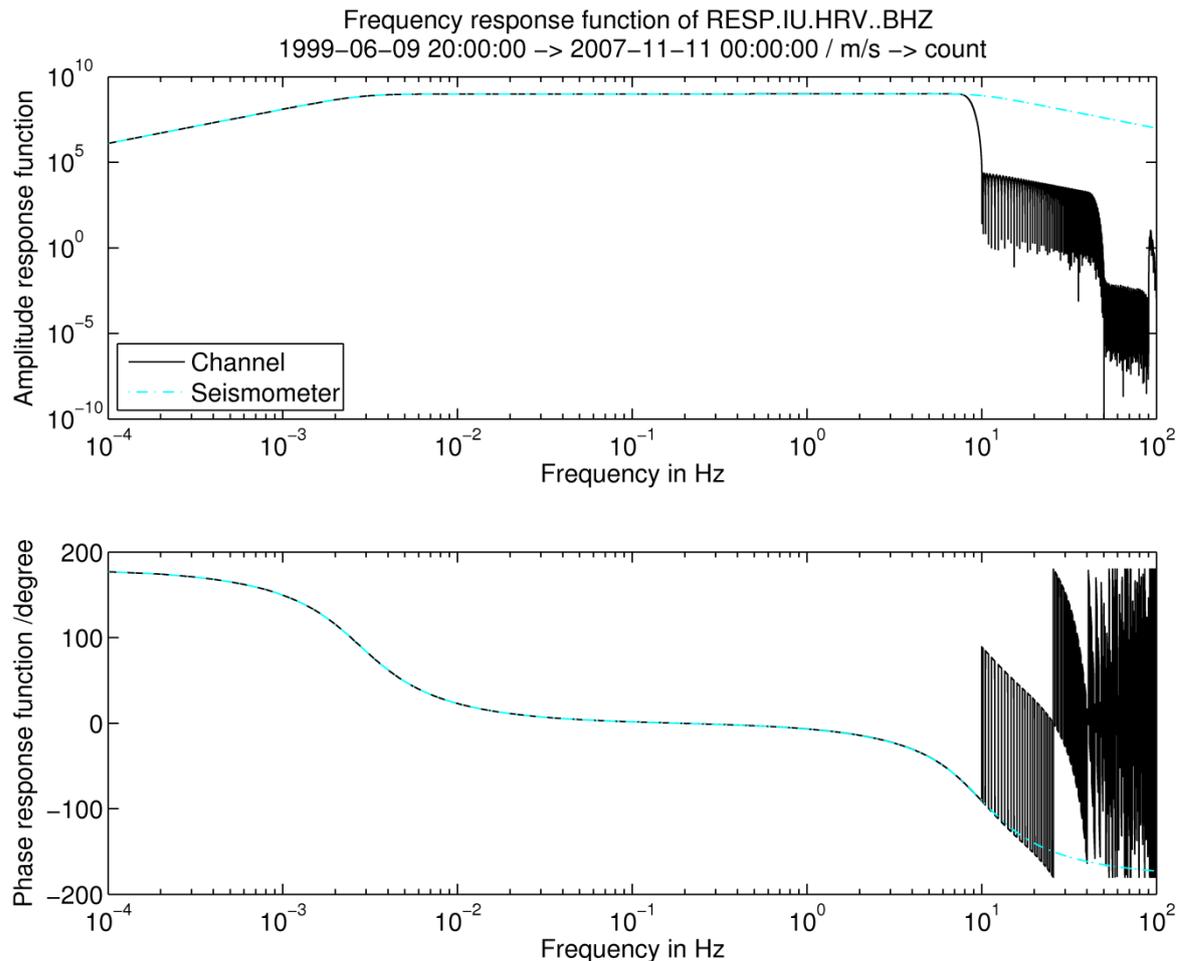


Figure 3.4: Frequency responses of the complete BHZ channel and the seismometer of GSN station HRV. The complete amplitude response (top) and phase response (bottom) spectra of the BHZ channel are plotted in black. The corresponding spectra of the seismometer frequency response are plotted in light blue. The channel frequency response function is dominated by the seismometer frequency response. An influence of the data-logger below the Nyquist frequency of the BHZ channel (here 10 Hz) can be observed only in the amplitude response spectrum.

3.3.3 The transfer function: Description of the seismometer

A LTI system such as a seismometer can be described in the frequency domain with the *frequency response function* $T(i\omega)$ of the measuring system by

$$X(i\omega) = T(i\omega)G(i\omega), \quad (3.1)$$

with $G(i\omega)$ the Fourier transform of the input signal $g(t)$ and $X(i\omega)$ the Fourier transform of the output signal $x(t)$ (Scherbaum, 2001). The transfer can be described also with the *transfer function* $T(s)$ using the more general Laplace transform as

$$X(s) = T(s)G(s), \quad (3.2)$$

with the complex variable $s = \sigma + i\omega$. The frequency response function $T(i\omega)$ can be obtained by evaluating the transfer function along the imaginary ω axis of the s -plane as the Laplace transform equals the Fourier transform if $s = i\omega$ ($\sigma = 0$). In practice, the discrete frequency response function is evaluated from the transfer function and the Discrete Fourier Transform (DFT) is used to remove the instrument response from the discrete time series.

All LTI systems can be described by a differential equation with constant coefficients (Scherbaum, 2001) which can be Laplace transformed to

$$[A_n s^n + \dots + A_1 s + A_0]X(s) = [B_m s^m + \dots + B_1 s + B_0]G(s). \quad (3.3)$$

The transfer function $T(s)$ is then given in its numerator-denominator representation by

$$T(s) = \frac{X(s)}{G(s)} = \frac{B_0 + B_1 s + \dots + B_m s^m}{A_0 + A_1 s + \dots + A_n s^n}, \quad (3.4)$$

which can be factorised and written in the pole-zero-gain representation

$$T(s) = \frac{X(s)}{G(s)} = \text{Gain} \frac{(s-q_1)(s-q_2)\dots(s-q_m)}{(s-p_1)(s-p_2)\dots(s-p_n)}, \quad (3.5)$$

with the poles p and the zeros q of the transfer function.

Both representations (amongst others) are used in practice to define transfer functions of LTI systems. The pole-zero representation is very common as the positions of the zeros and poles in the s -plane give easy accessible information about important characteristics of the system such as causality or stability (Scherbaum, 2001). Another possibility to describe a LTI system is its impulse response function which can be denoted as the ‘time domain representation’ of the frequency response function. The frequency response function $T(i\omega)$ is the Fourier transform and the transfer function is the Laplace transform of the impulse response function (Scherbaum, 2001).

As discussed above, the typical seismological measuring system consists of at least two parts, the seismometer and the data-logger. All components of the LTI system ‘seismological measuring system’ are, by definition, also LTI systems. Therefore, the transfer from $g(t)$ to $r[t]$ can be described also by the multiplication of several frequency response functions. To stay in the analogue domain it is assumed that an amplifier with frequency response function $T_A(i\omega)$ is installed at the output of the seismometer with frequency response function $T_S(i\omega)$ and $x(t)$ is now the output signal of the amplifier. The transfer from $g(t)$ to $x(t)$ can be described now by

$$X(i\omega) = T_A(i\omega)T_S(i\omega)G(i\omega). \quad (3.6)$$

In terms of the SEED format the two parts $T_S(i\omega)$ and $T_A(i\omega)$ are denoted as ‘stages’ of the complete frequency response function $T(i\omega)$ (SEED Reference Manual, 2009). As already visualised in Figure 3.3 the complete frequency response function $T(i\omega)$ of a seismological measuring system is subdivided in more than two stages with own frequency response functions for the different components. This segmentation of the complete frequency response function of the system in stages allows the description of every component of the complete system in the most suitable domain (analogue/discrete) and representation (poles-zeros/numerator-denominator). In practice, the complete frequency response function of a channel is then finally obtained by multiplying the discrete frequency response functions of all stages. Analogue components such as the sensor or an optional amplifier are commonly described by the poles and zeros of their transfer functions in the s -plane. The poles and zeros of the seismometers relevant for this thesis are given in Appendix B.

Digital components which are applied to the already digitised data stream are commonly described by the numerator and denominator coefficients of the *discrete transfer function* $T(z)$ in the ‘digital’ z -plane (Scherbaum, 2001; SEED Reference Manual, 2009). The discrete transfer function $T(z)$ is discussed together with the low-pass decimation filters of the data-logger in the next section.

3.3.4 The digital low-pass decimation filter

The digital low-pass decimation, or Digital Anti-Alias (DAA), filter is also a LTI system and the definition of the filter with a frequency response function, transfer function and impulse response function is in principal the same as for the analogue seismometer. The difference is the usage of the discrete instead of the analogue transforms. The Laplace-transform is replaced by the z-transform and the Fourier transform is replaced by the discrete Fourier transform (DFT). The DFT equals the z-transform evaluated on the unit circle in the z-plane (Scherbaum, 2001).

The bilateral z-transform of a discrete sequence $x[t]$ (t denotes now the sample index) is defined as

$$X(z) = \sum_{t=-\infty}^{\infty} x[t]z^{-t} \quad (3.7)$$

with the complex variable $z = e^{s\Delta t}$ (Scherbaum, 2001). The complex variable s is already known from the Laplace-transform and variable Δt is the time interval between two samples of $x[t]$.

The output $r[t]$ is related to the input $x[t]$ by the linear difference equation

$$\sum_{k=0}^N a_k r[t-k] = \sum_{l=0}^M b_l x[t-l], \quad (3.8)$$

which leads to the rational transfer function $T(z)$ as

$$T(z) = \frac{\sum_{l=0}^M b_l z^{-l}}{\sum_{k=0}^N a_k z^{-k}}. \quad (3.9)$$

The digital filters discussed in the following are defined by their numerator coefficients b_l and their denominator coefficients a_k which is a commonly used representation of digital filters. This form of representation is also recommended by the SEED format for the description of DAA filters.

In filter theory two important classes of digital filters are distinguished. It can be derived from equation (3.8) that the output signal $r[t]$ depends not only on the input signal $x[t]$ but also on earlier samples $r[t-k]$ of the output signal if the number of denominator coefficients a_k is larger than one. These filters are denoted as recursive, or *Infinite Impulse Response* (IIR), filters. Filters with only one denominator coefficient $a_0=1$ are denoted as non-recursive, or *Finite Impulse Response* (FIR), filters. A discussion of the differences between IIR and FIR filters is far beyond the scope of this text. Important is the fact that DAA filters in modern data acquisition systems are realised as FIR filters due to some significant advantages (Scherbaum, 2001). Amongst other advantages a FIR filter is always stable and it is easy to design a FIR filter which fulfils desired characteristics.

The FIR filters can be further subdivided into several classes of which two are important for the discussion of seismological data acquisition systems. FIR filters can be designed as *minimum-phase* and *linear-phase* filters in respect to their phase response spectrum. Loosely-speaking, the minimum-phase filters are designed to meet the desired amplitude characteristics by introducing as less phase distortions to the signal as possible. The linear-phase filters are designed with a linear phase response which means that they cause no phase distortions of the signal and produce only a constant time shift of the signal. A linear-phase filter is denoted as *zero-phase filter* if the time shift is removed from the signal after the filtering. Linear-phase filters are often denoted also as *acausal filters* and minimum-phase filters are often denoted also as *causal filters* which will become

apparent from the following discussion. The underlying difference causes some significant advantages and disadvantages of both FIR filter types if they are used as DAA filters in a seismological data acquisition system. The most important differences are discussed below. A comprehensive treatment of FIR filters can be found in Scherbaum (2001).

The difference is discussed with the help of the first (stage 3) and the last (stage 6) DAA filter of the BHZ channel of station HRV (Figure 3.3). The coefficients b_l and the phase response spectra of both DAA filters are presented in Figure 3.5. The filter of stage 3 (decimation from 2000 Hz to 400 Hz) is shown on the left side of the figure in comparison to the filter of stage 6 (decimation from 100 Hz to 20 Hz) on the right side. All linear-phase filters such as the DAA of stage 6 (right side of the figure) have symmetric coefficients b_l and are therefore denoted as two-sided or *acausal* filters. It is a consequence of this symmetry, that onsets of very impulsive signals (such as a P wave onset) may be obscured by precursory ('acausal') oscillations (Scherbaum, 2001). Such oscillations may be identified by mistake as a seismic signal and cause misleading seismological interpretations. Furthermore, the exact picking of first onset times is biased by such precursory signals. A comprehensive treatment of this subject is given by Scherbaum & Bouin (1997).

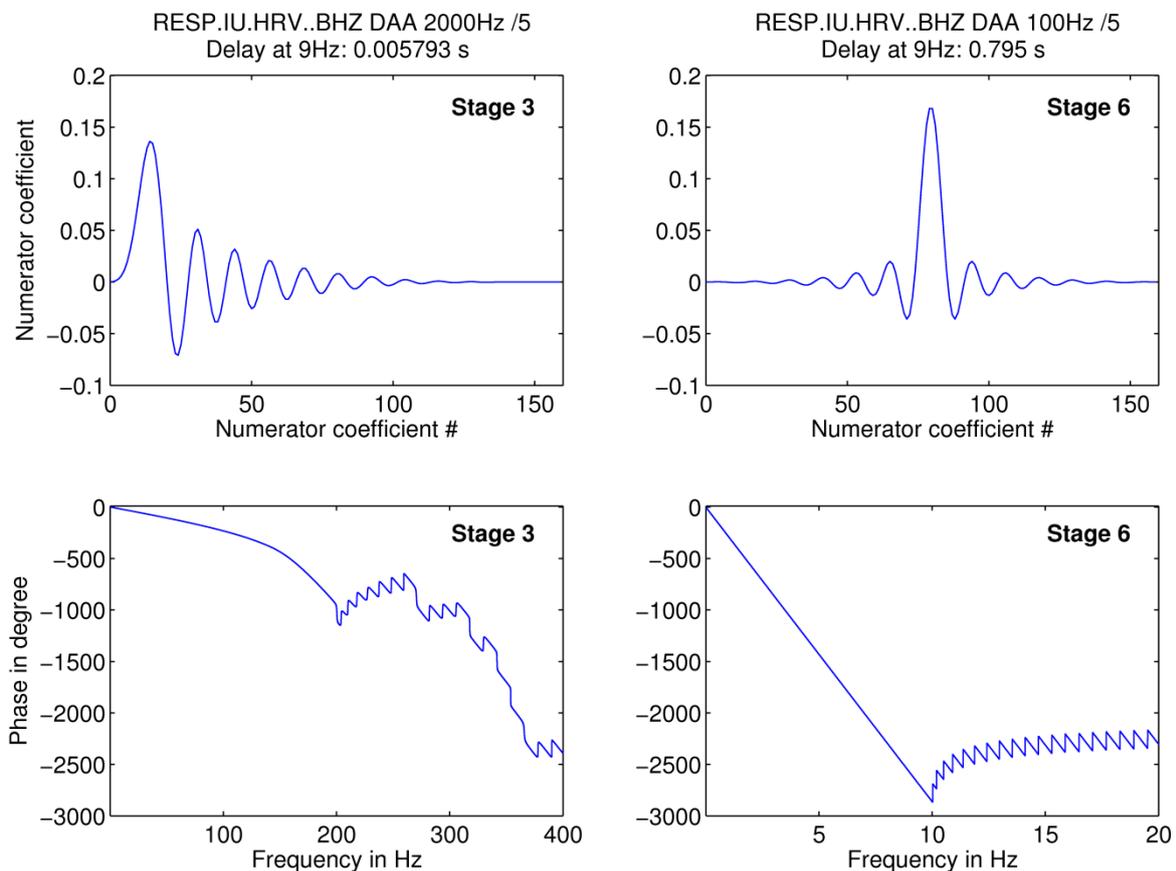


Figure 3.5: Filter coefficients b_l and phase response spectra of two DAA stages of HRV.BHZ. On the left side the first (minimum-phase) DAA filter (stage 3) and on the right side the last (linear-phase) DAA filter (stage 6) of the BHZ channel are shown.

The advantage of such a symmetric filter is the linear response spectrum below the Nyquist frequency of 10 Hz of the obtained time series after the decimation from 100 Hz to 20 Hz. This linear-phase spectrum corresponds to a constant time shift, or *delay*, of the entire time series of 0.795 seconds which is automatically removed by the data-logger.

The delay D in seconds of a symmetric linear-phase FIR filter with M coefficients applied to a time series with a sampling interval of Δt in seconds can be easily calculated by

$$D = \frac{(M-1)}{2} \Delta t. \quad (3.10)$$

The advantage of the asymmetric FIR filter of stage 3 on the left side of Figure 3.5 is that this type of FIR filter causes no spurious ‘acausal’ signals before impulsive signals. The disadvantage of this filter type is the non-linear phase response spectrum below the Nyquist frequency of 200 Hz of the decimated time series. Nevertheless, the low frequency part of the phase response spectrum is linear below the Nyquist frequency of the finally desired time series (10 Hz). This linear part of the phase response corresponds to a delay of 0.0058 s of the signals below 10 Hz which is also removed automatically by the data-logger. It is common practice to use asymmetric FIR filters for the first DAA filter stages at high sampling frequencies to avoid problems with artificial ‘acausal’ signals. This is done as long as the phase response spectrum is linear below the finally desired Nyquist frequency of the channel. That means that at least the last DAA filter stage is in general realised as a symmetric zero-phase filter to avoid phase distortions below the Nyquist frequency of the channel. This is the most common DAA filter configuration for seismic data-loggers. For some specific applications also the last DAA filter stage can be realised as an asymmetric FIR filter. This is done for example for data channels with very high sampling rates above 80 Hz which are used only for the picking of onset times with very high precision.

Concluding, the definition of a DAA filter stage consists of the filter coefficients b_i and the delay D which is corrected automatically by the data-logger. It is well defined in the SEED manual how the coefficients (page 73, SEED blockette 54) and the corrected delay (page 77, SEED blockette 57, field 8) of a DAA filter stage have to be given in the instrument response metadata.

3.3.5 The inconsistencies

The usual way to define DAA FIR filters in the SEED metadata is to give the numerator coefficients of the transfer function in the z -plane using SEED blockette 54. The SEED convention of blockette 54 regarding the filter coefficients is to give them in *forward order*, which means from b_0 to b_M .

Information about the input sampling rate, the decimation factor and the delay of the DAA filter stage is given with SEED blockette 57. Regarding the time delay of the filter stage two fields are provided by blockette 57. Field 7 provides the ‘Estimated delay (seconds)’ and field 8 provides the ‘Correction applied (seconds)’ of the filter stage. Relevant for the calculation of the frequency response is only field 8. It is mandatory for the correct calculation of the frequency response function of the channel that the time shift which was automatically corrected by the data-logger is given in field 8. It is furthermore crucial that the correct sign in terms of the SEED definition is used. All FIR filters **delay** the signal, which means that the signal is shifted towards *later* times by the filtering. This ‘estimated delay’ of the filter can be given in field 7, although this information is not necessary to calculate the frequency response. After the filtering the data-logger corrects the time series and shifts the time series back towards *earlier* times. This time shift to correct the time series has to be given correctly in field 8 ‘correction applied (seconds)’. The SEED manual defines both time shifts, as they are described above, to be positive.

Concluding, both delay values in the fields 7 and 8 have to be positive and larger than zero if they represent a real FIR filter in a well-designed data-logger and follow the SEED definitions. The ‘correction applied’ value in field 8 is mandatory to calculate the correct frequency response. Without the correct delay value in field 8 the corresponding FIR filter described by blockette 54 acts as an linear-phase filter when the frequency response

function is calculated and not as a zero-phase filter as it is actually realised in the data-logger. The FIR filter coefficients have to be given necessarily in forward order with blockette 54.

The inconsistencies: Blockette 57 Field 8 ‘correction applied (seconds)’

As discussed above the value of the ‘correction applied (seconds)’ information in the SEED metadata has to be positive and larger than zero. Two common errors regarding this field can be found in SEED metadata.

The first common error is a different sign convention to describe the time shifts which means that a negative value is given instead of a positive in field 8. It is not unusual that a positive delay value is given in field 7 and a negative delay value in field 8. The time delay of the DAA filter stage is not corrected but doubled if the frequency response function is calculated in accordance to the SEED definitions in this case. The entire time series is shifted towards earlier times by the removal of such an incorrect frequency response. The time shift depends on the DAA filter configuration of the data-logger and can be expected between 1 s and 2 s.

Another common error is a ‘correction applied’ value of zero. This error is most often found with symmetric FIR filter stages and may be caused by the fact that these filters are commonly denoted as ‘zero-phase’ filters. Nevertheless, if the correction is given as zero in field 8 the frequency response function of a linear-phase filter is calculated and the time series is shifted also towards earlier times if the incorrect frequency response is removed from the data. The time shift can be expected between 0.5 s and 1 s.

The inconsistencies: Order of the filter coefficients

While working on the GSN data set it became apparent to the author that the filter coefficients of the DAA filter stages of all channels of station HRV are given in *reverse order*, which means from b_M to b_0 . It is obvious, that the coefficient order is only relevant for the asymmetric FIR filters. A ‘reversed’ asymmetric filter has the same amplitude response spectrum as the original filter but a significantly different phase response spectrum. This is demonstrated for the stages 3 and 5 of the channel HRV.BHZ in Figure 3.6. The filter coefficients (top), amplitude response (middle) and phase response (bottom) of the correct FIR filter (forward order, black) as well as of the incorrect FIR filter (reverse order, red) are presented for stage 3 on the left side and for stage 5 on the right side. The amplitude response spectra don’t change due to the coefficient order in contrast to the phase spectrum. The delay of the reversed filters below the finally desired Nyquist frequency of 10 Hz increases significantly from 0.0058 s to 0.0737 s and from 0.0218 s to 0.4532 s. The reversed order of the first DAA stage is not a problem as the increased time delay of the filter is smaller than 2 samples of the resulting BHZ channel. The reversed order of the last DAA stage instead causes a serious problem. The introduced time shift of ~0.4 s is larger than the time differences between modelled and observed arrival times which are commonly analysed in seismology.

A summary of all relevant delay times of channel IU.HRV.BHZ is given in Table 3.1. The delay times of the correct FIR filters with the coefficients in forward order equal the delay times which are given in the SEED metadata for the ‘corrected delay’ in field 8. The delay times of the filters actually defined incorrectly in the SEED metadata are significantly larger due to the reversed coefficient order.

Channel IU.HRV.BHZ (1999-06-09 20:00 – 2007-11-11 00:00)

Delay in s of	Stage 3	Stage 4	Stage 5	Stage 6	SUM
1 Filter with forward order	0.0058	0.0105	0.0218	0.795	0.8331
2 SEED info 'corrected delay'	0.0058	0.0105	0.0211	0.795	0.8324
3 Filter defined in metadata	0.0737	0.2269	0.4532	0.795	1.5488
Difference (3-2)	0.0679	0.2164	0.4321	0	0.7164

Table 3.1: Summary of the delay times of channel IU.HRV.BHZ.

The delay times in lines '1' and '3' are calculated from the phase response spectra of the correct filter in forward order (1) and the filter given in the SEED metadata in incorrect reverse order (3). The delay times given in field 8 'corrected delay' of the SEED metadata are given in line (2) for completeness and equal the delay times of the correct filters in line 1. The difference between the added delays of the correct filters (1) and the added delays of the reversed filters (3) causes a time shift of the time series of 0.7164 s towards earlier times.

If the ground motion is recovered from the recordings with the incorrect frequency response function a time shift is introduced which shifts the entire time series 0.7164 s towards earlier times (Table 3.1). In other words: Onset arrival times are picked significantly earlier than they really are. The largest time shift occurs if the order of the filter coefficients is reversed and the 'correction applied' value in field 8 is given with a negative sign. In this case a time shift towards earlier times of more than 2 s may be introduced to the time series.

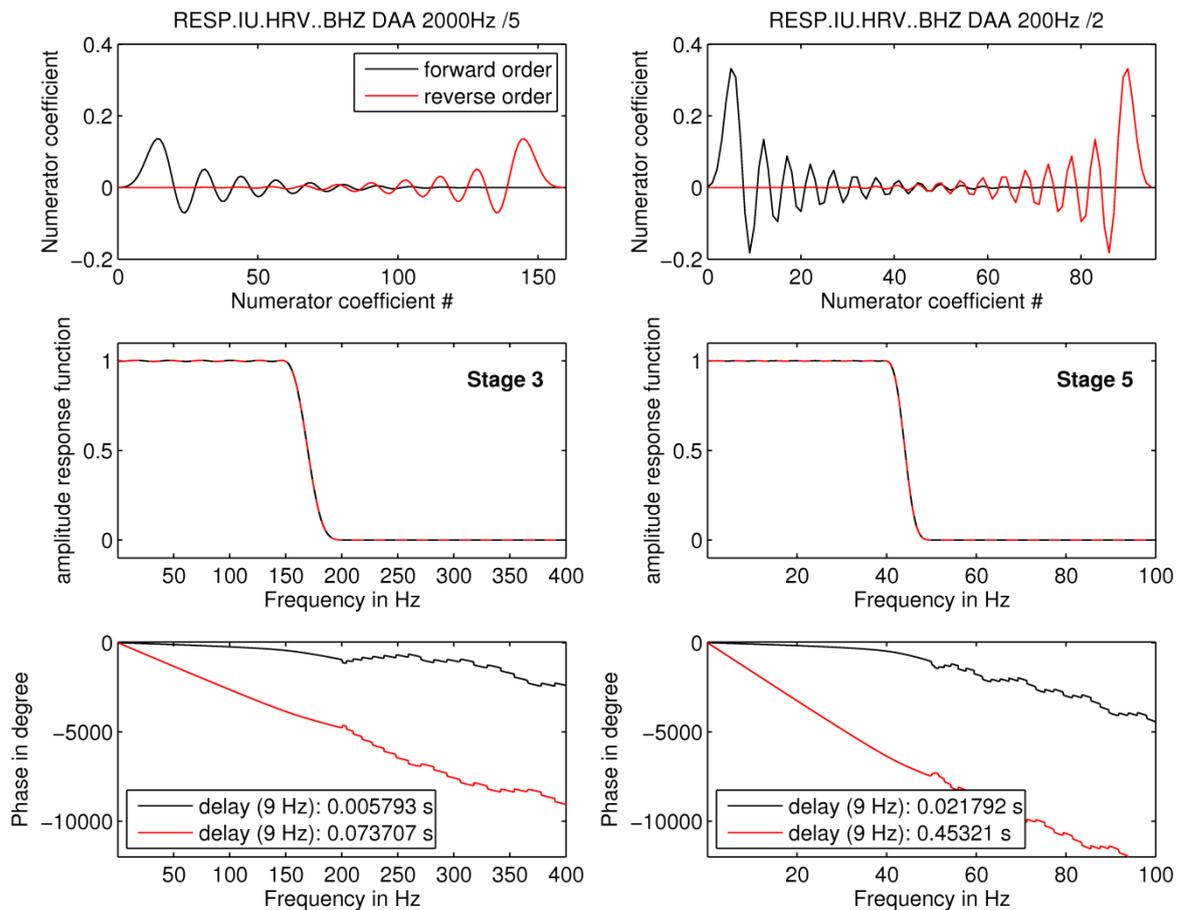


Figure 3.6: Filter coefficients b_i with amplitude and phase response spectra of stages 3 and 5 of HRV.BHZ. On the left side the first DAA filter (stage 3) and on the right side the third DAA filter (stage 5) of channel HRV.BHZ are shown with correct forward (black) and false reverse (red) coefficient order. Presented from top to bottom are the filter coefficients, the amplitude response spectra and the phase response spectra.

3.3.6 Some remarks about *evalresp*

The above calculated and estimated time shifts due to the described data inconsistencies imply a calculation of the frequency response in accordance to the SEED definitions. The situation is getting more complicated if the frequency response functions are calculated with the wide-spread software *evalresp*, especially with versions prior to 3.3.0. Version 3.3.0 of *evalresp* was released in July 2009 due to a correction of the frequency response function calculation of FIR filter stages. Versions prior to 3.3.0 ignored the 'correction applied' information in field 8 of blockette 57. The delay of all FIR filter stages was estimated by the simple way presented in equation (3.10) for symmetric linear-phase filters. Consequently, the frequency response functions of asymmetric FIR filters were incorrect due to a significantly overestimated time delay. In the case of the example IU.HRV.BHZ (with correct coefficient order!) a time shift towards **later** times of about 0.36 s is introduced to the time series by the application of the incorrect frequency response function. Consequently, *evalresp* versions prior to 3.3.0 are by definition not suitable to calculate frequency response functions of channels with asymmetric FIR filter stages defined in the SEED metadata. This affects several of the seismic networks distributed over IRIS such as the Caltech Regional Seismic Network (CI). The behaviour of *evalresp* in previous versions regarding the asymmetric FIR filter stages was documented in the manual but without a clear remark about the introduced time shift. Furthermore, no warning was or is echoed by *evalresp* if a channel response contains one or more asymmetric FIR filter stage(s). It can be assumed that the problem is and was not apparent for many users.

As mentioned above, the calculation of the frequency response of the FIR stages was corrected with version 3.3.0. Version 3.3.0 and the recent version 3.3.3 (March 2010) use the 'correction applied' information in blockette 57 field 8 (B57.F8) to calculate the correct frequency response of asymmetric FIR filters. The delay of symmetric FIR filters is still estimated by the number of coefficients and the sampling rate which effectively avoids problems with 'correction applied' values set incorrectly to zero (see also the next section). The recent version 3.3.3 is now furthermore echoing warnings if the 'correction applied' value is negative as it was suggested also by the author in September 2009.

Concluding, the observed metadata inconsistencies and the different *evalresp* versions cause a complex and difficult situation. Corrections and changes of the software as well as the metadata over the years lead to a phenomenon known as 'data aging'. This term describes the situation that the same fundamental data (the raw seismic recordings) yield different results regarding only to the point in time when the software and the (meta)data were obtained by the user. Data aging is a serious problem in seismology.

3.3.7 Occurrence of the inconsistencies

Large data centres such as the IRIS DMC provide 'dataless SEED volumes' which contain no time series but the complete instrument response history of entire seismic networks. The dataless SEED volumes of several networks were obtained in September 2009 from the ftp-archive of the IRIS DMC (http://ftp.iris.washington.edu/pub/RESPONSES/DATALESS_SEEDS/) for a simple spot check evaluation of the frequency of occurrence of the identified SEED metadata inconsistencies.

The evaluation is limited to the responses of the BHZ channels valid at the reference date 2004-June-01 12:00 UTC of selected seismic networks due to the large amount of available stations and channels. In total 26 networks are selected containing the most important large networks such as major parts of the GSN (IU, II, IC, ...), USArray (US), Geofon (GE) and Geoscope (G) as well as regional networks such as the Berkeley Digital Seismic Network (BK), Caltech Regional Seismic Network (CI) and several national networks (e.g. DK, CH, CZ, TW, JP). The selection is neither comprehensive nor representative and is therefore only capable to provide a first impression about the situation for some of the most important networks which is given in Table 3.2.

2004-June-01 12:00
Download September 2009

Number of channels affected by inconsistencies

Network [total number of analysed channels]	IU [118]	II [67]	CZ [5]	GE [53]	G [27]	US [32]	CI [156]	BK [24]	OE [6]
False coefficient order	1		1	13		10	126		1
B57.F8: negative					6		153	24	
B57.F8 zero (asym. FIR)				13				8	1
B57.F8 zero (sym. FIR)		67		24	5	32			2

Table 3.2: Summary of the instrument response metadata analysis.

The numbers of the affected channel responses (valid at 2004-June-01 12:00) are given for the different inconsistencies and nine of the 26 analysed networks. All networks which contain channel responses with FIR filter coefficients in reverse order and/or negative 'correction applied' values in field 8 of blockette 57 (B57.F8) are included in this table. This analysis is not representative as it is only a spot check of the metadata available at the IRIS DMC.

In total the responses of 559 BHZ channels were obtained from IRIS for the reference date. The table is limited to the most important and/or affected networks of the 26 analysed networks. All networks which are found to contain channel responses with incorrect FIR coefficient order and/or a negative 'correction applied' value in field 8 of blockette 57 are included in Table 3.2. This analysis demonstrates that inconsistencies in the SEED instrument response metadata are not uncommon. Significant differences between the networks are noticeable and can be explained by the fact that they are maintained by different operators. The responsibility for the correctness of the metadata lies solely with the network operators (personal communication with IRIS DMC staff). The inconsistencies identified by this analysis were reported to the IRIS DMC and several network operators.

One example of a common systematic discrepancy from the SEED conventions is the definition of the 'correction applied' value (B57.F8) to zero. This is especially the case for symmetric FIR filter stages (see networks II and US). This is not in accordance with the SEED conventions but fortunately of less practical relevance as the time delay of a symmetric FIR filter can be easily determined from the number of coefficients and the corresponding input sampling rate for the calculation of the correct frequency response function. All versions of *evalresp* are doing this by default. The corrected time delays of the asymmetric FIR filter stages are in far most cases smaller than the sampling interval of the finally obtained channel as they are used only for the first DAA filter stages at high sampling rates.

More critical is the consequent definition of negative 'correction applied' values as it was observed for the networks BK and CI. The impact of this inconsistency can be assumed to be significantly moderated in practice as the wide-spread software *evalresp* ignores the delay values of symmetric FIR filters. Nevertheless, other software packages calculating frequency response functions from SEED RESP files may be affected. Negative delay values defined for asymmetric FIR filter stages lead also the recent *evalresp* version to calculate incorrect frequency response functions. Fortunately, a negative delay value causes *evalresp* now to echo a warning in the recent version 3.3.3 from March 2010. Version 3.3.0 calculates the incorrect frequency response function without warnings. Older versions are by default not suitable to calculate frequency response functions of asymmetric FIR filters as discussed above. The total time shift introduced by a cascade of asymmetric FIR filters with negative delay values is around two or three times the sampling interval. This time shift may be neglected for most applications in seismology. The negative delay values in the metadata of network BK were already corrected in 2010 and the metadata of network CI is under revision.

The inconsistency with the most significant practical relevance (incorrect coefficient order, time shifts ~0.5-1 s for BHZ channels) is observed for several networks (CI, GE, US, CZ,

IU, OE). Nearly all stations of the very homogeneous network CI are affected as they have an identical hardware configuration. A detailed analysis of all channels found with reversed filter coefficients revealed a common systematic cause for this type of metadata inconsistency. All affected stations were equipped at the reference date with data-loggers of the manufacturer *Quanterra*. The stations were (and partly still are) equipped with data-loggers of the models *Q4120* (CI, GE, OE, CZ, IU) and *Q730N* (US). Both data-logger models use the same FIR filters for the DAA stages. The manuals of these data-loggers are not freely available but their FIR filters can be obtained as ASCII files from the ftp-archive of the *Quanterra User Group* (<ftp://quake.geo.berkeley.edu/pub/quanterra/>). The provided ASCII files contain the FIR filter coefficients in reverse order. It can be speculated that in some cases these ASCII files were used for the compilation of SEED metadata without checking of the coefficient order.

The SEED metadata of network GE was already corrected in 2010 and the metadata of network CI is under revision. The problem with the reversed FIR filter coefficients was reported to the IRIS DMC.

3.3.8 Summary

The SEED format is the common standard for the distribution of digital seismological time series. It provides the time series together with the complete metadata of the seismic measuring system. While working on the GSN data set (see section 3.2) the author became aware of data format inconsistencies in the SEED metadata of the GSN station HRV. All inconsistencies affect the description of FIR filters which are commonly used for the decimation stages of digitisers.

The filter coefficients of asymmetric FIR filters are given sometimes in reverse order instead in forward order. This is an inconsistency with practical relevance as the filter coefficients in reverse order bias the phase response spectrum of the channel. The time series is shifted towards earlier times by the removal of the biased frequency response function. The amount of the time shift depends on the FIR configuration of the channel and can be up to 1 s. The FIR filter stages of 559 BHZ channels distributed by the IRIS DMC were analysed in September 2009 and 152 BHZ channels were found to contain FIR filter stages with reversed filter coefficients.

Another common inconsistency affects the value for the corrected filter delay time of a FIR filter stage. The corrected filter delay time has to be positive in every case following the SEED definition but is often found as a negative value (183/559 analysed BHZ channels) or to be zero (153/559 analysed BHZ channels). This inconsistency is fortunately of less practical relevance as the wide-spread program *evalresp*, which calculates frequency response functions from the SEED metadata, is not using the given delay time to correct symmetric FIR filter stages. If the false values of symmetric FIR filters are used to calculate the frequency response function a time shift up to 2 s is introduced to the time series. The time shift due to asymmetric FIR filter stages with false time delay values is in general very small (2 to 3 times the sampling interval) and can be neglected in far most cases.

The problems with the reversed FIR filter coefficients and the time delay values were reported to the IRIS DMC and several network operators. The SEED metadata of several affected networks is already corrected.

4 Time series classification

In this chapter a time series classification approach is introduced to analyse large seismic noise data sets such as the URS data set (section 3.1; Groos & Ritter, 2009) by classifying and quantifying seismic noise time series in the time domain. This approach combines amplitude information with few distinct noise classes (Gaussian distribution, presence of large transient or periodic signals) characterising the seismic noise. The purpose of this approach is manifold.

First of all, the combination of quantification and classification enables one to display the spatial and temporal noise conditions in any frequency band of interest. This information in combination with a common spectral time-frequency analysis can improve the identification of noise sources and their temporal and spatial varying influence. This application of the proposed time series classification is demonstrated by the analysis of the URS data set which is discussed in chapter 5.

Furthermore, the approach is capable to identify some common technical artefacts occurring during (mobile) passive seismic measurements (e.g. data gaps, direct mechanical impacts to the sensor). The classification allows next to the analysis of the seismic noise conditions also an automatic data selection from large data sets for a consecutive seismic noise processing (e.g. H/V, seismic interferometry) by identifying corrupt and/or inappropriate (e.g. dominating periodic or transient signals) time windows of seismic noise. This application of the time series classification is demonstrated by the realisation of a data selection approach for seismic interferometry (see chapter 6).

As highlighted in chapter 2, seismic noise has to be described as superposition of signals emitted by numerous independent, time-varying and spatially distributed sources. Many of the signals themselves are deterministic, but the underlying physics are mostly unknown (source parameters, wave propagation, ...). Therefore, seismic noise should be considered as a temporal and spatial non-stationary random process. In the ideal case of an infinite large number of contributing independent signals, seismic noise in total would be Gaussian distributed according to the Central Limit Theorem (CLT, Bendat & Piersol, 1994). Nevertheless, in reality one has to analyse a finite time series of seismic noise with a finite number of contributing signals. The number and the amplitude distribution of the contributing signals finally control the convergence of the time series towards the Gaussian distribution. Non-Gaussian time series of seismic noise are not an exception due to the occurrence of single dominating signals (e.g. earthquakes, man-made signals) with very large amplitudes in comparison to the remaining background signals. To quantify such seismic noise recordings in the time domain, the character (Gaussian, presence of large transients, ...) of the analysed sample of seismic noise is important additional information.

The classification for seismic noise time series introduced in this chapter provides this additional information about the statistical properties of the seismic noise. The deviations from the Gaussian distribution for the sample value distributions (histograms) from finite digital time series are used for the classification. The time series parameters utilised for the classification are introduced in section 4.1 prior to the time series preparation (section 4.2). The typically observed deviations from the Gaussian distribution in the case of broadband seismic data are discussed in section 4.3. Afterwards the classification scheme is introduced in section 4.4. A test of the noise classification with synthetic data is discussed in section 4.5.

4.1 Time series properties used for quantification and classification

The proposed statistical classification method is based on ratios of amplitude intervals (I), or alternatively percentiles (P), to classify a given time series. The application of statistical

higher order moments like kurtosis and skewness proved to be problematic to reliably determine small deviations from the Gaussian distribution for broadband seismic time series. Main problem in this case is the decreased performance of the common higher moment estimators for time series with sample numbers as large as 1.44 million samples (Hyvärinen et al., 2001). Therefore, specific ratios between percentiles of amplitudes are exploited to classify the seismic noise time series. In the case of a zero mean Gaussian distribution, 68% of the measurements lie within an interval of one standard deviation away from zero (168). 95.45% are within two times the standard deviation (195) and 99.73% are within three times the standard deviation (199) (Figure 4.1). This is also known as the $2\text{-}\sigma$ and $3\text{-}\sigma$, or the “empirical”, rule. The ratios between these intervals, or the corresponding percentiles, of the time series are used to identify and quantify deviations from the Gaussian distribution. First, the ratios σ_2 (195/168), σ_3 (199/168) and the peak factor pf (199/195) are introduced for the time domain classification (Table 4.1). Basic idea is that these ratios increase in the case of a positive kurtosis and decrease in the case of a negative kurtosis of the histogram in comparison to the values 2 (σ_2), 3 (σ_3) and 1.5 (pf) expected for a Gaussian distributed time series. Furthermore, the ratio P84STD between the upper boundary of 168, the 84-percentile (P84), and the standard deviation σ of the time series is utilised to ensure the reliable identification of Gaussian distributed time series. The standard deviation of the time series is calculated with the common statistical estimator for σ (Bendat & Piersol, 1994).

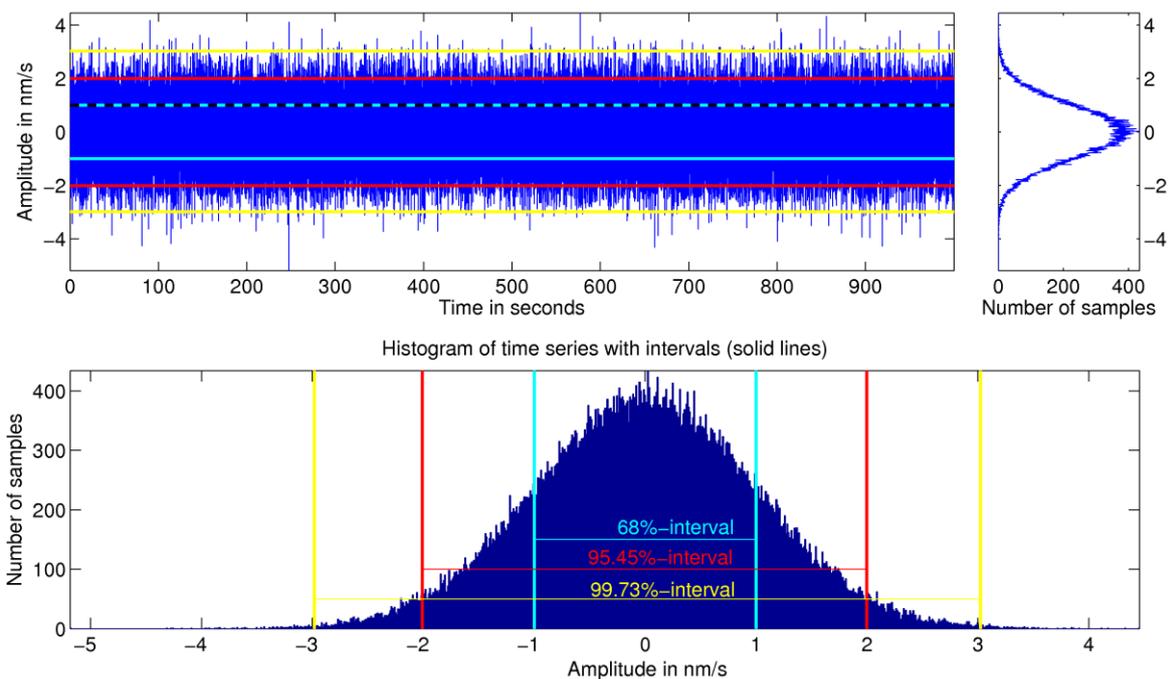


Figure 4.1: Standard Gaussian distributed stochastic process with histogram. Solid lines indicate the margins of the 68% (cyan), 95.45% (red) and 99.73% (yellow) intervals. The dashed black line indicates the standard deviation of the stochastic process. In the case of a symmetric distribution the upper and lower interval margins are symmetric around the mean. In the case of a Gaussian distribution, the standard deviation equals the spread of the 68% interval around the mean. The ratios of the ranges for the 68%, 95.45% and 99.73% intervals are 2, 3 and 1.5 in the case of a Gaussian distribution (empirical rule, see section 4.1).

The ratios σ_2 , σ_3 and P84STD are used to identify the most probably Gaussian distributed time series. In the case of deviations from the Gaussian distribution the peak factor is used as a single measure to quantify the positive kurtosis ($pf > 1.5$) or the negative kurtosis ($pf < 1.5$) of the time series. Similarly, ratios of time series percentiles are used to assess the symmetry of the time series histograms. The ratios of the lower and upper boundaries of the amplitude intervals (e.g. the 16-percentile and the 84-percentile for the 68%-interval) reveal a possible skewness of the histogram and can be used as a

symmetry measure. The ratios SI68 and SI95 are introduced to identify asymmetric time series (Table 4.1). The quantification of the seismic noise time series is realised by the amplitude intervals. In the following the range of the 68%-interval is used for quantification and called 'noise amplitude'.

Parameter		Description	Gaussian
σ_2	I95.45/I68	ratio between the range of the 95.45% and the 68%-interval	2
σ_3	I99.73 /I68	ratio between the range of the 99.73% and the 68%-interval	3
peak factor (pf)	I99.73/I95.45	ratio between the range of the 99.73% and the 95.45%-interval	1.5
P84STD	P84/ σ	ratio between the 84-percentile and the standard deviation of the time series	1
symmetry of I68 (SI68)	P84/P16	ratio between the upper and lower boundary of the 68%-interval	1
symmetry of I95 (SI95)	P97.725/P2.275	ratio between the upper and lower boundary of the 95.45%-interval	1

Table 4.1: The time series parameters used for the classification (with abbreviations). The theoretical values of the parameters in the case of a Gaussian distribution are given in the right column.

By this simple approach, the final classification scheme based on the ratios of percentiles can be easily adjusted to the actual classification problem or noise environment. Typical data problems like data gaps or single extreme transients can be identified by unusual absolute percentiles or percentile ratios. Therefore, this classification approach exhibits also important data quality control abilities for automated processing.

4.2 Preparation of the time series prior to the classification

The time series classification requires a preparation of the seismic noise time series. In accordance to the actual task (analysis or data selection, frequency range of interest, ...) suitable frequency ranges and a time window length have to be selected. As an example, a time window length of 4 hours and 8 frequency ranges are selected based on a spectral time-frequency analysis for the classification analysis of the URS data set. The selection of the time window length and the 8 frequency ranges (see Table 5.2) is discussed in detail in section 5.2. Examples obtained from the URS data set as well as the chosen time window length and frequency ranges are used in the following for the introduction and evaluation of the time series classification.

Prior to the classification means, linear trends and instrument responses are removed from originally five hours long raw broadband time series of seismic noise. A 20-percent cosine window is used to taper the time series during the processing. The broadband time series are filtered forward and reverse in the time domain with a second order Butterworth band pass filter to obtain the seismic noise in the frequency bands of interest. After the band pass filtering the first and last half hour of the time series are cut to obtain 4 hours of seismic noise for the classification without unwanted disturbances of the preprocessing.

4.3 Observed deviations from the Gaussian distribution

The typically observed deviations of seismic noise time series histograms from the Gaussian distribution are illustrated with the help of the URS data set. Figure 4.2 displays

several examples of seismic noise time series (4 hours) measured in the metropolitan area of Bucharest. Together with the time series, the corresponding histograms are shown. For comparison, Gaussian distributions are displayed together with the histograms (green lines). These Gaussian distributions are estimated from the mean and the upper boundary of the 68%-interval (equals the standard deviation in the case of a Gaussian distribution) of the time series. Most of the analysed time series (see chapter 5 for the detailed discussion) exhibit a bell-shaped distribution (Figure 4.2a-c) like the Gaussian distribution (Figure 4.2a).

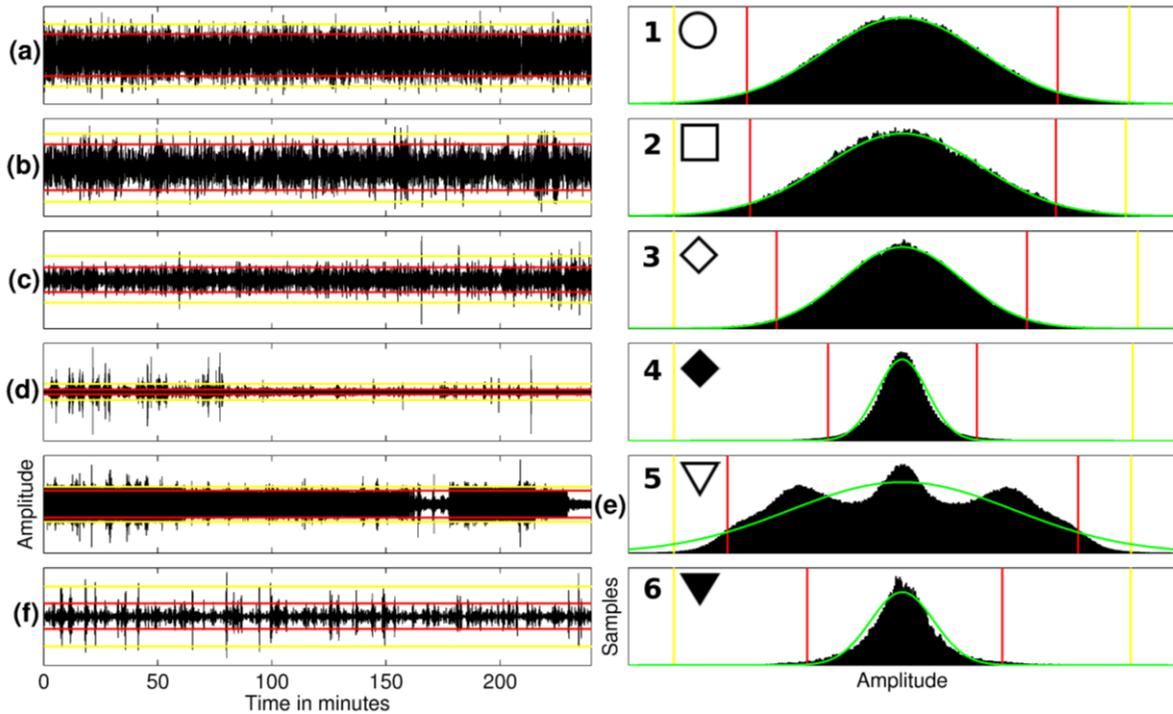


Figure 4.2: Time series of vertical component urban seismic noise with their histograms.

The urban seismic noise was recorded in the Bucharest area during the URS project. The time series (left) and their histograms (distribution of sample values) (right) are plotted together with the best-fitting Gaussian distributions (green lines) estimated from the mean and the upper boundary of the 68%-interval of the corresponding time series. The 95.45- (red) and 99.73- (yellow) amplitude intervals are indicated, corresponding to the $2\text{-}\sigma$ and $3\text{-}\sigma$ range for a Gaussian distributed time series. The noise classification (noise classes 1-6) of the time series is given in the histograms with the corresponding noise class symbols used in following figures (e.g. map of seismic noise amplitudes in Figure 5.3). **(a)** Gaussian distributed time series (NC1, circle) of the USN(0.25-0.6 Hz) at site URS02 with a peak factor of 1.47 **(b)** Nearly Gaussian distributed time series (NC2, square) of the USN(0.04-0.09 Hz) at site URS06 with a peak factor of 1.47. The minor deviations from the Gaussian distribution are small and caused by short transient events with double-amplitudes larger than the range of the 99.73%-interval (left panel). **(c)** Time series of the USN(0.18-0.25 Hz) at site URS04 dominated by short transient noise signals (NC3, diamond) resulting in a peak factor of 1.85. The histogram is slightly deformed at the tails in comparison to the estimated Gaussian distribution. **(d)** Time series of the USN(1-25 Hz) at site URS01 dominated by short transient noise signals (NC4, filled diamond) resulting in a peak factor of 3.1. The histogram is heavily deformed at the tails in comparison to the estimated Gaussian distribution. **(e)** Time series of the USN(1-25 Hz) at site URS20 with a peak factor of 1.3 (NC5, triangle) and a non bell shaped multi-modal distribution. The displayed time series is dominated by sinusoidal signals. **(f)** Time series of the USN(0.04-0.09 Hz) at site URS26 with an asymmetric distribution (NC6, filled triangle) due to the dominance of asymmetric signals.

Nevertheless, most time series exhibit a positive kurtosis (Figure 4.2c and d) due to the presence of short (in comparison to 4 hours) transient signals with large amplitudes (in comparison to the background signal). With an increasing amplitude difference between the 95.45%-interval and the 99.73%-interval the positive kurtosis of the histogram increases. Next to bell-shaped distributions, also multi-modal distributions can be observed (Figure 4.2e). Time series with such multi-modal distributions are most commonly dominated by large-amplitude periodic signals, which sometimes are even

sinusoidal. As a result, the time series histograms are also dominated by the distribution of the dominating signal. Furthermore time series with asymmetric distributions are observed due to the dominance of signals which themselves exhibit asymmetric distributions (Figure 4.2f). The variability of the observed waveforms and histograms highlights the difficulty to quantify (urban) seismic noise in the time domain with one simple amplitude measure like the standard deviation. The major task of the proposed time series classification is to provide additional information to complement the amplitude information. To do so, the classification should be able to differ between Gaussian, symmetric and asymmetric distributed time series. For the most common symmetrically distributed time series the sign and extent of the kurtosis is of interest for describing noise characteristics.

4.4 Noise classification scheme

In total 10 noise classes are introduced based on the identified time series properties. These noise classes (NCs) account for corrupt (NC10-NC13) and non-corrupt (NC1-NC6) time series. All time series with a high probable Gaussian distribution are classified as NC1. If small deviations from the Gaussian distribution occur, the corresponding time series are classified as NC2. Noise classes 3 and 4 account for time series which exhibit distributions with a minor (NC3) or distinct (NC4) positive kurtosis. All time series with a negative kurtosis are classified as NC5. The noise classes 1-5 account for time series with symmetric distributions around their mean. Asymmetric time series are classified as NC6. The classification process itself is a decision tree with several consecutive decision steps (Table 4.2). If the criteria of a decision step are fulfilled, the classification of the analysed time series is finished and no further steps are applied.

4.4.1 Corrupt time series

First, the time series are checked for several criteria of exclusion to identify time series with a high probability for corrupt data. The noise classes 10 and 11 account for time series consisting mainly of zeros. This artefact can occur if missing data is zero padded or if raw time series containing large offsets or offset differences due to technical problems of the sensor are filtered. For the broadband ground motion velocity data of the URS data set a threshold of 3 nm/s for the range of the 68%-interval was chosen which proved to be suitable for all data sets measured with the KARlsruhe BroadBand Array (KABBA). This threshold has to be chosen carefully to stay below the 68%-interval of non-corrupt measurements of seismic noise in all analysed frequency bands. Time series with negative interval ratios are classified as noise class 13. Such negative ratios can also arise if large offset differences due to technical problems (sensor, data acquisition, filtering) occur in the time series. Time series with extreme large amplitudes ($>10^6$ nm/s) due to direct mechanical impacts on the sensor or other technical problems are classified as NC12. Also time series with a range just below the 'clipping amplitude' of 10^6 nm/s but extreme large interval ratios are classified as noise class 12 due to the presence of extreme transients. The criteria for noise class 12 are chosen very conservatively to exclude only extreme and obviously biased time series. This selection is not critical as time series with extreme transients, passing these criteria, are classified reliably as noise class 4 or noise class 6 in the consecutive classification. Furthermore, short but extreme large transients are not critical for quantification, if the 68%-interval range is used. All time series passing the first eight decision steps are treated as non-corrupt time series of seismic noise.

4.4.2 Non-corrupt time series

The first decision step for non-corrupt time series classifies time series which are highly probable Gaussian distributed as NC1 (step 9 in Table 4.2). To do so, the interval ratios σ_2 , and σ_3 , the percentile ratios SI68 and SI95 and the ratio P84STD are applied. To derive the criteria for these parameters 1586 5 hours long time series (4+1 hours, see

section 4.2) of synthetic Gaussian random noise (Marsaglia & Tsang, 2000) are processed to obtain in total 12688 4 hours long time series in all 8 frequency ranges. Afterwards, the resulting 12688 time series are tested with a chi-square goodness of fit test (χ^2 -test) for the null hypothesis that the time series are Gaussian distributed (Bendat & Piersol, 1994). The criteria chosen for NC1 are selected to cover 99% of all synthetic time series identified as Gaussian distributed by the χ^2 -test with a significance level α of 0.05.

Identification of corrupt time series

Step	NC	Comment	σ_2	σ_3	pf	/nm/s I68	/nm/s I100
1	10	zero trace				$<10^{-5}$	
2	11	recorder noise				<3	
3	13	tech. artefacts	≤ 0				
4	13	tech. artefacts		≤ 0			
5	13	tech. artefacts			≤ 0		
6	12	extreme values	>40				
7	12	extreme values		>60			
8	12	clipping					$>10^6$

Classification of symmetric time series

Step	NC	Comment	σ_2	σ_3	pf	P84STD	SI68	SI95
9	1	Gaussian	2 ± 0.05	3 ± 0.15		1 ± 0.01	1 ± 0.015	1 ± 0.015
10	2	nearly Gauss.			1.5 ± 0.1	1 ± 0.06	1 ± 0.03	1 ± 0.047
11	3	pos. kurtosis			$1.5 < pf \leq 2$		1 ± 0.03	1 ± 0.047
12	4	pos. kurtosis			$2 < pf$		1 ± 0.03	1 ± 0.047
13	5	neg. kurtosis			$pf < 1.4$		1 ± 0.03	1 ± 0.047

Identification of asymmetric time series

Step	NC		1-SI68	1-SI95
14	6		>0.03	
15	6			>0.047

Table 4.2: The classification scheme used to classify time series of seismic noise. The scheme consists of 15 consecutive decision steps. Details about the used parameters are shown in Table 4.1. The classification is finished if a time series matches all criteria of a decision step.

From the same test the criteria for the symmetry measures SI68 and SI95 are derived to distinguish between symmetric (NC2-NC5) and asymmetric (NC6) time series. The criteria for symmetric time series (SI68: 1 ± 0.03 ; SI95: 1 ± 0.047) are chosen to contain 97% of the 12688 synthetic band pass filtered time series. All asymmetric time series are classified as noise class 6. The same synthetic data set was used to verify the noise classification (see section 4.5).

Time series that do not agree with the strict criteria for NC1, but showing only small deviations from the Gaussian distribution in terms of the peak factor and the ratio between the 84-percentile and the standard deviation, are classified as NC2 (Table 4.2). Time series classified as NC2 exhibit no clear positive or negative kurtosis, as stated by the small allowed deviations (± 0.1) of the peak factor from 1.5. In the consecutive classification steps the peak factor is used exclusively to distinguish between symmetric time series with positive (NC3 and NC4) and negative (NC5) curved distributions. To display the extent of the positive kurtosis, time series are separated for moderate positive kurtosis ($1.5 \leq pf \leq 2$) and distinct ($2 < pf$) positive kurtosis. The differentiation between time series with moderate and distinct positive kurtosis at a peak factor of 2 is arbitrary and may be adjusted regarding to the noise conditions or the requirements of the following noise analysis. Time series with a peak factor between 1.5 and 2 are observed to exhibit in general no distinct transient signals emerging clearly above the background signal. Therefore time series which are obviously not Gaussian distributed but however not clearly dominated by single transient signals are separated as NC3. Time series with clearly visible transient signals are observed to exhibit in general a peak factor larger than 2 and are classified as NC4.

4.5 Classification of synthetic data

The 12688 4 hours long synthetic time series (1586 Gaussian time series filtered in 8 frequency ranges) are now classified to test the proposed time series classification. The noise class distributions for the synthetic data set in the 8 frequency ranges are given in percent in Table 4.3. Furthermore, the amount of time series accepted (0) and rejected (1) by the chi-square goodness of fit test (χ^2 -test) with the null hypothesis, that the time series are Gaussian distributed, are given in Table 4.3.

In the frequency ranges above 1 Hz all time series (100%) are classified as NC1. Furthermore more than 94% of the time series filtered in the frequency ranges 1-25 Hz and 25-45 Hz are identified as Gaussian distributed (0) by the χ^2 -test. Minor deviations from the Gaussian distribution occur increasingly in the frequency ranges below 0.25 Hz. This effect is indicated by the decreasing amount of time series classified as NC1 from 100% (1-25 Hz) to 5% (0.008-0.04 Hz). Concurrently, the amount of time series identified as Gaussian distributed (0) by the χ^2 -test decreases from 95% to 0%. The difference between the numbers of time series classified as NC1 and identified as Gaussian distributed by the χ^2 -test is significant, especially in the frequency ranges below 0.25 Hz. This demonstrates that the proposed classification is not as strict as the χ^2 -test by identifying time series as Gaussian distributed. This is not critical as the χ^2 -test is known to increasingly reject the null hypothesis for in fact acceptable distributions at large sample sizes (Bull et al., 1992). This disadvantageous high sensitivity occurs for time series with 1.44 million samples. Nevertheless, the increasing occurrence of small deviations from the Gaussian distribution towards lower frequencies is present and resolved by the increasing amount of time series classified as NC2. In the frequency range between 0.04 Hz and 1 Hz more than 98% of the time series are classified as NC1 or NC2. This indicates that the deviations from the Gaussian distribution are rather small. Nevertheless, the power to resolve originally Gaussian distributed seismic noise reliably as NC1 after the filtering is decreased in the frequency ranges below 0.25 Hz. Furthermore, the classification test

reveals significant deviations from the Gaussian distribution in the lowest frequency range 0.008-0.04 Hz. Only 70.9% of the time series are classified as NC1 or NC2, leaving 29.1% with larger deviations from the Gaussian distribution. Most of the time series with larger deviations from the Gaussian distribution are classified as NC6 (24%). This is observed to be due to differences between the absolute upper and lower boundaries of the 95%-interval that are larger than 4.7% (see SI95 in Table 4.2).

Frequency band in Hz	Noise class distribution in %						Hypothesis of statistical test in %	
	NC1	NC2	NC1+NC2	NC3	NC5	NC6	0	1
0.008-0.04	4.9	66	70.9	3.7	1.4	24	0	100
0.04-0.09	35.8	63.1	98.9	0.8		0.3	0	100
0.09-0.18	73	27	100				7.3	92.7
0.18-0.25	77.7	22	99.7	0.3			31	69
0.25-0.6	98.2	1.8	100				72.4	27.6
0.6-1	99.7	0.3	100				85.7	14.3
1-25	100	0	100				94.3	5.7
25-45	100	0	100				95.4	4.6

Table 4.3: Results of the classification test with synthetic Gaussian noise.

The 1586 time series generated with a random number generator were filtered in the 8 used frequency ranges and afterwards classified by the time domain classification. The amount of time series classified as NC1 decreases towards lower frequencies. On the right side of the table the results of a chi-square goodness of fit test (χ^2 -test) are shown. Due to the minor but increasing deviations from the Gaussian distribution the amount of time series rejected by the test is significantly increasing towards lower frequencies. The χ^2 -test is very sensitive to smallest deviations from the theoretical distribution due to the large sample size of the time series with more than 1 million samples.

A second classification test with longer Gaussian distributed time series of 16 hours duration was conducted by the author to investigate this effect. This second synthetic test reveals that more than 99% of the longer time series are classified as NC1 or NC2 in all frequency ranges. This result indicates that the occurrence of larger deviations in the frequency range 0.008-0.04 Hz is most probably related to the shorter time series length of 4 hours. Nevertheless, an increasing occurrence of small deviations from the Gaussian distribution (NC2) towards lower frequencies is also observed with the longer time series.

Concluding, the noise class distributions for time series of 4 hours are biased in the lowest frequency range 0.008-0.04 Hz. The increased occurrence of larger deviations from the Gaussian distribution is related to the length of the time series and may be furthermore influenced by the filtering. In the frequency range between 0.04 Hz and 0.25 Hz differences between time series classified as NC1 and NC2 are less significant than for the higher frequency ranges and should not be used for an interpretation. Therefore, the amount of time series classified as NC1 or NC2 is additionally given in the noise class distributions in Table 4.3 (synthetic data) and the tables for measured data in chapter 5. Nevertheless, the rather small deviations from the Gaussian distribution of time series classified as NC2 are acceptable in terms of the quantification and can be assumed to be acceptable for most applications of seismic noise like H/V or seismic interferometry.

4.6 Summary of chapter 4

The time series classification introduced in this chapter uses ratios between time series amplitude intervals as well as time series percentiles to identify deviations of a time series distribution from the Gaussian distribution with high sensitivity. The ratio between the 99.73% amplitude interval and the 95.45% amplitude interval is introduced as the quantity *peakfactor*. The peakfactor (pf) equals 1.5 in the case of a Gaussian distributed time series and increases/decreases with increasing/decreasing kurtosis of the time series.

The time series classification is capable to identify different types of corrupt time series (e.g. technical problems with the sensor). Regarding the non-corrupt time series six noise classes are introduced to classify the typically observed deviations of seismic noise time series from the Gaussian distribution. Gaussian distributed time series are classified as noise class 1 (NC1). Non-Gaussian but symmetric time series are classified as NC2-NC5. Time series which exhibit determinable but rather small and unspecific deviations from the Gaussian distribution ($pf \ 1.5 \pm 0.1$) are classified as noise class 2 (NC2). Time series with a gentle peaked histogram in comparison to the Gaussian distribution ($1.6 < pf \leq 2$) 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 of the time series as noise class 4 (NC4). Symmetric time series with a flattened histogram in comparison to the Gaussian distribution ($pf < 1.4$) are classified as noise class 5 (NC5). All time series which are not identified as symmetric time series are classified as noise class 6.

The influence of the applied band pass filters and the time window length on the time series classification is evaluated with a data set of synthetic time series. A time window length of at least 200 times the longest contained period is recommended to ensure an unbiased statistical time domain analysis.

The time series classification is applied in the next chapter to analyse the statistical properties of the urban seismic noise in Bucharest. In chapter 6 the time series classification is used to introduce an automated data selection to the calculation of seismic noise cross-correlation functions for the estimation of Green's functions.

5 Analysis of urban seismic noise

The efforts which are undertaken concurrently in seismology to turn ambient seismic noise into signal help to overcome obstacles hampering established methods of active source and passive earthquake seismology (see chapter 2). At the same time the social and economical importance of seismic hazard assessment and mitigation in (mega)cities is rapidly increasing due to the exploding urbanisation, especially close to major fault systems (United Nations, 2006; Montgomery, 2008). Site effect analysis, wave propagation scenarios and early warning concepts are high-priority issues for such urban regions. Therefore, the number of passive seismic measurements, such as the URS project in the metropolitan area of Bucharest (section 3.1), increases with the aim to record ambient seismic noise in urban environments (e.g. Milana et al., 1996; Scherbaum et al., 2003; Fäh et al., 2008). These recordings can provide the required information for seismic hazard assessment. Another important task affected by urban seismic noise is the seismological monitoring of geothermal power plants which are increasingly installed in densely populated areas to be economically successful (Giardini, 2009; Groos & Ritter, 2010).

Seismic noise in cities is a complex superposition of seismic signals emitted by man-made (e.g. traffic, industry) and natural (wind-, ocean wave- or earthquake-induced tremor) seismic sources (Groos & Ritter, 2009). The high temporal and spatial variability of the urban seismic noise (USN) is a fact which must be addressed by all methods utilising seismic noise (see section 2.3) or hampered by seismic noise (e.g. monitoring of induced micro-earthquakes). A good knowledge of the utilised seismic noise wave field is necessary in every individual case as recalled and demanded also by Bonnefoy-Claudet et al. (2006a).

In this chapter a detailed discussion of the broad-band USN in the metropolitan area of Bucharest is provided which improves also the understanding of the urban seismic noise in general. The data set of the URS project is a rare opportunity to analyse the urban seismic noise continuously in a broad frequency range (0.008-45 Hz) and over a long time (9 months). Typical seismic measurements in large cities are short campaigns (hours to days) with short-period sensors (> 1 Hz) to obtain information about the underground close to the surface using seismic noise array techniques (section 2.3.2) or the H/V technique (2.3.1).

The temporal variability and the typical sources of the USN in Bucharest are discussed in section 5.1 of this chapter with the help of a time-frequency analysis. Purpose of the time-frequency analysis is the identification of suitable frequency bands and time windows for the consecutive analysis with the time series classification introduced in chapter 4. The frequency bands and time windows as well as the selection of 11 working days for the further analysis are discussed in section 5.2. The temporal and spatial variability of the vertical-component USN in Bucharest is discussed in detail with the help of the selected 11 working days in section 5.3. The purpose of the selection is to analyse predominantly the USN caused by sources in the metropolitan area with a focus on man-made sources. Therefore working days are selected which are not affected by global seismicity and as far as possible not by storm-induced ocean-generated microseism originating from the Mediterranean and Black Sea. The analysis of the complete URS data set with the time series classification including the horizontal component USN is discussed afterwards in section 5.4. The analysis of such a large and complex noise classification data set by an analyst involves a considerable effort. The next following step is therefore the analysis of an obtained seismic noise classification data set with machine learning and pattern recognition techniques to support the human analyst. A first feasibility study with the Self-Organizing Map (SOM) technique, which is a neural network technique, is presented in section 5.5. The chapter is concluded with a short summary in section 5.6.

5.1 Time-Frequency analysis

To identify the frequency-dependent behaviour of the time-variable sources of the seismic noise long-term spectrograms are calculated for time-series of up to 28 days duration (Figure 5.1a and b). The spectrograms are used for the preceding frequency-time analysis to determine reasonable frequency bands and time windows for the time domain classification and quantification. The long-term spectrograms are calculated from continuous seismic time series sampled at 100 Hz. Before filtering means and linear trends are removed. All data are filtered with a 4th order Butterworth band pass filter in the time domain between the fundamental frequency of the seismometer (0.008, 0.01, 0.03 and 0.2 Hz respectively, see section 3.1) and the Nyquist frequency (50 Hz) of the recorder system. Actually, the spectrograms are limited to an upper frequency of 45 Hz due to the steep cut-off around 45 Hz caused by the decimation filter of the data-logger.

After filtering power spectral density (psd) spectra for consecutive time windows with a length of 480 s are computed using the adaptive multiple taper method (Percival & Walden, 1998) with a time-bandwidth product (NW) of 3. The effective frequency resolution of the obtained spectra is 0.0065 Hz. To furthermore reduce the noise of the psd estimates and to reduce the amount of data points 4 consecutive psd spectra are averaged for the 28 days long-term spectrograms resulting in a temporal resolution of 32 minutes for the spectrograms. For the shorter spectrogram of 7 days (Figure 5.1c) 2 consecutive psd spectra are averaged to obtain a temporal resolution of 16 minutes. Every averaged psd spectrum is normalized with the maximum power spectral density of all psd spectra belonging to the complete spectrogram.

To avoid confusion between seismic waves induced by earthquakes and signals induced by other noise sources, expected P-wave onset times for maybe detectable earthquake waves are calculated using iasp91 (Kennett & Engdahl, 1991) and marked in the spectrograms as stars. To identify all earthquakes potentially observable at the URS network, the NEIC (USGS, <http://neic.usgs.gov>) and ROMPLUS (Oncescu et al., 1999) catalogues are analysed for distant (distance $\Delta > 30^\circ$), regional ($30^\circ \geq \Delta > 5^\circ$) and local ($5^\circ \geq \Delta$, Vrancea subduction zone) earthquakes, respectively. All catalogued regional and local earthquakes are considered. The selection of potential time windows with teleseismic earthquake waves is done by conservatively estimated distance-magnitude relations based on the experience from previous experiments and a cross-check with the actual URS recordings (Table 5.1). In total 972 local, regional and distant earthquakes have to be considered, and their onset times are marked by red (local) and green (regional and distant) stars on the time axes of the spectrograms.

$\Delta / ^\circ$		Catalogue	Magnitude M_w	Events #
	$\Delta \leq 5$	Romplus	all	59
5 <	$\Delta \leq 30$	NEIC	all	38
30 <	$\Delta \leq 100$	NEIC	> 4.9	680
100 <	Δ	NEIC	> 5.4	195

Table 5.1: Selection criteria for earthquakes which may be observed at the URS network. These criteria are applied to the catalogues ROMPLUS (Oncescu et al., 1999 and NIEP (<http://www.infp.ro/eqsinfo.php>)) for local earthquakes in the Vrancea subduction zone and NEIC (<http://neic.usgs.gov>) for regional and teleseismic earthquakes between 2003-Oct-01 and 2004-Sep-01. Δ is distance in degree.

The frequency-time analysis for the ‘microtremor’ (>1 Hz, see section 2.2) is discussed with Figure 5.1a in section 5.1.1. The lower frequencies are discussed with Figure 5.1b and c in the sections 5.1.2 (transitional range, 0.6-1 Hz) and 5.1.3 (‘microseism’, <0.6 Hz). In Figure 5.1c the discussed influences of the dominating natural sources on the seismic

noise below 2 Hz are marked with numbers. Based on the frequency-time analysis several frequency bands and time windows per day are selected for the statistical time domain analysis which are discussed in section 5.2.

5.1.1 Urban seismic noise above 1 Hz ('microtremor')

The long-term spectrogram of the vertical ground motion velocity in Figure 5.1a reveals a regular pattern of high (red to yellow) psd at daytimes and low (green to blue) psd at nighttimes in the city centre at site URS12. The psd difference between day and night in the frequency range 1-25 Hz varies from 2-25 dB at the station sites in the metropolitan area and is 11 dB in spatial average. The nightly psd lows of ~5 hours are much shorter than the daily highs of 19 hours. This distinct daily pattern demonstrates the dominant influence of human activity on the urban seismic noise in the frequency range 1-25 Hz. Besides this daily pattern, the USN also contains a weekly pattern at most locations in the metropolitan area. Working days with high psd can be distinguished from weekends with reduced USN. The weekends are marked with bars at the top in Figure 5.1a. In the frequency range above 1 Hz traffic-induced seismic waves are excited by road traffic (Hao & Ang, 1998; Coward et al., 2003) and trains (Fiala et al., 2007; Chen et al., 2004) or can be generated by traffic induced bridge oscillations (Chen et al., 2007).

These contribute significantly to USN in a broad frequency range from ~1 Hz to more than 45 Hz with maximum amplitudes between 1-10 Hz. Long lasting and very narrow-band signals above 1 Hz, recognised as horizontal lines of increased psd in the spectrograms (e.g. 16.7 Hz) are sinusoidal-type seismic waves most probably excited by rotating machinery at sharp frequencies. Examples are electrical motors and gear boxes of industrial machinery (Plesinger & Wielandt, 1974; Bokelmann & Baisch, 1999; Kar & Mohanty, 2006), power generators and building services machinery (Coward et al., 2005). Due to gear boxes and frequency converters such sinusoidal signals can be observed in the whole frequency range from 1 Hz up to the power frequency (50 Hz) but predominantly around 12.5 Hz (8 poles engines), 16.67 Hz (6 poles), 25 Hz (4 poles) and 50 Hz (2 poles). A significant drop of psd by up to 20 dB can be observed above 25 Hz towards higher frequencies at most sites in the Bucharest area. Therefore, two frequency ranges (1-25 Hz and 25-45 Hz) are chosen for the statistical analysis to get a deeper insight into the temporal and spatial changes of amplitude and statistical properties of the high-frequency USN.

5.1.2 Urban seismic noise 0.6-1 Hz (natural and man-made sources)

In contrast to the dominating man-made USN above 1 Hz, USN becomes more complicated towards frequencies below 1 Hz due to overlapping man-made and natural signals. In the frequency range 0.6-1 Hz both man-made and natural sources contribute significantly to the USN in Bucharest. One of the most powerful noise sources observed in the frequency range 0.6-1 Hz are seismic waves generated by local wind (Withers et al., 1996; Ritter & Groos, 2007) and wind-induced oscillations of buildings and structures (Ward & Crawford, 1966). In Figure 5.1c the correlation between the noise psd at station site URS21 in the outskirts of Bucharest and the wind velocity is displayed. For wind velocities exceeding 3-4 m/s increased noise amplitudes are observed between 0.6 Hz and 1.2 Hz. The same increase of noise amplitudes can be observed in the inner city area (compare Figure 5.1c with the last week in Figure 5.1b). At the same time, the influence of man-made seismic noise can still be observed due to the regular pattern of higher psd at daytimes and lower psd at nighttimes down to 0.6 Hz. Therefore the frequency range 0.6-1 Hz is chosen for the time domain analysis to cover the transition between microtremor and microseism and the frequency band dominantly influenced by wind-induced microseism.

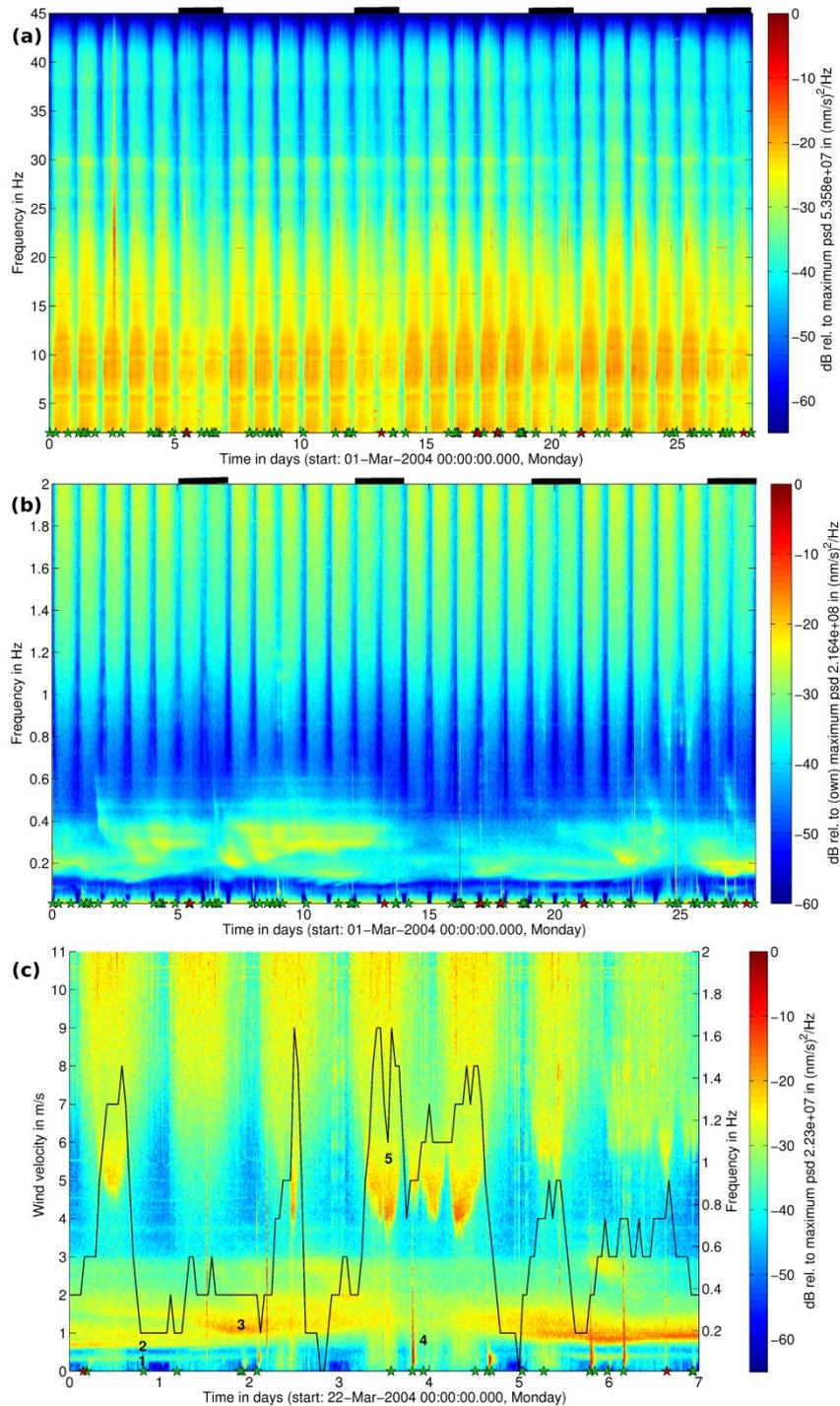


Figure 5.1: Spectrograms of the vertical-component USN in Bucharest during March 2004. **(a)** The spectrogram displays the psd in decibel of the USN (2-45 Hz) at station site URS12 in the city centre of Bucharest in the first 4 weeks of March 2004. The observed USN at this site is typical for the inner city area of Bucharest. Stars at the bottom indicate onset times of world-wide (green stars) and local (red stars) earthquakes that potentially could be observed in Bucharest. Weekends are marked with bars on the top. Periodic signals excited by electrical machines with discrete frequencies can be recognised as horizontal lines at distinct frequencies (e.g. 16.7 Hz and 30 Hz). The effective frequency resolution of the spectrogram is 0.0065 Hz and the temporal resolution is 32 minutes. **(b)** Same Spectrogram as in (a), but the lower frequency range 0.008-2 Hz is displayed. **(c)** Correlation of wind with the vertical-component urban seismic noise. The spectrogram displays the USN (0.008-2 Hz) at station site URS21 in the southern outskirts of Bucharest in the last week of March shown also in (a) and (b). The hourly averaged wind velocity in the metropolitan area of Bucharest is shown as overlay. Selected events are outlined **1**: primary ocean-generated microseism originating in the North Atlantic Ocean **2**: secondary ocean-generated microseism originating in the North Atlantic Ocean **3**: ocean-generated microseism originating in the Mediterranean and the Black Sea **4**: seismic waves excited by earthquakes **5**: wind induced seismic noise.

5.1.3 Urban seismic noise below 0.6 Hz ('microseism')

The daily pattern vanishes in the frequency band 0.09-0.6 Hz due to the dominance of the ocean-generated microseism (Figure 5.1b and c). Temporal changes of the psd reaching about 20 dB in the frequency range 0.09-0.6 Hz are not unusual. Numerous of such events with complex frequency-dependence were observed during the URS experiment. These events persist from several hours up to several days and are observed concurrently at all sites in the metropolitan area. Such dynamic microseismic events in the frequency range 0.09-0.6 Hz, independent from daytime and observed in near-coastal regions, are well known to be secondary (double-frequency) ocean-generated microseism (Bromirski & Duennebier, 2002; Bonnefoy-Claudet et al., 2006a; Bromirski et al., 1999). To confirm this explanation array processing (FK analysis, see section 2.3.2) is applied at the URS stations for several of the observed microseism events. In the frequency band 0.18-0.6 Hz Rayleigh waves are observed which approach from several recurring backazimuth (BAZ) ranges corresponding to the nearby Black Sea (minimum distance 200 km, BAZ 60°-150°), Marmara Sea (minimum distance 450 km, BAZ 150°-170°) and Mediterranean Sea (minimum distance 450 km, BAZ 170°-280°). The FK analysis for the microseismic event on March 23rd (marked with 3 in Figure 5.1c) is shown in Figure 5.2b. The waves approach from a BAZ of ~200° and with a slowness of ~73 s/° typically for short-period Rayleigh waves in sediments. Further narrowband microseism events (0.09-0.18 Hz, marked with 2) can be observed at lower frequencies next to the microseismic events above 0.18 Hz, originating from the Mediterranean and Black Sea. The FK analysis reveals these microseismic events as intermediate-period (slowness ~30 s/°) Rayleigh waves approaching from north-western and northern directions with BAZ between 270-50° (Figure 5.2a). Very often narrowband microseismic events with half the frequency (0.04-0.09 Hz, marked with 1) but the same BAZ can be observed, too. The most probable source of these microseism events are primary and secondary (double-frequency) ocean-generated microseism generated by the swell of distant but very strong storms (Bromirski & Duennebier, 2002) over the North Atlantic Ocean (NAO) and the northern seas.

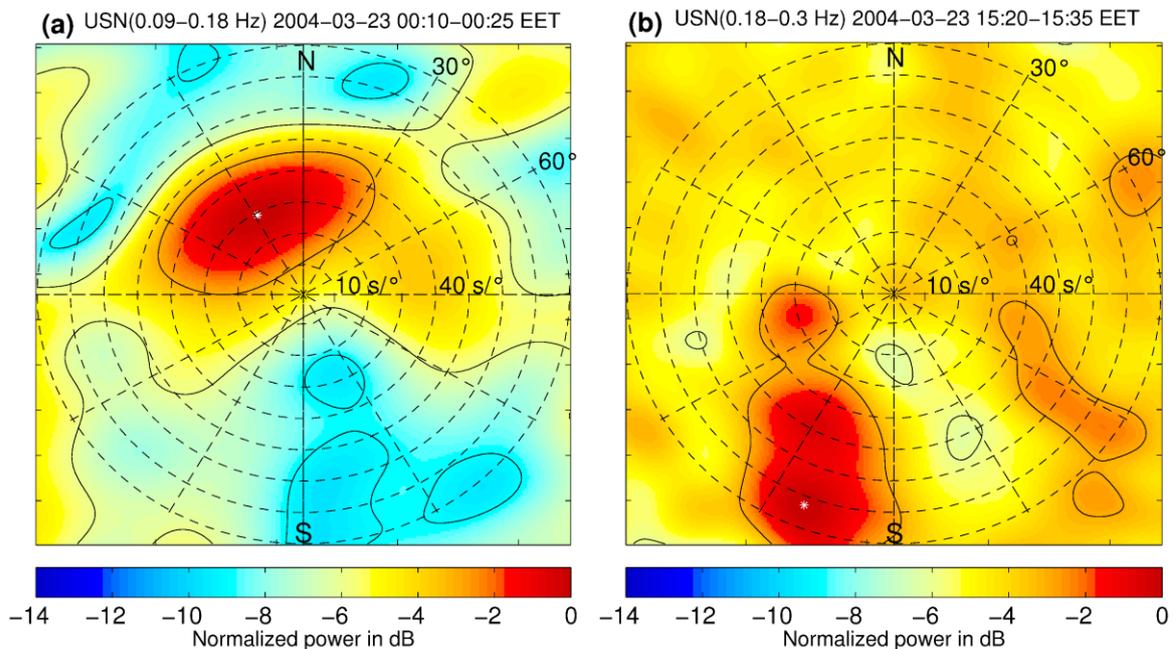


Figure 5.2: Frequency-wave number (FK) analysis of vertical-component USN in Bucharest. The FK analysis is applied to 15 minutes of seismic noise in the frequency ranges (a) 0.09-0.18 Hz and (b) 0.18-0.3 Hz. In the frequency range 0.09-0.18 Hz Rayleigh waves approaching from NNW directions (BAZ ~330°) with a slowness of ~30 s/° are observed. These Rayleigh waves are most probably secondary ocean-generated microseism originating from the northern seas. In the frequency range 0.18-0.3 Hz Rayleigh waves approaching from SSW directions (BAZ ~200°) with a slowness of ~73 s/° are observed. These Rayleigh waves are most probably secondary ocean-generated microseism originating from the Mediterranean seas.

Earthquake waves contribute to the USN in Bucharest in the observable frequency range from 8 mHz to ~25 Hz. Below 0.09 Hz they are the most powerful seismic energy in Bucharest (marked with 4). Surprisingly, also large variations with daytime can be observed at frequencies below 0.04 Hz at several inner city sites in Bucharest as shown for site URS12 in Figure 5.1b. The USN in Bucharest below 0.09 Hz is a dynamic superposition of seismic waves generated by the oceans, wind, earthquakes and man-made signals.

5.2 Analysed frequency bands and time windows

Based on the analysis of the long-term spectrograms finally 8 different frequency ranges are chosen and summarised with their specific dominating noise sources in Table 5.2. The different frequency ranges between 0.04 Hz and 0.6 Hz are selected to reflect the frequency bands of the distant primary (0.04-0.09 Hz) and secondary (0.09-0.18 Hz) as well as the regional (0.18-0.6 Hz) ocean-generated microseism. The frequency band of the regional ocean-generated microseism originating from the Mediterranean and Black Sea is divided into the two frequency ranges 0.18-0.25 Hz and 0.25-0.6 Hz to distinguish between the permanent and the storm-induced ocean-generated microseism.

Frequency /Hz	seismometer types	Dominating noise sources
0.008 - 0.04	STS-2	wind, human activity, earthquakes
0.04 - 0.09	STS-2, G40T/3ESP	ocean-generated microseism (Atlantic Ocean), human activity, wind, earthquakes
0.09 - 0.18	STS-2, KS-2000, G40T/3ESP	ocean-generated microseism (Atlantic Ocean) (Mediterranean and Black Sea)
0.18 - 0.25	STS-2, KS-2000, G40T/3ESP	ocean-generated microseism (Atlantic Ocean) (Mediterranean and Black Sea)
0.25 - 0.6	all	ocean-generated microseism (Mediterranean, Marmara and Black Sea)
0.6 - 1	all	wind, human activity
1 - 25	all	human activity (traffic, industry, ...)
25 - 45	all	human activity (traffic, industry, ...)

Table 5.2: Frequency ranges for the statistical time-series analysis of urban seismic noise.

Depending on the frequency range, recordings from different seismometers are used and different noise sources dominate the urban seismic noise in Bucharest.

A time window length of 4 hours is selected to cover the time window with the lowest constant noise conditions at nighttime which is observable from the time-frequency analysis between 0-4 local time (EET). Furthermore, every day can be divided into six consecutive 4 hours time windows covering night (0-4 EET), morning (4-8 EET), the working hours (8-12, 12-16, 16-20 EET) and the evening (20-24 EET) for the analysis of the complete data set (see sections 5.4 and 5.5). For the analysis of the selected working days (see below and section 5.3) only three daily 4 hours time windows are selected between 0-4, 8-12 and 13-17 EET. This is a trade-off to be able to compare the USN at nighttime and in the working hours, to analyse the consistency of the USN during the working hours and to find several working days which are not affected by earthquake waves, strong storm-induced ocean-generated microseism as well as strong wind in all of the three daily time windows.

The influence of the applied band pass filters on the statistical properties of the seismic noise time series increases with decreasing lower corner frequency and decreasing time window length (see details in section 4.5). At a time window length of 4 hours the classification of the time series in the lowest frequency band 0.008-0.04 Hz is already biased by the filtering. A time window length of at least 200 times the longest contained period is recommendable to ensure an unbiased statistical time domain analysis. A considerable and reasonable shortening of the time window length is possible, if the lowest frequency band 0.008-0.04 Hz would be dismissed. Nevertheless, the time window length of 4 hours is selected to cover the whole span of time with the lowest constant

noise conditions at night and the frequency band 0.008-0.04 Hz is included in the discussion of the selected working days in section 5.3.

As a first step the vertical-component USN is analysed at 11 working days distributed over the time span (~285 days) of the URS data set (Table 5.3). Earthquakes and ocean-generated microseism are well known sources of seismic energy (see chapter 2) whose influence on the USN is illustrated very well by the time-frequency analysis (Figure 5.1). The task of the first noise classification analysis is to reveal the temporal and spatial variability of the USN due to local sources in the metropolitan area and with a focus on man-made sources. Therefore working days are selected which are not affected by seismic waves excited by local, regional or teleseismic earthquakes (see Table 5.1) as well as storm-induced ocean-generated microseism originating from the Mediterranean Sea and the Black Sea. Nine of these working days exhibit average wind velocities below 3 m/s in all three daily time windows. Two more working days are selected for comparison with moderately increased wind velocities in the forenoon hours (8-12 EET) and very large wind velocities in the afternoon hours (13-17 EET). These days are included to analyse the influence of the wind induced seismic noise.

In total 1709 consecutive four hour time windows between 27th October 2003 00:00 EET and 7th August 2004 00:00 EET are processed for the analysis of the complete URS data set in sections 5.4 and 5.5. The time window 28th March 2004 00-04 EET is excluded from the analysis due to the change from winter to summer time in this time window.

Wind velocities (min/max/mean) in m/s				
Date	Weekday	Night 00-04 EET	Morning 08-12 EET	Afternoon 13-17 EET
Working days, average to low wind velocities				
2003-12-02	Tuesday	2/3/2.8	2/2/2	1/2/1.2
2003-12-15	Monday	2/4/3	2/3/2.5	1/2/1.5
2004-02-02	Monday	2/2/2	0/2/1.5	1/3/2
2004-02-03	Tuesday	2/4/2.5	1/2/1.2	1/2/1.5
2004-02-19	Thursday	1/1/1	0/1/0.5	2/2/2
2004-04-01	Thursday	0/1/0.5	0/2/1	1/2/1.5
2004-06-23	Wednesday	1/3/1.5	1/2/1.5	1/1/1
2004-06-24	Thursday	1/1/1	1/1/1	1/2/1.8
2004-07-23	Friday	0/0/0	0/0/0	0/1/1.5
Working days, high wind velocities				
2004-03-25	Wednesday	2/2/2	2/5/3.8	5/9/6.8
2004-05-07	Friday	1/2/1.8	0/5/2.8	7/8/7.8

Table 5.3: All working days selected for the first noise classification analysis. Additionally, the daily time windows (Eastern European Time, EET) selected for the noise classification are given with the corresponding minimum, maximum and mean wind velocities (Institutul National de Meteorologie, Hidrologie INMH, Romania www.inmh.ro). The wind velocities are given in m/s.

5.3 Analysis of selected working days

The results of the time series classification of the vertical-component USN at the 11 selected working days with low and increased wind velocities are presented in the following including a detailed discussion of the spatial variability of the USN in Bucharest. The noise class distributions in the 8 selected frequency bands (Table 5.2) for the 9 working days with low wind conditions (in total 5656 time series) are shown in Table 5.4 and discussed in detail in the sections 5.3.1 to 5.3.3.

Noise class distribution in %

Freq. band /Hz	Time (EET)	#	NC1	NC2	NC1 NC2	NC3	NC4	NC5	NC6	NC11-13
0.008-0.04	0-4	162	0.62	5.56	6.17	20.37	16.05	1.23	52.47	3.70
0.008-0.04	8-12	161	0	8.07	8.07	13.66	10.56	0.62	55.90	11.18
0.008-0.04	13-17	162	0	6.17	6.17	8.02	6.79	0.62	70.99	7.41
0.04-0.09	0-4	186	11.29	39.78	51.08	20.43	24.73	0	1.61	2.15
0.04-0.09	8-12	185	4.32	20.00	24.32	23.24	37.84	0	6.49	8.11
0.04-0.09	13-17	186	2.69	19.89	22.58	25.27	43.01	0	4.84	4.30
0.09-0.18	0-4	249	52.61	32.93	85.54	7.63	4.82	0	0.80	1.20
0.09-0.18	8-12	247	31.58	21.05	52.63	16.19	23.08	0	0.81	7.29
0.09-0.18	13-17	248	32.26	25.81	58.06	14.52	22.98	0	0	4.44
0.18-0.25	0-4	249	71.89	23.69	95.58	2.01	1.20	0	0	1.20
0.18-0.25	8-12	247	51.82	21.86	73.68	12.55	6.48	0	0.40	6.88
0.18-0.25	13-17	248	57.26	21.37	78.63	9.68	7.26	0	0	4.44
0.25-0.6	0-4	261	80.08	14.94	95.02	1.92	1.92	0	0	1.15
0.25-0.6	8-12	259	66.80	5.41	72.20	11.58	10.42	0	0.39	5.41
0.25-0.6	13-17	260	65.00	14.23	79.23	11.15	5.77	0	0	3.85
0.6-1	0-4	261	16.09	20.31	36.40	54.79	7.66	0	0	1.15
0.6-1	8-12	259	38.22	22.78	61.00	27.80	6.18	0	0	5.02
0.6-1	13-17	260	35.00	26.54	61.54	27.31	7.69	0	0	3.46
1-25	0-4	261	0	0.77	0.77	11.88	83.14	1.15	1.92	1.15
1-25	8-12	259	0	0.39	0.39	38.22	54.44	0.39	0.39	6.18
1-25	13-17	260	0	0	0	35.00	60.77	0.00	0.00	4.23
25-45	0-4	261	1.53	6.90	8.43	18.01	65.52	3.83	1.92	2.30
25-45	8-12	259	0	0.39	0.39	10.42	81.47	0.77	3.09	3.86
25-45	13-17	260	0	0	0	10.77	82.69	0.77	2.69	3.08

Table 5.4: The noise class distributions for the vertical-comp. USN in Bucharest (selected working days)

The table displays the NC distributions in the 8 frequency ranges (Table 5.2) in the night (0-4 EET), forenoon (8-12 EET) and afternoon (13-17 EET) hours of the 9 working days with low to moderate wind velocities (Table 5.3). The noise class distributions are given in percent related to the absolute number (third column, #) of analysed time series in the corresponding time window and frequency range.

The approach to classify the seismic noise time series with 6 noise classes proved to be reasonable. Around 42% of the analysed time series are classified as NC1 or NC2 and can supposed to be Gaussian or nearly Gaussian distributed (Table 5.4). The most common larger deviations from the Gaussian distribution are a minor (NC3) or distinct (NC4) positive kurtosis due to transient signals accounting for 47% of the analysed time series. Less than 7% are classified as NC5 or NC6 due to a negative kurtosis and

asymmetric amplitude intervals respectively. In total 4% of the time series are classified as corrupt. The amount of time series classified as corrupt (NC11-NC13) differs in the different frequency ranges due to the frequency dependent influence of single extreme transients.

In all frequency bands above 0.04 Hz systematic differences between the noise class distributions of day- and nighttime are observed and discussed below. The results obtained from the morning (08-12 EET) and afternoon (13-17 EET) hours are quite similar and summarised as 'daytime' results in the following. Most Gaussian distributed time series (NC1 and NC2) are observed in the frequency bands dominated by ocean-generated microseism between 0.09-0.6 Hz. The amount of time series with larger deviations from the Gaussian distribution increases towards lower and higher frequencies.

5.3.1 Microtremor (1-45 Hz)

In Figure 5.3 the classification and noise amplitudes (ranges of the 68%-intervals) of the USN (1-25 Hz) (top) and USN (25-45 Hz) (bottom) are displayed for daytime (left) and nighttime (right). Discrete transients (NC3 and NC4, diamonds in Figure 5.3) dominate the USN above 1 Hz at day- and nighttime (Table 5.4). Due to these most probably man-made transient noise signals the Central Limit Theorem (CLT) is not applicable to the USN (1-45 Hz) at the inner city sites. Furthermore, a small amount of time series (<4%) with a negative kurtosis (NC5) can be observed for the USN (1-45 Hz) especially at nighttime due to dominating periodic signals. This observation can be explained with 2- to 8-pole electrical motors operating at frequencies above 10 Hz which are widely used in industry and housing. But also distinct differences between the noise class distributions for the USN (1-25 Hz) and USN (25-45 Hz) can be observed. The total amount of time series with a distinct positive kurtosis (NC4) reduces from nighttime to daytime at 1-25 Hz. This behaviour indicates that the convergence to the Gaussian distribution is enhanced at daytime due to the numerous overlapping noise signals and their amplitude distribution. The decreased number of transients at nighttime results in lower noise amplitudes but a more distinct positive kurtosis of most time series. The inverse effect is observed at 25-45 Hz. The influence of transient signals increases from night- to daytime in contrast to USN(1-25 Hz). This is indicated by a higher amount of time series classified as NC4 at daytime and the observation of a few time series (~7%) with only small deviations from the Gaussian distribution (NC2) at nighttime. Also the amount of asymmetric time series (NC6) increases from night- to daytime for the USN(25-45 Hz), indicating that few time series are dominated by asymmetric signals.

The noise amplitudes in the frequency ranges 1-25 Hz and 25-45 Hz increase towards the city centre at day- and nighttime (Figure 5.3). The difference between the rural areas outside the city (URS06) and the city centre (URS12) reaches up to 16 dB at daytime and 30 dB at nighttime. Inside the inner city area a heterogeneous spatial distribution of the highest noise amplitudes at day- and nighttime is observed. The amplitude increase corresponds in principle very well with the increasing population density towards the lively city centre (Figure 3.1) with increased car, bus and tram traffic. However, at some sites distinct higher amplitudes are observed. Except the city centre sites (URS09, URS12, URS29) all sites with high amplitudes (URS32, URS25, URS13, URS02, URS04) are in vicinity (<500 m) of busy heavy industry areas. At station sites in vicinity of industrial areas the concentration of higher amplitudes seems more pronounced for the USN (25-45 Hz) than for the USN(1-25 Hz). This observation might be biased due to the higher attenuation in the frequency range 25-45 Hz compared to lower frequencies. Therefore, it can be expected that preferably high-frequency signals excited nearby the instruments are observed in the frequency range 25-45 Hz. Especially the time series of USN (25-45 Hz) classified as NC5 or NC6 may be dominated by man-made signals excited inside the buildings where the instruments were located (e.g. air conditioner, heating systems, water pumps, footsteps, ...).

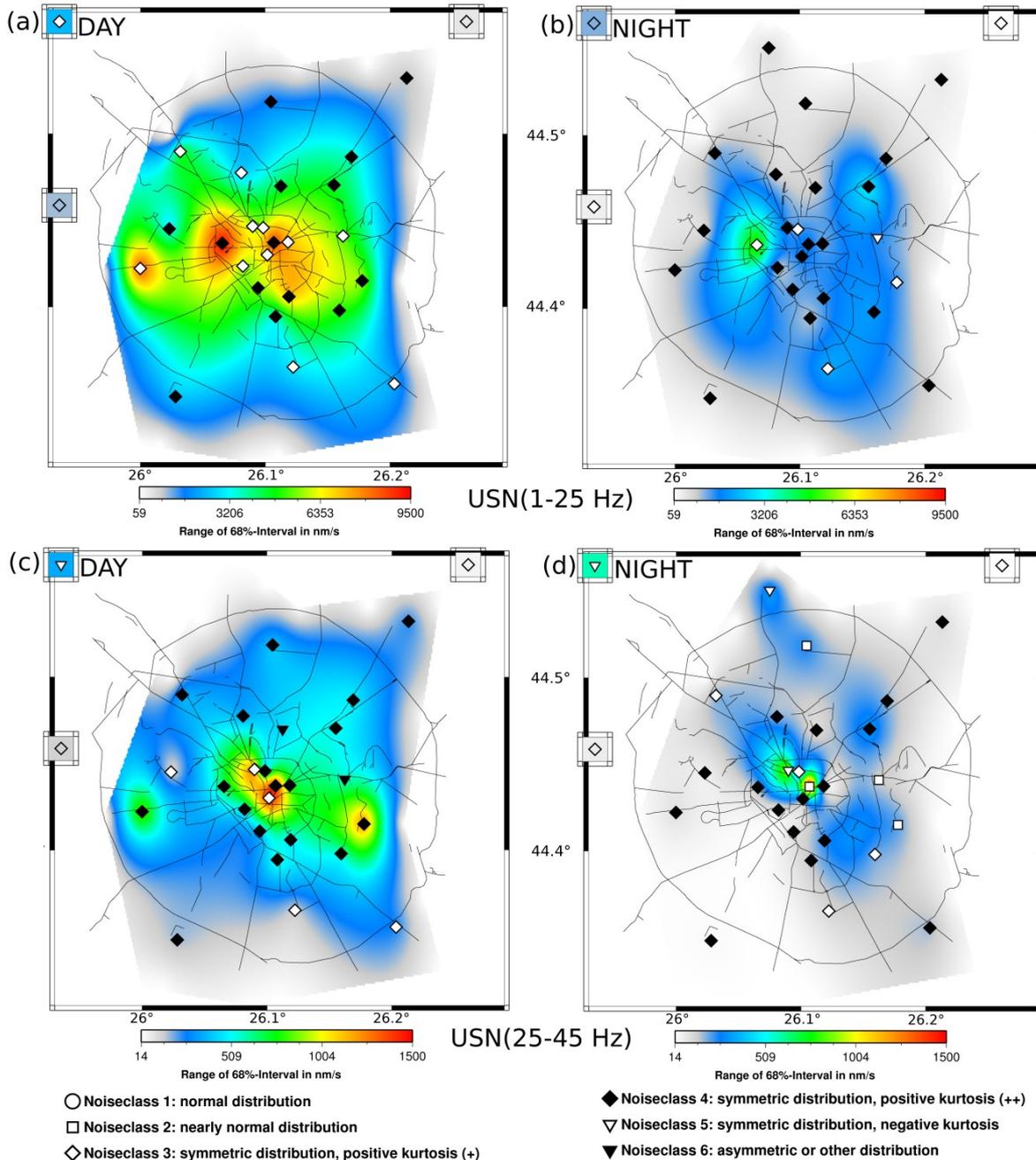


Figure 5.3: Maps with seismic noise amplitudes (colour) and classes (symbols) at the URS station sites. Displayed are the noise amplitudes (range of the 68%-interval) of the ground motion velocity in nm/s in the frequency range 1-25 Hz (top) on 2004-Feb-03. (a) daytime (13-17 EET, 1-25 Hz). (b) nighttime (00-04 EET, 1-25 Hz) and in the frequency range 25-45 Hz (bottom) on 2004-Feb-03. (c) daytime (13-17 EET, 25-45 Hz). (d) nighttime (00-04 EET, 25-45 Hz). Stations URS06, URS17 and URS19 (outside the city area, see Figure 3.1) are displayed in separate boxes at the edges. Surface gridding was done by GMT (www.gmt.soest.hawaii.edu) with the continuous curvature surface gridding algorithm (Smith & Wessel, 1990) under a tension of 0.1. This interpolation preserves measured amplitude values at the station sites. The lower limit of the interpolation output is 90 percent of the smallest measured noise amplitude. The upper limit is the largest measured noise amplitude. Between station sites interpolation effects may occur due to partly large inter-station differences in amplitude. Real noise amplitudes cannot be derived between station sites. These maps are mainly visualisation tools to analyse large amounts of output data and search for temporal and spatial variations.

5.3.2 Transitional range (0.6-1 Hz)

The USN (0.6-1 Hz) exhibits distinct day/night differences (average 5-8 dB, Figure 5.4 and Figure 5.5) and workday/weekend differences (average 3-4 dB, Figure 5.1b) on the vertical-component in the city centre. During daytime most time series are classified as NC1 or NC2 (~60%), indicating that the Central Limit Theorem (CLT) is applicable for 52

these time series due to numerous overlapping noise signals (Table 5.4). All time series with larger deviations from the Gaussian distribution exhibit a minor (NC3) or distinct (NC4) positive kurtosis. The amount of time series classified as NC1 or NC2 decreases significantly from ~60% at daytime to ~36% at nighttime. At the same time the amount of time series classified as NC3 increases from day- to nighttime. As for higher frequencies above 1 Hz, there are fewer signals contributing to USN at night. The superposition of a lower amount of different signals at night exhibits lower average noise amplitudes and is predominantly non-Gaussian distributed due to few strong transient signals emerging from the in general lower noise amplitudes in the metropolitan area.

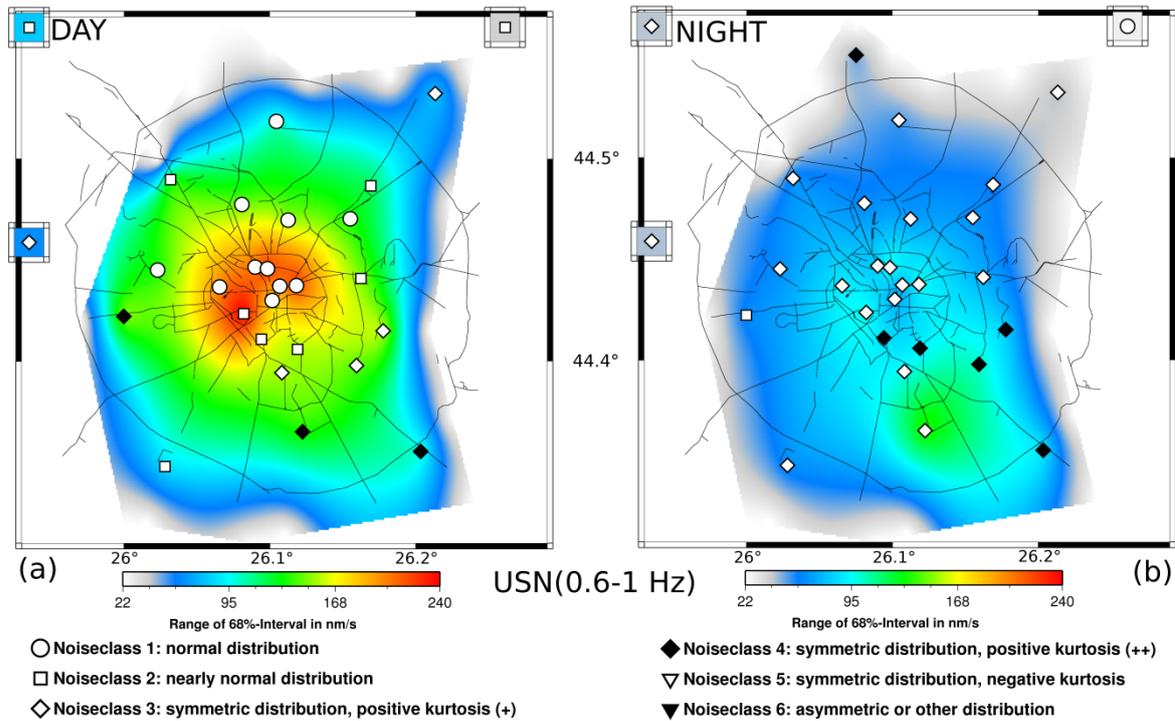


Figure 5.4: Maps with seismic noise amplitudes and classes at the URS station sites (0.6-1 Hz). Displayed are the noise amplitudes (range of the 68%-interval) of the ground motion velocity in nm/s in the frequency range 0.6-1 Hz on 2004-Feb-03. (a) daytime (13-17 EET). (b) nighttime (00-04 EET).

During working days with low wind conditions (wind velocity <3m/s, e.g. 2004-Feb-03) the USN (0.6-1 Hz) amplitudes predominantly increase towards the inner city area. The amplitude behaviour displayed in Figure 5.4 demonstrates the higher level of man-made USN during the day in the inner city area (mean noise amplitudes of about 150-240 nm/s) versus the quieter period at night (about 80-100 nm/s). Excepted for this overall trend, the strongest man-made signals are often recorded in the vicinity of a busy industrial area south-east of the city centre at station site URS23.

At all stations of the URS network (e.g. URS12 and URS21 in Figure 5.5) unusually large USN(0.6-1 Hz) amplitudes are observed during the afternoon (13-17 EET) hours on March 25th and May 7th. This observation correlates well with the increased wind velocities larger than 5 m/s.

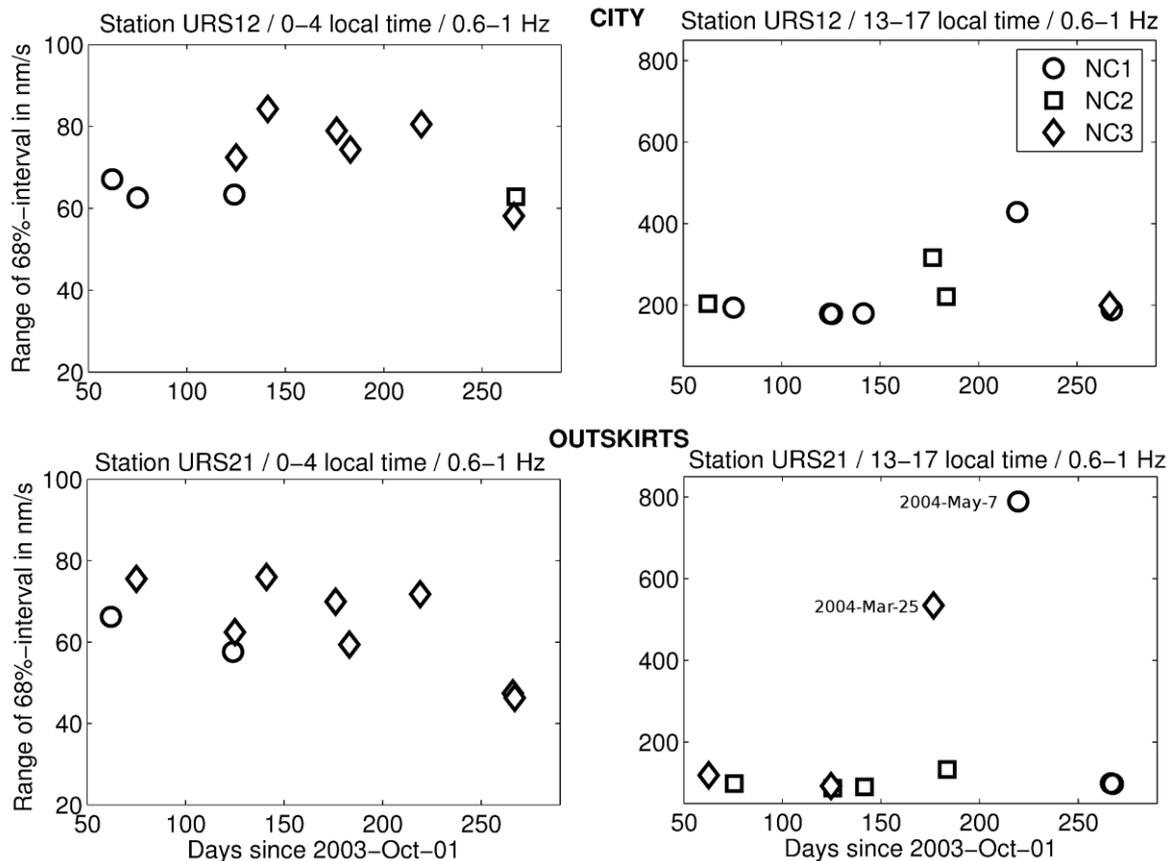


Figure 5.5: Seismic noise amplitudes and classes in the frequency range 0.6-1 Hz. Displayed are the noise amplitudes (range of 68%-interval) of the vertical ground motion velocity in nm/s for the 11 analysed days at nighttime (00-04 EET) (left) and daytime (13-17 EET) (right) for the two stations sites URS12 (top) and URS21 (bottom). The x-coordinate is in days after 2003-Oct-01, the start of the URS experiment. The noise classes are indicated by symbols (NC1 (circle), NC2 (square), NC3 (white diamond), NC4 (black diamond), NC5 (white triangle), NC6 (black triangle)).

5.3.3 Microseism (0.04-0.6 Hz)

Most of the analysed time series in the frequency ranges between 0.04 Hz and 0.6 Hz are dominated by ocean-generated microseism and classified as NC1 or NC2 showing only small deviations from the Gaussian distribution. Nevertheless, significant variations of the noise class distributions with daytime and frequency occur. Regarding the noise classification two major trends are observed in the frequency ranges between 0.04 Hz and 0.6 Hz. First, the total amount of time series classified as NC1 or NC2 decreases from night- to daytime (Table 5.4). Second, the occurrence of larger deviations from the Gaussian distribution increases towards lower frequencies.

Regarding the noise amplitudes no systematic amplitude differences between day- and nighttime are observed in the frequency range 0.09-0.6 Hz. As an example the noise amplitudes at the inner city site URS12 are given in Figure 5.6. At low frequencies (0.04-0.09 Hz) larger amplitude differences with daytime occur in the metropolitan area compared to the higher frequency ranges between 0.09 Hz and 0.6 Hz. Furthermore a trend towards a seasonal decrease of noise amplitudes is observed from winter to the summer months. This effect is possibly caused by the calmer weather conditions at the surrounding seas during summer.

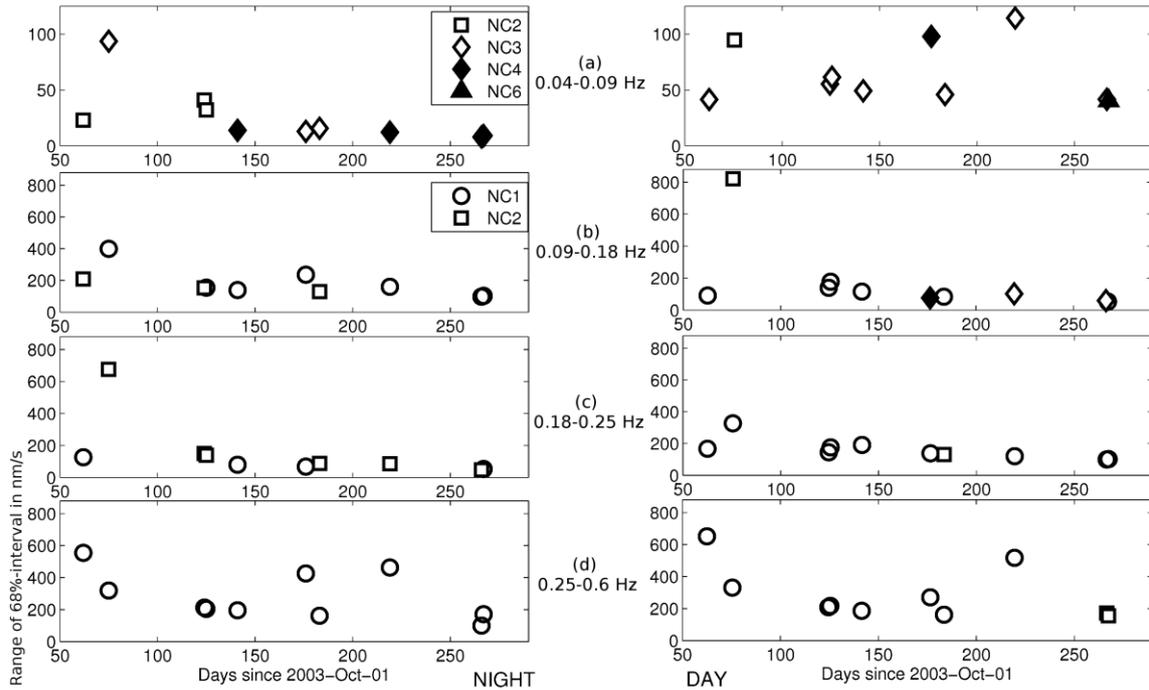


Figure 5.6: Seismic noise amplitudes and classes for the vertical-component USN (0.04-0.6 Hz). Displayed are the noise amplitudes (range of 68%-interval) in several frequency ranges at station site URS12 in the city centre at nighttime (00-04 EET) (left) and daytime (13-17 EET) (right). (a) USN(0.04-0.09 Hz). (b) USN(0.09-0.18 Hz). (c) USN(0.18-0.25 Hz). (d) USN(0.25-0.6 Hz).

5.3.3.1 Frequency range 0.25-0.6 Hz

The low human influence on USN (0.25-0.6 Hz) is shown by slight day/night differences of less than 3 dB at the URS sites on average (Figure 5.6d and Figure 5.7). The USN (0.25-0.6 Hz) noise amplitudes at nighttime vary between 600 nm/s and 100 nm/s from winter to summer in the inner city area. These dominant long-term temporal variations of noise amplitude are observed concurrently at all sites in the metropolitan area. The long-period (several days) changes of noise amplitude without pronounced variations with daytime indicate the dominance of large scale (e.g. meteorological) processes on the USN(0.25-0.6 Hz) as observed from the frequency-time analysis. Spatial amplitude variations of the USN(0.25-0.6 Hz) within the metropolitan area are observed but less systematic and pronounced as for the higher frequencies above 0.6 Hz. Time series with deviations from the Gaussian distribution are predominantly observed at sites in the inner city area of Bucharest at daytime indicating a man-made influence on the USN (0.25-0.6 Hz). Larger deviations from the Gaussian distribution (NC3) are also observed in time windows with wind velocities larger than 5 m/s (Figure 5.7). Concurrently, the spatial variations of noise amplitudes in the metropolitan area increase significantly. Both observations indicate a spatially varying influence of wind on the USN (0.25-0.6 Hz) in the metropolitan area which deserves further analysis.

5.3.3.2 Frequency range 0.09-0.25 Hz

The temporal variations of noise amplitudes and statistical properties of the USN are very similar in the frequency ranges 0.09-0.18 Hz and 0.18-0.25 Hz. At nighttime 85-95% of the analysed time series are classified as NC1 or NC2 at working days with low wind conditions (Table 5.4). Towards daytime the amount of time series with larger deviations from the Gaussian distribution due to transient signals (NC3 and NC4) increases significantly, especially in both frequency bands. At daytime the amount of time series classified as NC3 or NC4 equals ~37% and ~17%, respectively. The long term variations of noise amplitudes are similar at all sites in the metropolitan area. Generally no significant changes of noise amplitude with daytime are observed at most sites in the metropolitan area (Figure 5.8a and b). Nevertheless, distinct variations of noise amplitude

with daytime are observed erratically at sites predominantly in the inner city area. As for the higher frequency range 0.6-1 Hz the occurrence of unusual large noise amplitudes and larger deviations from the Gaussian distribution are often observed simultaneously.

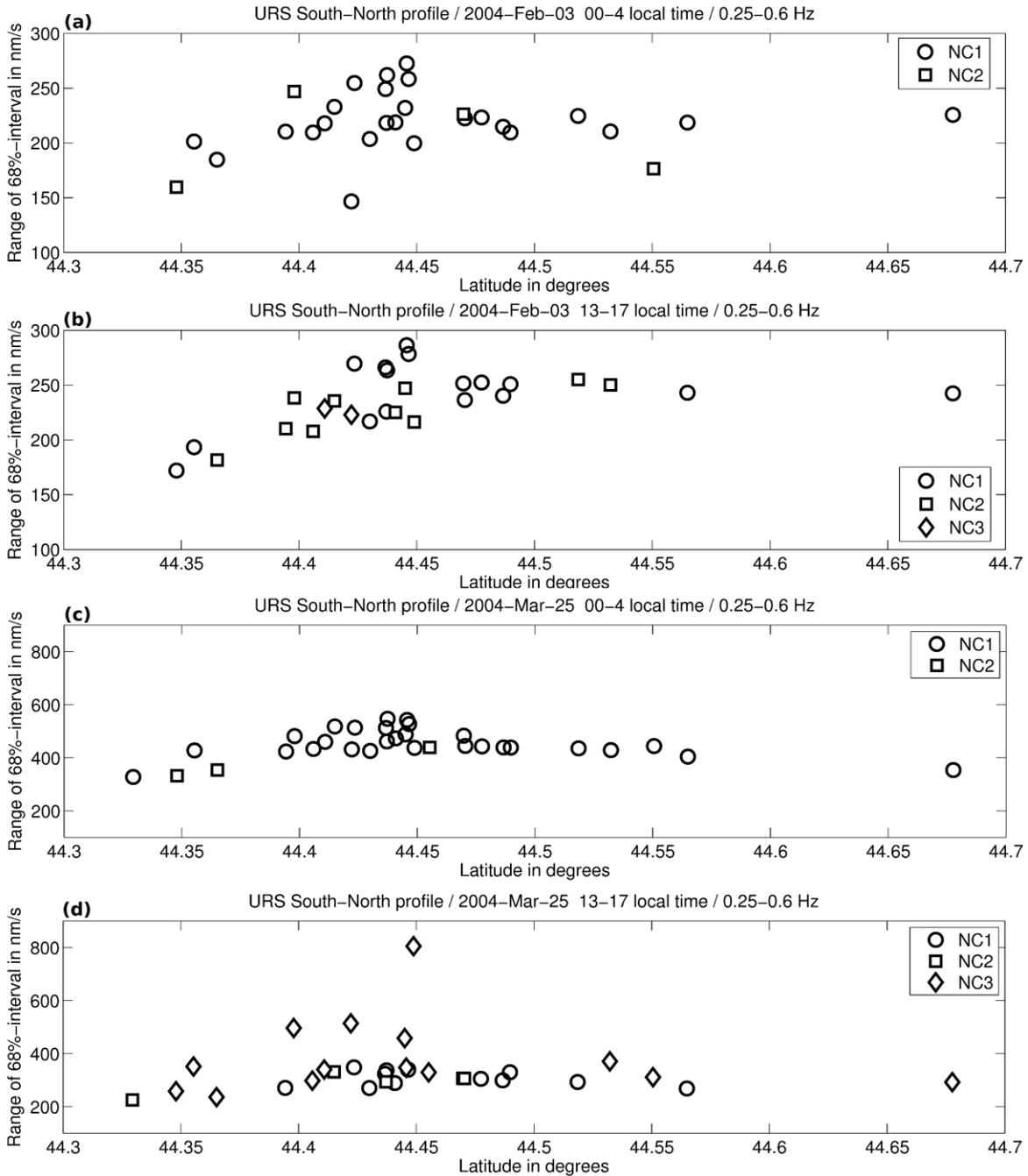


Figure 5.7: South-North profiles of the seismic noise amplitudes and classes (0.25-0.6 Hz). The seismic noise amplitudes (range of 68%-interval) and noise classification (symbols) in the frequency range 0.25-0.6 Hz at day and night are displayed for two days with different wind conditions. (a) 2004-Feb-03 at nighttime (00-04 EET) with a low average wind velocity of 2.5 m/s. (b) 2004-Feb-03 at daytime (13-17 EET) with a low average wind velocity of 1.5 m/s. (c) 2004-Mar-25 at nighttime (00-04 EET) with a low average wind velocity of 2 m/s. (d) 2004-Mar-25 at daytime (13-17 EET) with a high average wind velocity of 6.8 m/s.

Regarding the noise amplitudes a trend of decreasing noise amplitudes from the North towards the South is observed at day- and nighttime and all analysed days in the frequency ranges between 0.09-0.25 Hz (Figure 5.8a+b). An amplitude and site effect study by Sudhaus and Ritter (2009), using waveforms of the URS data set, found the same effect of latitude-dependent ground motion amplification across the Bucharest area. Sudhaus and Ritter (2009) relate this effect to resonance effects in the unconsolidated

sediments above the distinct dipping Neogene-Cretaceous boundary from 800 m depth in the South to 2000 m depth in the North (Mandrescu et al., 2004; Sèbe et al., 2009). Also for the URS data set Ziehm (2006) applied a comprehensive spectral H/V analysis for the Bucharest area. She observed a distinct and stable peak in the H/V ratio in the frequency range 0.09-0.25 Hz. This peak frequency changes smoothly from 0.25 Hz in the South of Bucharest to 0.09 Hz in the North. A more detailed comparison of the spectral H/V ratio observed by Ziehm (2006) with the time domain H/V ratio obtained from the noise classification is discussed in section 5.4.6.

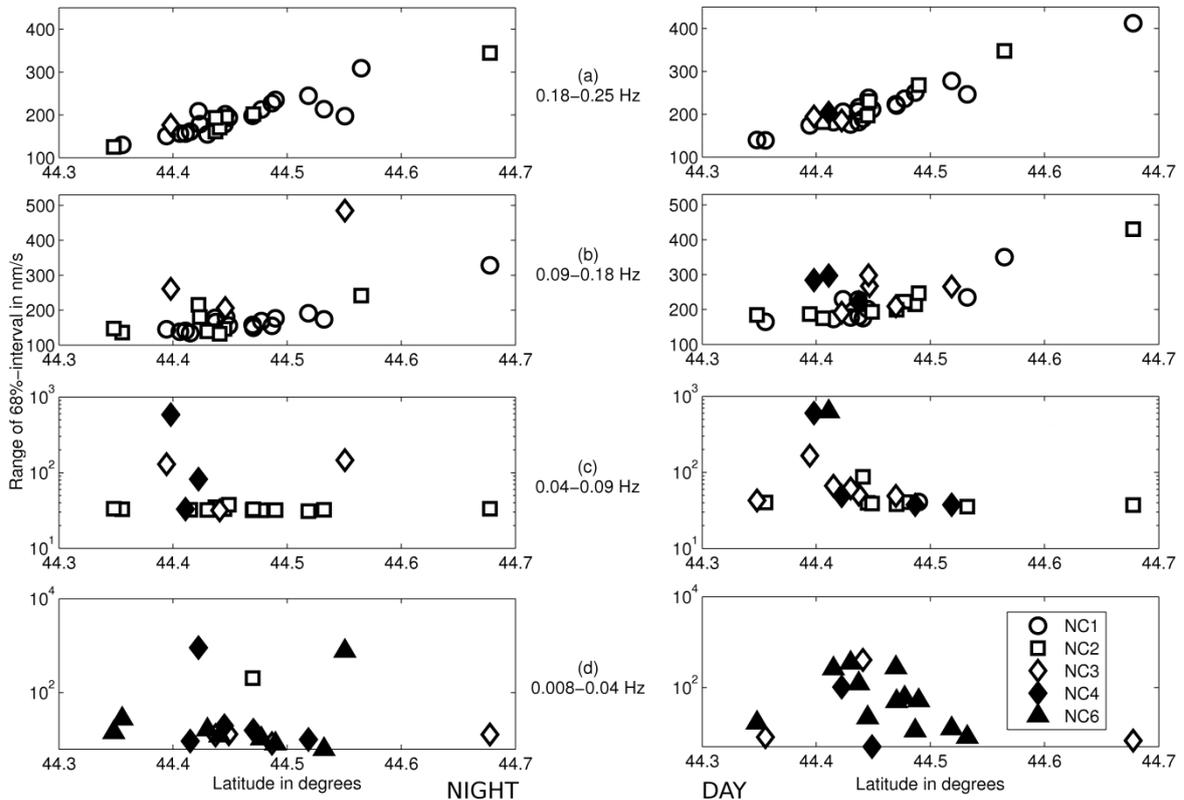


Figure 5.8: South-North profiles of the seismic noise amplitudes and classes (0.008-0.25 Hz). The seismic noise amplitudes (range of 68%-interval) and noise classification (symbols) on 2004-Feb-03 at nighttime (00-04 EET) (left) and daytime (13-17 EET) (right) are displayed for several frequency ranges. (a) 0.18-0.25 Hz. (b) 0.09-0.18 Hz. (c) 0.04-0.09 Hz. (d) 0.008-0.04 Hz.

5.3.3.3 Frequency range 0.04-0.09 Hz

The long-term temporal pattern of noise amplitudes of the USN (0.04-0.09 Hz) at nighttime is similar to the noise amplitudes of the USN (0.09-0.18 Hz) indicating a similar composition of contributing sources at nighttime (Figure 5.6). Around 50% of the analysed time series at nighttime are classified as NC1 or NC2 (Table 5.4). Most common deviations from the Gaussian distribution are a minor (NC3) or distinct (NC4) positive kurtosis due to transient signals. The occurrence of time series with larger deviations from the Gaussian distribution increases from ~45% at nighttime to more than 60% at daytime. Also an increasing amount (up to 6%) of time series classified as NC6 occurs especially at daytime. This effect cannot be explained with filtering effects occurring for non-corrupt time series (see section 4.5). A manual analysis of the waveforms revealed that time series classified as NC6 in the frequency range 0.04-0.09 Hz exhibit short transients with large amplitudes and asymmetric distributions by far most cases as the example shown in Figure 4.2f.

Variations of noise amplitude with daytime are observed sporadically at several stations in the metropolitan area (Figure 5.6 and Figure 5.8c). Noise amplitudes and their temporal variations are less systematic than at higher frequencies and differ strongly between the station sites. Nevertheless, large noise amplitudes, noise amplitude variations with

daytime and deviations from the Gaussian distribution are observed predominantly in the inner city area. These observations indicate a spatially strongly varying human influence on the USN (0.04-0.09 Hz). In the analysed time windows at the 9 working days chosen for this first analysis no significant global or local seismicity is observed. Nevertheless, earthquake induced seismic waves are a significant source of seismic energy in the frequency range 0.04-0.09 Hz (see 5.1.3). The influence of the global seismicity on the USN conditions is also discussed in sections 5.4.2 and 5.5.3.

5.3.4 Frequency range 0.008-0.04 Hz

The noise class distributions of the USN (0.008-0.04 Hz) at daytime (morning and afternoon hours) are less stable than for the higher frequency bands (Table 5.4). Furthermore, no systematic differences in the noise class distributions between day- and nighttime are observed. Less than 10% of the analysed time series are classified as NC1 or NC2. More than 50% of the time series are classified as NC6. The remaining time series exhibit predominantly a positive kurtosis and are classified as NC3 or NC4. The extent of time series classified as NC6 is not explainable by the filter effects alone. As for the higher frequency band 0.04-0.09 Hz most time series classified as NC6 are dominated by transients with large amplitudes and exhibit distinct asymmetric distributions.

Furthermore the noise amplitudes and their variations with daytime differ tremendously in the metropolitan area of Bucharest, especially in the inner city area (Figure 5.8d). These strong spatial variations are pointing towards local sources in the vicinity of the affected sites. In fact, the spatial distribution of large noise amplitudes is well correlated with zones of heavy industry connected to the railway network (compare Figure 5.9 and Figure 3.1). Next to various production sites of heavy industries, accelerating and decelerating trains (Karlström, 2006) can be assumed to induce such very low frequency vibrations. Furthermore, slow movements of heavy masses and their induced ground tilt (Forbriger, 2007) are probable sources of this very located low-frequency USN. Also possible is tilting of high-rise buildings under wind load or tilting due to decoupled basement floors acting as a membrane under air pressure load (Beauduin et al., 1996). To reliably identify such sources of USN a further analysis involving the horizontal components of USN (0.008-0.04 Hz) and correlations with corresponding parameters like wind velocity and air pressure directly at the station sites are necessary. Unfortunately, these parameters are not available for the URS project.

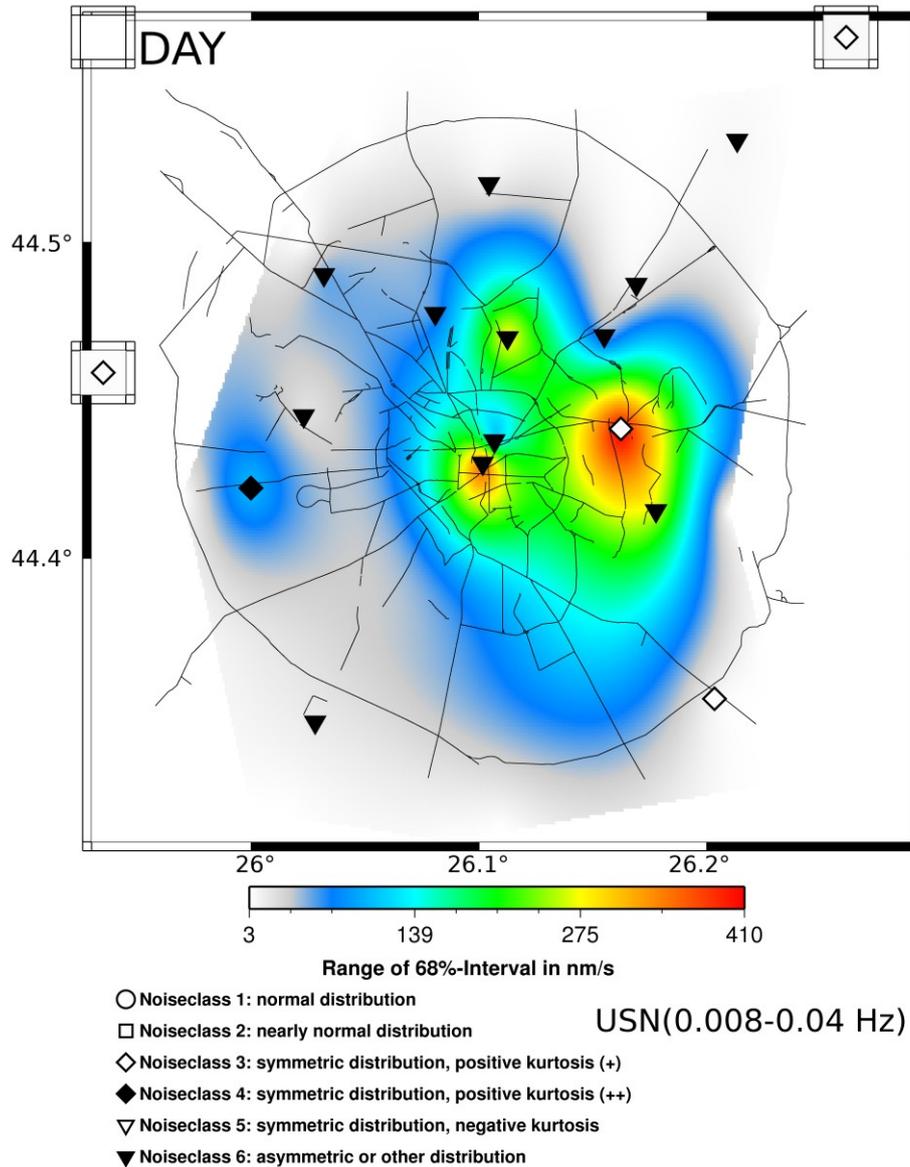


Figure 5.9: Map of seismic noise amplitudes and classes in Bucharest (0.008-0.04 Hz). Displayed are the noise amplitudes (range of the 68%-interval) of the ground motion velocity in nm/s in the frequency range 0.008-0.04 Hz at day 2004-Feb-03 during daytime (13-17 EET) at the URS station sites.

5.4 Analysis of the complete URS data set

The results of the complete analysis of the URS data set with consecutive 4 hour time windows (0-4, 4-8, 12-16, 16-20 and 20-24 EET, see section 5.2) are discussed in comparison and as extension of the detailed discussion of the selected working days above. The seismic noise amplitudes represented again by the ranges of the 68% amplitude intervals are discussed exemplary with the help of the USN (0.6-1 Hz) in section 5.4.1. The results obtained from all working days with low wind conditions but without the exclusion of the seismic waves excited by earthquakes or the ocean-generated microseism are discussed in comparison to the selected working days in section 5.4.2. Based on the results of the complete data set also a comparison of the USN statistical properties at working days and Sundays (with low wind conditions) can be provided (section 5.4.3) as well as a comparison of working days with low and high wind conditions (section 5.4.4). The discussions in sections 5.4.1 to 5.4.4 are limited without disadvantage to the results obtained for the vertical component USN. The differences between the statistical properties of the vertical and horizontal USN are discussed in section 5.4.5. The amplitude ratio between the vertical and horizontal USN (H/V ratio) is discussed and compared with the results of Ziehm (2006) in section 5.4.6.

5.4.1 General discussion of the seismic noise amplitudes

The analysis of the complete URS data set also confirms the observations of the time-frequency analysis in section 5.1. Regarding the man-made contributions to the USN the seismic noise amplitudes at nighttimes (0-4 EET) on Sundays are the lower boundary and the amplitudes at daytimes (8-16 EET) on working days are the upper boundary of the typically observed seismic noise amplitudes as discussed in detail in section 5.3. The seismic noise amplitudes in the time windows 4-8, 16-20, 20-24 EET are in general between these lower and upper amplitude boundaries with similar amplitudes in the morning (4-8) and evening (20-24) hours. The seismic noise amplitudes at the late afternoon/early evening (16-20) are in general higher than in the later evening and morning hours and slightly lower than in the main working hours (8-16 EET). These observations are especially true for the frequency range above 0.6 Hz.

Variations of the seismic noise outside of this man-made temporal pattern are in general caused by the natural sources of seismic noise. As discussed above, local wind is observed as the dominating natural source of seismic noise in the frequency range of 0.6-1 Hz. In the frequency bands below 0.6 Hz the ocean-generated microseism as well as seismic waves excited by earthquakes cause the largest seismic noise amplitudes as discussed in section 5.1. The same temporal and frequency dependent behaviour of the noise amplitudes is observed for the horizontal components. The H/V amplitude ratio is discussed in section 5.4.6.

5.4.2 Comparison of all / selected working days (low wind velocities)

In this section the results obtained from the nine selected working days (Table 5.3) with low wind conditions, low regional ocean-generated microseism and without seismic waves excited by earthquakes (see sections 5.2 and 5.3) are compared with the results obtained from all 156 working days with low wind conditions (average wind velocity < 3 m/s). Main differences between both data sets are therefore the influence of the global seismicity and the storm induced regional ocean-generated microseism on the USN in the Bucharest area. The noise class distributions obtained from both data sets are presented in Table 5.5 for the time windows 0-4 EET and 8-12 EET. The noise classes 1 and 2 as well as 3 and 4 are summarised and noise classes 5 and 6 (together less than 10%) are excluded to improve the readability of the table. The complete results are given in Appendix C.

Freq. band /Hz	EET	Working days, low wind conditions, vertical component			
		selected days (9)		all days (156)	
		NC1+2	NC3+4	NC1+2	NC3+4
		%	%	%	%
0.04-0.09	0-4	51.08	45.16	24.22	73.33
0.04-0.09	8-12	24.32	61.08	10.75	83.87
0.09-0.18	0-4	85.54	12.45	74.85	24.52
0.09-0.18	8-12	52.63	39.27	57.57	42.15
0.18-0.25	0-4	95.58	3.21	88.46	11.50
0.18-0.25	8-12	73.68	19.03	82.39	17.59
0.25-0.6	0-4	95.02	3.84	84.21	15.77
0.25-0.6	8-12	72.20	22.00	82.24	17.69
0.6-1	0-4	36.40	62.45	34.70	65.24
0.6-1	8-12	61.00	33.98	70.31	29.69
1-25	0-4	0.77	95.02	1.17	96.69
1-25	8-12	0.39	92.66	0.71	98.56
25-45	0-4	8.43	84.53	6.96	88.07
25-45	8-12	0.39	91.89	0.45	95.54

Table 5.5: Comparison of the noise class distributions for selected / all working days (low wind conditions)
The table displays the NC distributions in percent for the frequency bands between 0.04 Hz and 45 Hz for the vertical-component USN in the time windows 0-4 EET and 8-12 EET obtained from the nine selected working days (Table 5.3) and all 156 working days with low wind conditions (average wind velocity < 3 m/s). No significant differences are observed in the frequency bands above 0.6 Hz. Below 0.6 Hz an increased occurrence of transient signals is observed which is caused mainly by the global seismicity.

No significant differences between the 9-days and the 156-days data set are observed for the frequency bands above 0.6 Hz which are dominated by the man-made sources of seismic noise. The discussion of the nine selected working days in section 5.3 is therefore verified to be representative for the typical man-made contributions to the USN in Bucharest.

In the frequency bands 0.18-0.25 Hz and 0.25-0.6 Hz no distinct differences between the statistical properties of the USN at night- and daytime are observed in contrast to the analysis of the nine selected working days. The amount of time windows classified as NC1 or 2 decreases at nighttimes and increases at daytimes to ~83% for the complete data set. This effect is most probably caused by the influence of storm induced ocean-generated microseism from the Mediterranean and Black Sea which was excluded by the selection of the nine working days analysed in section 5.3. These storm induced events with large amplitudes depend not on the daytime and mask man-made transient signals with lower amplitudes. In the lower frequency bands 0.04-0.09 Hz and 0.09-0.18 Hz another effect is observed with the increased amount of time windows classified as NC3 and NC4 in the data set with all working days. Time windows containing several and/or strong transient seismic waves excited by earthquakes (see Figure 5.1 and Table 5.1) are classified to a significant amount as NC3 and NC4.

5.4.3 Comparison of working days / Sundays (low wind velocities)

In this section the classification results obtained from all time windows with low wind conditions (average wind velocity < 3 m/s) are discussed with focus on the USN above 0.6 Hz dominated by man-made sources. The discussion includes the time windows 4-8, 16-20 and 20-24 EET which are not analysed for the selected working days in section 5.3. The results are presented separately for the working days and the Sundays in Table 5.6. The noise classes 1 and 2 as well as 5 and 6 are summarised in Table 5.6 to improve the readability. Furthermore the table is limited to the frequency bands above 0.6 Hz as the following discussion focuses on this frequency range. Detailed results for all frequency bands are provided in Appendix C.

The differences of the USN conditions (statistical properties) in the time windows 4-8, 16-20 and 20-24 EET in comparison to nighttimes (0-4 EET) and the main working hours (8-16 EET) as well as working days and Sundays are strongly frequency dependent. Below 0.6 Hz the differences are rather small and unspecific as the dominating natural sources such as ocean-generated microseism and earthquakes depend not on daytime or weekday (see detailed tables in Appendix C). In general the observed (small) differences are largest between the nighttimes (0-4 EET) and the other daily time windows due to the remaining influence of the human activity.

Comp.	Freq. band /Hz	EET	#	Working days				#	Sundays			
				NC1+2	NC3	NC4	NC5+6		NC1+2	NC3	NC4	NC5+6
				%	%	%	%		%	%	%	%
Z	0.6-1	0-4	5006	34.70	48.64	16.60	0.06	915	34.86	37.92	27.10	0.11
Z	0.6-1	4-8	4842	10.90	78.87	10.22	0.00	1032	30.62	51.26	18.12	0.00
Z	0.6-1	8-12	4237	70.31	23.27	6.42	0.00	880	53.30	26.48	20.23	0.00
Z	0.6-1	12-16	3231	64.90	22.75	12.35	0.00	712	60.81	18.96	20.22	0.00
Z	0.6-1	16-20	3754	75.52	19.42	5.06	0.00	674	61.28	22.70	16.02	0.00
Z	0.6-1	20-24	4807	55.29	35.59	9.09	0.02	1004	61.85	28.69	9.36	0.10
Z	1-25	0-4	5041	1.17	10.47	86.21	2.14	915	2.84	15.52	79.13	2.51
Z	1-25	4-8	4847	0.72	19.25	79.88	0.14	1033	0.77	11.62	87.03	0.58
Z	1-25	8-12	4230	0.71	39.24	59.31	0.73	880	0.80	22.73	75.91	0.57
Z	1-25	12-16	3247	1.02	36.86	61.66	0.46	719	1.39	28.93	69.54	0.14
Z	1-25	16-20	3752	0.75	35.10	63.83	0.32	712	1.40	29.07	69.52	0.00
Z	1-25	20-24	4807	0.96	23.40	75.43	0.21	1003	1.00	23.23	75.37	0.40
Z	25-45	0-4	4955	6.96	18.79	69.28	4.96	886	7.00	21.11	67.27	4.63
Z	25-45	4-8	4785	0.71	9.68	86.81	2.80	1033	1.16	12.78	83.35	2.71
Z	25-45	8-12	4211	0.45	12.51	83.02	4.01	879	0.34	8.76	89.19	1.71
Z	25-45	12-16	3234	0.40	11.13	84.76	3.71	718	0.00	11.28	88.02	0.70
Z	25-45	16-20	3745	0.67	16.29	80.61	2.43	713	0.56	13.60	84.29	1.54
Z	25-45	20-24	4801	1.87	15.46	80.63	2.04	1001	1.90	14.79	81.32	2.00

Table 5.6: Comparison of the noise class distributions for working days / Sundays (low wind conditions) The table displays the NC distributions in percent for the frequency bands between 0.6-45 Hz for the vertical-component USN obtained from all working days and all Sundays with low wind conditions (average wind velocity < 3 m/s). No significant differences are observed in the frequency band above 25 Hz. Below 25 Hz an increased occurrence of transient signals is observed at Sundays. The most significant differences between working days and Sundays are marked by red coloured values in the noise class distributions of the Sundays.

Above 25 Hz the statistical properties of the USN differ also less in the time windows between 4-24 EET due to numerous transient signals (large amount of time series classified as NC4) in this time range. Only during the night hours (0-4 EET) the amount of transient signals is significantly reduced leading to lower noise amplitudes and to less time windows classified as NC4. The statistical properties of the USN (25-45 Hz) differ furthermore less between working days and Sundays. This demonstrates that transient signals occur frequently above 25 Hz any time but that they are too few to superimpose to a Gaussian distributed time series in general.

The temporal variations of the USN with daytime and weekday are most pronounced in the frequency bands 0.6-1 Hz and 1-25 Hz. Significant differences between the classification results of working days and Sundays are indicated by red values in Table 5.6 and discussed below. The median 68%-interval ranges (noise amplitudes) are given for the six daily time windows in the frequency bands 0.6-1 Hz and 1-25 Hz in Table 5.7.

The statistical properties of the USN in the time window 16-20 EET are very similar (0.6-1 Hz) or identical (1-25 Hz) to the conditions in the main working hours (8-16 EET) (Table 5.6). The USN (16-20 EET) amplitudes are in general slightly lower than in the working hours and higher than in the evening and morning hours (Table 5.7). The statistical properties in the morning (4-8 EET) and evening (20-24 EET) hours differ both from the night (0-4 EET) and the working hours and reflect the change of the human activity from night to day and vice versa in these time windows. In both frequency bands more transient signals are observed in the morning hours than in the other time windows at daytime which is illustrated by the higher amount of time windows classified as NC4 in the morning hours. The noise amplitudes in the morning and evening hours are in general very similar (Table 5.7).

Vertical Component		Median noise amplitudes in nm/s (median 68% interval range)	
Freq. band /Hz	EET	Working days	Sundays
0.6-1 Hz	0-4	63	56
0.6-1 Hz	4-8	103	77
0.6-1 Hz	8-12	160	102
0.6-1 Hz	12-16	157	105
0.6-1 Hz	16-20	135	98
0.6-1 Hz	20-24	96	81
1-25	0-4	1070	1040
1-25	4-8	2473	1524
1-25	8-12	5135	2485
1-25	12-16	4952	2531
1-25	16-20	3891	2266
1-25	20-24	2270	1822

Table 5.7: Median noise amplitudes on working days and Sundays. The table displays the median ranges of the 68% amplitude intervals for all working days and Sundays with low wind conditions (average wind velocity < 3 m/s). The lowest noise amplitudes are observed in the night hours. The amplitudes in the evening and morning hours are similar and smaller than the amplitudes during the working hours (8-20 EET).

The USN on working days and Sundays differs in its statistical properties as well as noise amplitudes in the frequency bands 0.6-1 Hz and 1-25 Hz. In general lower noise amplitudes but more transient signals are observed between 8-16 EET on Sundays than on working days. Due to the decreased human activity less transient signals are excited in total on Sundays which causes the lower noise amplitudes. The remaining transient signals lead in general to a higher kurtosis of the seismic noise time series indicated by the higher amount of time windows classified as NC4. The USN conditions (statistical properties and amplitudes) in the Sunday morning hours equal the conditions at nighttimes of working days in both frequency ranges (Table 5.6 and Table 5.7). The statistical properties of the USN in the Sunday evening hours differ less (0.6-1 Hz) or not (1-25 Hz) from the evening hours at working days (Table 5.6). This observation corresponds well with the typical weekend behaviour of most people with less/no activity at Sunday mornings and increased (social/weekend commuter) activity at Sunday evenings.

Concluding, the daily and weekly rhythms of human live affect significantly the seismic noise amplitudes and statistical properties especially in the frequency range 0.6-25 Hz. In general, the seismic noise amplitudes are lower and the frequency of transient signals is higher on Sundays and at nighttime than on working days and at daytime.

5.4.4 Comparison of working days with high / low wind velocities

In this section the classification results of all working days with low wind conditions (average wind velocity < 3 m/s) are compared with the results of all working days with high wind conditions (average wind velocity > 6 m/s). Differences of the USN statistical properties due to the wind conditions are observed especially in the frequency ranges 0.04-0.18 Hz and 0.6-25 Hz which are therefore presented in Table 5.8 and discussed below. The statistical properties of the USN in the frequency ranges 0.25-0.6 Hz and 25-45 Hz are observed to be hardly influenced by the wind conditions (Appendix C).

Comp.	Freq. band /Hz	EET	Working Days									
			#	Average wind vel. < 3 m/s				#	Average wind vel. > 6 m/s			
				NC1+2	NC3	NC4	NC5+6		NC1+2	NC3	NC4	NC5+6
Z	0.04-0.09	0-4	3596	24.22	22.41	50.92	2.45	44	0.00	43.18	54.55	2.27
Z	0.04-0.09	4-8	3355	14.43	30.01	50.73	4.83	43	2.33	79.07	16.28	2.33
Z	0.04-0.09	8-12	2976	10.75	26.61	57.26	5.38	65	1.54	67.69	29.23	1.54
Z	0.04-0.09	12-16	2262	14.28	26.66	54.33	4.73	177	1.13	45.76	49.72	3.39
Z	0.04-0.09	16-20	2672	17.10	32.63	45.17	5.09	109	3.67	41.28	50.46	4.59
Z	0.04-0.09	20-24	3408	17.31	32.10	45.63	4.96	21	0.00	95.24	4.76	0.00
Z	0.09-0.18	0-4	4804	74.85	13.36	11.16	0.62	58	13.79	75.86	10.34	0.00
Z	0.09-0.18	4-8	4575	66.30	14.73	18.14	0.83	57	22.81	73.68	3.51	0.00
Z	0.09-0.18	8-12	4028	57.57	18.77	23.39	0.27	86	20.93	61.63	17.44	0.00
Z	0.09-0.18	12-16	3059	57.57	18.76	23.34	0.33	233	25.75	49.79	24.03	0.43
Z	0.09-0.18	16-20	3541	65.55	16.38	17.74	0.34	144	18.06	51.39	30.56	0.00
Z	0.09-0.18	20-24	4591	69.57	15.16	14.62	0.65	28	25.00	64.29	10.71	0.00
Z	0.6-1	0-4	5006	34.70	48.64	16.60	0.06	59	55.93	38.98	5.08	0.00
Z	0.6-1	4-8	4842	10.90	78.87	10.22	0.00	59	57.63	35.59	6.78	0.00
Z	0.6-1	8-12	4237	70.31	23.27	6.42	0.00	89	69.66	26.97	3.37	0.00
Z	0.6-1	12-16	3231	64.90	22.75	12.35	0.00	240	52.08	41.67	6.25	0.00
Z	0.6-1	16-20	3754	75.52	19.42	5.06	0.00	148	42.57	45.95	11.49	0.00
Z	0.6-1	20-24	4807	55.29	35.59	9.09	0.02	29	68.97	24.14	6.90	0.00
Z	1-25	0-4	5041	1.17	10.47	86.21	2.14	59	1.69	38.98	59.32	0.00
Z	1-25	4-8	4847	0.72	19.25	79.88	0.14	59	1.69	45.76	49.15	3.39
Z	1-25	8-12	4230	0.71	39.24	59.31	0.73	89	1.12	48.31	49.44	1.12
Z	1-25	12-16	3247	1.02	36.86	61.66	0.46	240	1.67	38.75	59.58	0.00
Z	1-25	16-20	3752	0.75	35.10	63.83	0.32	147	2.04	42.18	55.10	0.68
Z	1-25	20-24	4807	0.96	23.40	75.43	0.21	29	6.90	55.17	37.93	0.00

Table 5.8: Comparison of the noise class distributions for working days with low / high wind velocities
The table displays the NC distributions in percent for the frequency bands between 0.04-0.18 Hz and 0.6-0.25 Hz for the vertical-component USN obtained from all working days with low and high wind conditions. In general the influence of strong transient signals typically present in the USN on the time series distributions is reduced by the superposition with the wind induced seismic noise (e.g. 1-25 Hz). At the same time the superposition of typically Gaussian distributed USN with wind induced seismic noise produces time series which are not Gaussian distributed (e.g. 0.09-0.18 Hz).

The differences between the statistical properties of the vertical and the horizontal USN (see the next section) are the same also for high and low wind conditions. The discussion is therefore limited to the vertical component USN.

In the frequency band 0.04-0.09 Hz the USN is a complex superposition of man-made seismic noise as well as ocean-generated microseism and seismic waves excited by earthquakes (see section 5.3.3.3). The amount of time series classified as NC1 and 2 is significantly decreased due to wind induced seismic noise from 10-24% (low wind) to 0-4% (high wind). At the same time the influence of single transient signals (earthquakes, man-made signals) is reduced by the wind induced noise which is illustrated by the lower amount of time windows classified as NC4 in several of the daily time windows. The same

effect is observed in the frequency band 0.09-0.18 Hz although the amount of (nearly) Gaussian distributed time series (NC1 and NC2) is in general higher in this frequency band.

The situation is more complicated in the frequency band 0.6-1 Hz due to the larger human influence on the USN. In the frequency band 0.6-1 Hz the same situation as for the two lower frequency bands is observed in the daytime hours 8-20 EET. The amount of time windows classified as NC1 and 2 (Gaussian distributed) as well as NC4 (strong influence of transients) is reduced at daytimes due to the superposition of the man-made seismic noise with the wind induced seismic noise. In the time windows with less human influence (night, morning and evening) the superposition of the wind induced noise with the lower man-made seismic noise is less deviated from the Gaussian distribution than the USN (0.6-1 Hz) on working days with low wind conditions.

In the frequency band 1-25 Hz the influence of man-made transient signals on the statistical properties of the USN is reduced by the wind induced seismic at all daytimes. The amount of time windows classified as NC4 is lower with high wind conditions. Furthermore, the typically observed differences between the USN statistical properties at night (0-4 EET), morning (4-8 EET) and working hours (8-16 EET) are not observed for the time windows with high wind conditions. Again, the influence of dominant man-made transient signals on the USN is reduced by the wind induced seismic noise.

Concluding, wind induced seismic noise due to average wind velocities larger than 6 m/s can be expected to cause deviations of the USN from the Gaussian distribution. Most time series during high wind conditions are classified as NC3. At the same time, the amount of time series classified as NC4 due to strong (man-made) transient signals is reduced as these transients are superimposed with the wind-induced seismic noise.

5.4.5 Comparison of the vertical and horizontal components

The classification results for all time windows classified as NC1-6 of the vertical, North-South and East-West components of the USN are presented in Table 5.9. The statistical properties of the North-South and East-West components of the USN in Bucharest are in general identical (Table 5.9). The differences between the statistical properties of the vertical component and the horizontal component USN are discussed for the complete data set as the differences between the components are not observed to be influenced by the weekday or the wind conditions.

In Table 5.9 the noise classes 1 and 2, 3 and 4 as well as 5 and 6 are summarised to improve the readability of the table. The detailed results are provided in Appendix C. The largest differences between the vertical and horizontal components of the USN are observed in the frequency range below 0.18 Hz. In the frequency range 0.18-0.6 Hz the statistical properties of the vertical and horizontal components are nearly identical (Table 5.9). Minor differences are observed in the frequency range above 0.6 Hz. (Table 5.9).

In the frequency bands below 0.18 Hz more time windows classified as NC3 and 4 instead of NC1 and 2 are observed for the horizontal components of the USN. This indicates the presence of a larger amount of transient signals on the horizontal components than on the vertical components. This effect may be caused by the properties of the contributing noise sources inside as well as outside the Bucharest area.

Freq. band /Hz	EET	Z			N			E		
		NC1+2	NC3+4	NC5+6	NC1+2	NC3+4	NC5+6	NC1+2	NC3+4	NC5+6
		%	%	%	%	%	%	%	%	%
0.04-0.09	0-4	23.82	71.68	2.48	5.42	84.24	6.93	5.99	84.94	6.35
0.04-0.09	4-8	15.13	76.60	4.66	3.20	78.52	10.62	3.26	81.07	10.42
0.04-0.09	8-12	10.26	81.94	5.11	1.13	80.84	11.17	0.86	81.17	11.35
0.04-0.09	12-16	11.80	81.05	4.82	1.16	82.56	9.38	1.08	83.16	9.44
0.04-0.09	16-20	14.20	78.88	5.01	1.88	81.26	11.71	2.06	82.59	10.65
0.04-0.09	20-24	18.17	75.77	4.45	4.04	80.97	11.84	4.21	82.54	10.70
0.09-0.18	0-4	73.15	25.72	0.59	67.65	31.53	0.20	67.00	32.42	0.12
0.09-0.18	4-8	66.11	32.26	0.73	58.10	40.39	0.29	59.14	39.99	0.08
0.09-0.18	8-12	56.34	42.13	0.47	43.82	54.16	0.16	44.91	53.47	0.17
0.09-0.18	12-16	58.03	40.48	0.34	44.53	53.10	0.21	46.00	52.18	0.19
0.09-0.18	16-20	62.87	35.12	0.52	50.05	47.56	0.33	51.28	46.43	0.25
0.09-0.18	20-24	70.40	28.71	0.74	63.45	35.84	0.08	63.74	35.79	0.14
0.18-0.25	0-4	88.50	11.38	0.04	87.33	12.22	0.03	86.75	13.10	0.03
0.18-0.25	4-8	84.64	14.80	0.04	84.21	14.95	0.11	85.08	14.39	0.07
0.18-0.25	8-12	82.66	16.36	0.05	78.96	19.79	0.13	80.62	18.67	0.04
0.18-0.25	12-16	82.71	16.40	0.05	77.84	20.90	0.04	79.05	19.65	0.07
0.18-0.25	16-20	84.79	14.48	0.03	81.35	17.39	0.03	81.98	16.66	0.03
0.18-0.25	20-24	88.17	11.70	0.09	88.57	11.03	0.05	88.37	11.50	0.07
0.25-0.6	0-4	83.14	16.10	0.01	83.79	15.32	0.01	84.20	15.13	0.01
0.25-0.6	4-8	82.29	17.12	0.06	79.33	19.89	0.03	81.29	18.19	0.01
0.25-0.6	8-12	79.37	19.58	0.06	77.66	21.20	0.09	79.42	19.89	0.03
0.25-0.6	12-16	80.76	18.07	0.03	77.61	20.86	0.04	78.79	19.90	0.00
0.25-0.6	16-20	83.96	15.16	0.01	80.47	18.36	0.00	81.78	17.28	0.01
0.25-0.6	20-24	86.07	13.85	0.06	86.62	13.06	0.00	87.83	12.12	0.00
0.6-1	0-4	34.87	64.51	0.05	40.21	58.98	0.06	40.53	58.95	0.00
0.6-1	4-8	16.60	83.24	0.00	13.56	85.99	0.00	12.60	87.27	0.00
0.6-1	8-12	65.84	33.49	0.00	67.45	31.67	0.01	71.64	27.82	0.00
0.6-1	12-16	63.27	35.86	0.00	66.55	32.34	0.00	70.01	29.10	0.00
0.6-1	16-20	68.46	30.83	0.00	63.96	35.20	0.01	66.42	32.94	0.00
0.6-1	20-24	56.99	42.95	0.02	47.54	52.12	0.02	43.30	56.56	0.09
1-25	0-4	1.47	96.19	2.21	3.79	91.90	3.77	5.84	88.31	5.43
1-25	4-8	0.76	98.80	0.34	1.72	96.22	1.54	1.94	96.02	1.84
1-25	8-12	0.64	97.81	0.68	7.16	91.06	0.74	7.24	91.22	0.82
1-25	12-16	0.97	97.96	0.42	7.17	91.29	0.62	6.75	91.94	0.79
1-25	16-20	0.91	98.48	0.34	5.23	93.81	0.55	5.78	93.18	0.91
1-25	20-24	1.06	98.55	0.33	2.44	95.91	1.34	3.04	94.34	2.57
25-45	0-4	6.83	86.64	4.63	7.21	89.09	2.67	5.73	85.95	7.05
25-45	4-8	0.84	95.52	2.67	1.51	95.99	1.32	2.73	91.30	4.96
25-45	8-12	0.40	95.04	3.39	0.51	95.82	2.33	1.10	90.48	7.48
25-45	12-16	0.31	95.55	2.96	0.29	96.70	1.85	1.36	91.12	6.61
25-45	16-20	0.62	97.02	1.99	1.00	97.40	0.95	2.81	91.82	5.00
25-45	20-24	1.76	95.94	2.15	3.07	95.11	1.31	4.61	90.43	4.80

Table 5.9: Comparison of the noise class distributions for the components Z, N and E.

The table displays the NC distributions in percent for the frequency bands between 0.04-45 Hz obtained from all time windows classified as NC1-6 for the vertical, North-South and East-West components of the USN. No significant differences are observed between the North-South and East-West components in general. Minor differences between the horizontal components N and E exist in the frequency range 25-45 Hz due to man-made sources of periodic signals with a dominant direction of vibration. The statistical properties of the vertical and horizontal components are identical in the frequency range 0.18-0.6 Hz. In the frequency bands below 0.18 Hz less time series of the horizontal component USN are classified as NC1 and NC2 than of the vertical component USN.

A first idea is that the effect might be caused by the surface waves excited by earthquakes. Nevertheless, this question needs further research and remains open at the moment.

In the frequency bands above 0.6 Hz only small differences between the vertical and horizontal components are observed. In the frequency band 1-25 Hz the amount of time series classified as NC1 and 2 is slightly increased on the horizontal components. This indicates that the amount of signals or the amplitude distribution of the signals is slightly different between the vertical and the horizontal components. This point needs also further research which should include a careful analysis of individual signals contributing to the USN in the frequency range 1-25 Hz.

In the frequency range 25-45 Hz the statistical properties of the vertical and the North-South component of the USN are nearly identical. Eye-catching is the slightly larger amount of time series classified as NC5 and 6 on the East-West component. It turned out that over 70% of the USN (25-45 Hz) East-West component time series classified as NC5 and 6 are measured by the stations URS19 and URS20. At both station sites periodic signals such as the example presented in Figure 4.2f are recorded with significantly larger amplitudes on the East-West component than on the North-South component. Electrical engines which have a dominant direction of vibration are the most probable source of such a signal.

Concluding, the differences between the statistical properties of the spatial components of the USN are rather small. The largest differences are observed below 0.18 Hz with a larger amount of transient signals on the horizontal components. Minor differences are observed between 1-25 Hz with a slightly increased amount of time windows classified as NC1 and NC2 on the horizontal components. Periodic signals with a dominating direction of vibration occur above 25 Hz. The observed differences have to be assumed to be related to the sources of the seismic noise such as surface waves excited by earthquakes (below 0.18 Hz) or rotating machinery (especially above 25 Hz). This topic needs further research.

5.4.6 Time-domain H/V ratio

Similar to the spectral H/V ratio (see 2.3.1) the ratio between the ranges of the 68% amplitude intervals of the horizontal components and the vertical component are used to discuss the results of the time domain classification of the URS data set.

Figure 5.10 shows the results of a simplified H/V analysis using the noise classification results in comparison to the spectral H/V analysis for the Bucharest area of Ziehm (2006). For this analysis the complete URS data set is analysed in the frequency bands 0.09-0.18 Hz, 0.18-0.25 Hz, 0.25-0.6 Hz, 0.6-1.2 Hz, 1.2-2.4 Hz, 2.4-4.8 Hz and 4.8-9.6 Hz. Above the frequency bands of the ocean-generated microseism (see section 5.1.3) four frequency bands of one octave each are used up to 9.6 Hz to obtain H/V results which can be roughly compared with the existing spectral H/V analysis of Ziehm (2006) in the frequency range 0.1-10 Hz.

The H/V ratios of the 68% intervals obtained from all time windows classified as NC1-6 and averaged over all URS stations are presented in Figure 5.10a. The H/V ratios are plotted at the centre frequency of the corresponding frequency band and the standard deviation of the H/V ratios in the corresponding frequency bands is indicated by the error bars. The largest average H/V ratio of ~ 2.9 with also the largest standard deviation is observed in the frequency band 0.09-0.18 Hz. The H/V ratio decreases below 1 at frequencies above 2 Hz with a local minimum in the frequency band 0.25-0.6 Hz and a local maximum in the frequency band 0.6-1.2 Hz. The spectral H/V ratio for the time window 0-3 EET on 3rd January 2004 at station URS21 is shown in Figure 5.10b which is taken from the diploma thesis of Julia Ziehm (Ziehm, 2006).

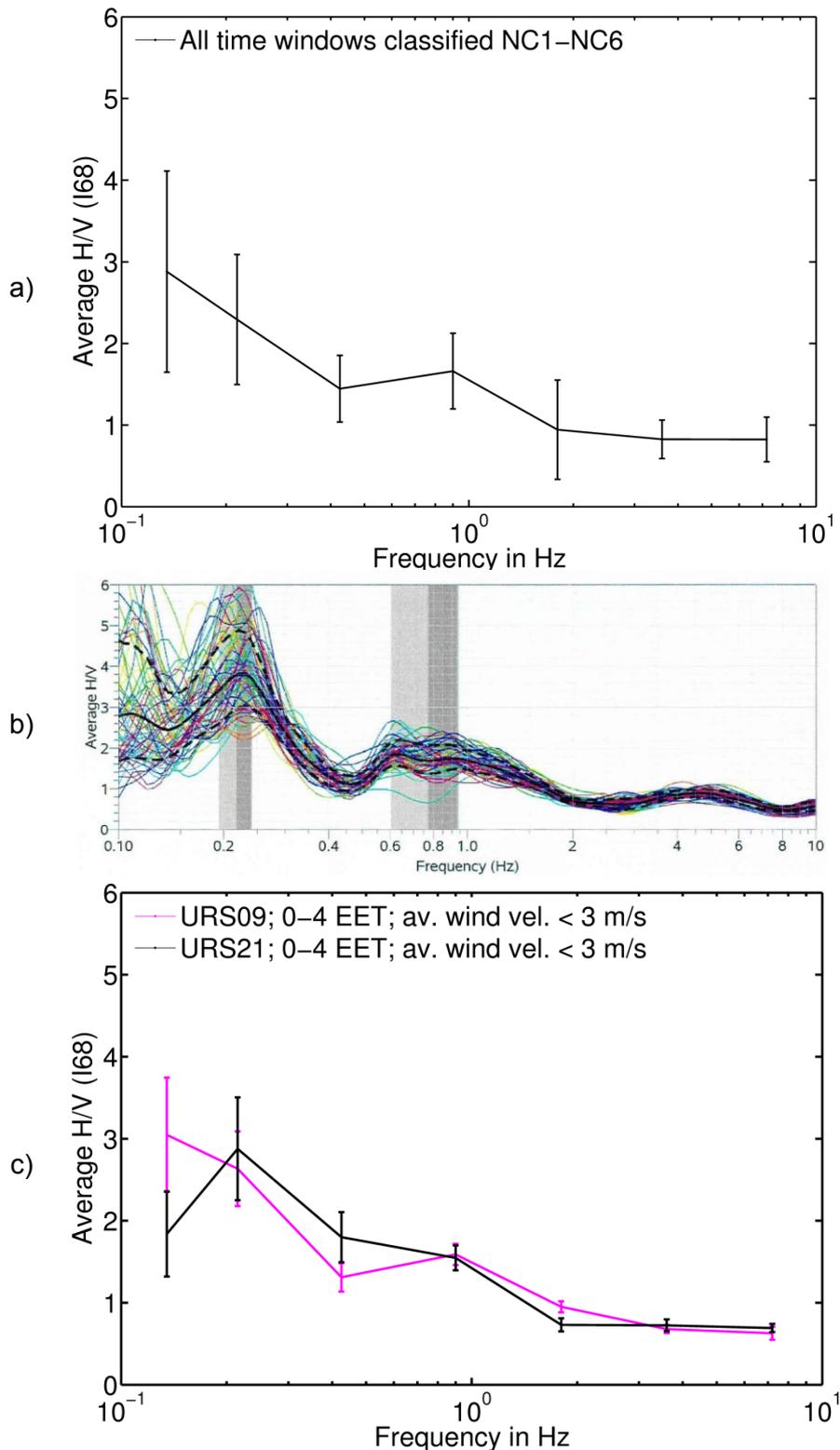


Figure 5.10: Time-domain and spectral H/V ratios observed in the Bucharest area. In **a)** the average ratios of the vertical and horizontal 68% amplitude intervals of all time windows classified as NC1-6 are presented together with the corresponding standard deviation in the frequency bands 0.09-0.18 Hz, 0.18-0.25 Hz, 0.25-0.6 Hz, 0.6-1.2 Hz, 1.2-2.4 Hz, 2.4-4.8 Hz and 4.8-9.6 Hz. The frequency bands are selected for the comparison with the spectral H/V ratio in **b)** of Ziehm (2006). In **b)** the spectral H/V ratio for the time window 0-3 EET (2004-01-03) at station URS21 is shown. This figure is taken from page 29 of the diploma thesis of Julia Ziehm (Ziehm, 2006). In **c)** the average time-domain H/V ratios obtained from all time windows at night (0-4 EET) with low wind velocities are shown for stations URS09 and URS21. The spectral and time-domain H/V ratios correspond quite well.

The coloured lines in Figure 5.10b illustrate the spectral H/V ratios of shorter time windows in the analysed time range from 0-3 EET at station URS21. The solid black line is the average spectral H/V ratio and the dashed black lines indicate its standard deviation. In Figure 5.10c the average time-domain H/V ratios for the stations URS09 and URS21 obtained from the time windows at night (0-4 EET) and with low wind conditions are presented. The time windows with moderate and high wind conditions are excluded as the H/V ratio of the USN is significantly influenced by wind induced seismic noise for wind velocities larger than 4 m/s (Ziehm, 2006).

Although the fairly simple time-domain H/V ratio presented here has a worse frequency resolution in comparison to the spectral H/V ratio it is in principle capable to resolve the most important features of the H/V ratio such as the peak around 0.2 Hz (URS21 in Figure 5.10c, see also section 5.3.3.3) and the local maximum observed for the frequency band 0.6-1.2 Hz (Figure 5.10a and URS09 in Figure 5.10c) which were observed and discussed by Ziehm (2006).

The peak around 0.2 Hz is not present in the average time-domain H/V ratio of all time windows (Figure 5.10a) and in the presented average H/V ratio of station URS09 in Figure 5.10c. This is most probably related to the worse frequency resolution of the presented simple time-domain H/V ratio (frequency bands 0.09-0.18 Hz and 0.18-0.25 Hz) on the one hand and the missing data selection on the other hand. The calculation of the spectral H/V ratio includes some criteria of data selection (Ziehm, 2006) which are not realised for the here presented time-domain H/V ratios. The results presented by Ziehm (2006) illustrate furthermore an increasing temporal variability of the spectral H/V ratio towards lower frequencies. This corresponds with the increasing standard deviation towards lower frequencies of the time-domain H/V ratios (Figure 5.10a and c). The frequency of the peak around 0.2 Hz is observed by Ziehm (2006) to change slightly with latitude in the Bucharest area. Ziehm (2006) assumes that the peak is related to the already discussed dipping Neogene-Cretaceous boundary which is located at 800 m depth in the South and at 2000 m depth in the North (see also section 5.3.3.2).

The local maximum observed for the frequency band 0.6-1.2 Hz (Figure 5.10a and URS09 in Figure 5.10c) corresponds to a stable maximum in the spectral H/V ratio around 0.7 Hz which was observed by Ziehm (2006) in the whole Bucharest area. Ziehm (2006) points out, that it was not possible to relate this very stable peak around 0.7 Hz to a geological discontinuity in the upper most 100 meters of the subsurface.

Concluding, the noise classification is in general capable to provide the same H/V information as the spectral H/V ratio. The important next step is to identify suitable frequency bands which should be as wide as necessary and as narrow as possible to enhance the frequency resolution of this 'time-domain H/V ratio'. Furthermore, a suitable time window length may allow an improvement of the obtained H/V ratio by an automated data selection based on the noise classification similar to the data selection implemented in the calculation of the spectral H/V ratio.

5.5 Analysis with Self-Organizing Maps (SOMs)

The detailed analysis of the classification results for the URS data set in the previous sections of this chapter demonstrates that the time series classification introduced in chapter 4 is capable to provide valuable information about the (urban) seismic noise conditions. Nevertheless, such a manual analysis of the classification data set in order to identify this information requires reasonable time and effort. Therefore, the next step is to improve and enhance this analysis process. Most promising for this task is the application of machine learning and pattern recognition techniques to support the seismologist analysing the data set. Suitable for the first exploration of an unknown data set are unsupervised pattern recognition techniques which are capable to provide a classification of a data set without a priori domain knowledge.

In this section a first feasibility study with a widespread unsupervised pattern recognition technique is presented to demonstrate the future potential of this approach. It has to be clarified at this point that the implementation of a comprehensive seismic noise analysis with pattern recognition techniques based on the seismic noise classification needs further research and especially evaluation. The author selected the Self-Organizing Map (SOM) method (Kohonen, 2001) for this feasibility study. This unsupervised neural network technique is widespread in science (see Kohonen, 2001) and was already applied successfully for seismological analyses (e.g. Tarvainen, 1999; Köhler et al., 2010). The SOM method is a powerful tool especially for the low-dimensional visualisation and automated clustering of high dimensional data. With the *SOM toolbox for MATLAB* (Vesanto et al., 2000) a widespread and powerful implementation of the SOM technique exists which is also used for this study.

5.5.1 The Self-Organizing Map

The SOM method is used to provide a two-dimensional representation (the ‘map’) of a high dimensional data set with d dimensions. Every neuron is represented by a d -dimensional *prototype vector* and corresponds to a so called *SOM unit* on a two-dimensional regular grid. Usually a rectangular grid with hexagonal SOM units is used to build the SOM (e.g. Köhler et al., 2010). This means that every neuron (SOM unit) has six adjacent neurons on the map. Every neuron is connected with the other neurons by a neighbourhood function which determines how strongly the neurons are connected. Initially, the prototype vectors of the SOM units are distributed in any order in the d -dimensional data space. Now the SOM is trained with an algorithm similar to vector quantization (Gray, 1984) using the d -dimensional data vectors of the (training) data set. For the training of the SOM the distances between data vectors and prototype vectors are calculated with a distance measure which is capable to handle missing values (see Vesanto et al., 2000 and below). The prototype vector closest to a data vector presented to the SOM is called the *Best-Matching Unit* (BMU). During training the BMU is stretched together with the prototype vectors of its neighbouring neurons (SOM units) towards the presented training data vector. The stretching of the neighbouring prototype vectors is controlled by the neighbourhood function. The training process organizes the prototype vectors in the data space preserving the topology of the map (neighbouring units on the SOM are neighbouring prototype vectors in the data space) and every prototype vector represents afterwards a group of close data vectors. New data vectors presented to the trained SOM can be classified by simply determining the BMU (closest prototype vector). For this study the commonly used Gaussian neighbouring function and the batch training algorithm are used to build and train the SOM (Köhler et al., 2010; Vesanto et al., 2000).

A trained SOM is commonly visualised by the so called *unified distance matrix* (U-matrix, see Figure 5.11a). The U-matrix visualises the distances between the SOM units (prototype vectors) by colour. Large distances (low data density) are indicated by red colours and low distances (high data density) are indicated by blue colours. For the U-matrix every SOM unit (except the units at the border of the map) is divided into seven

hexagonal subunits. The centre subunit shows the average distance of the SOM unit to all neighbouring SOM units by colour. The six surrounding subunits show the distance of the SOM unit to its corresponding adjacent SOM units. Therefore, the U-matrix visualises the cluster structure of the SOM. High values (red colours) in the U-matrix indicate the borders between clusters and low values (blue colours) indicate the clusters themselves. The clustering of the SOM prototype vectors can be done with a variety of well-established clustering algorithms. Following Köhler et al. (2010) an average linkage hierarchical clustering algorithm (Vesanto & Alhoniemi, 2000) is used for this study.

5.5.2 Application to the URS data set

For this study a data set of 1709 data vectors is built. The data vectors represent the 1709 analysed 4 hour time windows of the vertical component USN in Bucharest between October 2003 and August 2004. The frequency band 0.6-1 Hz is selected for this study as the USN in this frequency band is influenced significantly by man-made as well as natural seismic sources.

Nearly all URS stations are used for this study excluding only stations URS33, 34 and 35 from the data set. These stations were installed for only 2-3 months at the end of the URS project and station URS35 was installed at the top floor of a tall building (Appendix B).

The components of each data vector, also called the 'features' of the data, are the ranges of the 68% amplitude intervals of the 31 used URS stations as well as the corresponding peakfactors. The peakfactor is used instead of the noise class as it is the underlying measure and resolves the deviations from the Gaussian distribution with a higher resolution. Concluding, every data vector summarises 62 data features (31 interval ranges and 31 peakfactors) characterising the vertical component USN (0.6-1 Hz) in the Bucharest area in every 4 hour time window. The application of a statistical feature such as the peakfactor to characterise the seismic noise is a new approach. Typically, features obtained from array analyses, spectral analyses or polarisation analyses are used to characterise seismic noise for pattern recognition (Köhler et al., 2009). These features are in general not capable to capture the temporal and spatial variations of the statistical properties of the seismic noise caused by the superposition of signals emitted by varying noise sources.

Only the noise amplitudes and peakfactors of time series classified as NC1-6 are used to build up the data vectors in order to exclude corrupt time series which do not represent the USN. The components of a data vector are filled with NaNs (Not a Number) if the corresponding amplitude and peakfactor values of a station are missing (e.g. classified as NC12 in the corresponding time window or if the station was temporary not operational). The used distance measure allows the replacement of missing values by NaNs what significantly reduce the influence of missing values on the training of the SOM (Vesanto et al., 2000). The 31 amplitude components of the 1709 data vectors are visualised in Figure 5.12. The gaps in the plot indicate the missing values.

Finally, the data vectors are normalized prior to the training of the SOM. This normalisation is necessary as the variances of the 31 amplitude components of the data vectors are significantly larger than the variances of the 31 peakfactor components. Without normalisation the amplitude components would dominate the organisation of the map. The components are normalised with a so-called 'logistic' normalisation (Vesanto et al., 2000) in that way that all values between $-\infty$ and ∞ are within the range of [0,1].

The original values x of a given component are first scaled by a variance normalisation, using the mean value \bar{x} and the standard deviation σ_x , which leads to

$$\hat{x}=(x-\bar{x})/\sigma_x. \quad (5.1)$$

The normalised values x_n of the original values x of a given component are then obtained by

$$x_n=1/(1+e^{-\hat{x}}). \quad (5.2)$$

This normalisation is more or less linear around the mean value \bar{x} and has a smooth nonlinearity at both ends of the range of the component. Furthermore, this normalisation ensures that values normalised in the future (new data vectors analysed with the trained SOM) are within $[0,1]$ after the normalisation.

5.5.3 Results of the SOM analysis

The U-matrix of the SOM after the training with all 1709 data vectors of the vertical component USN (0.6-1 Hz) is shown in Figure 5.11a with a possible clustering with five clusters in Figure 5.11b. The determined clusters are numbered from 1 to 5 in Figure 5.11b and indicated by colours in this and the following figures. These clusters are used in the following to classify the time windows. The clustering of the SOM units is a critical point of the analysis. The clustering is affected by the chosen clustering algorithm as well as the number of assumed clusters. In general the analyst should not expect or accept one clustering of the SOM as the best solution (Köhler et al., 2010). In fact, the clustering of the SOM is that point where the human analyst with his domain knowledge enters the data exploration process. The further analysis includes the evaluation of the different possible clusters and the identification of meaningful clusters for the classification.

For this analysis the clustering with five clusters, as shown in Figure 5.11b, is comprehensive with the displayed U-matrix. A classification of the seismic noise time windows is introduced based on this clustering of the prototype vectors. Corresponding to the five clusters found in the SOM, five so-called *SOM classes* are defined. The cluster membership of the BMU of a given data vector is used to allocate a SOM class to the corresponding time window of seismic noise. The meaningful classification of the URS data set by the five SOM classes obtained from the presented clustering is demonstrated in the following.

In Figure 5.12 the amplitude components of the data vectors are plotted with the same colours of the clustering result in Figure 5.11b. This visualisation illustrates already the meaningful classification based on the five clusters of prototype vectors. Time windows with unusual high noise amplitudes (ranges of the 68% interval) at all URS stations are classified as SOM class 1 (2.57% of the 1709 time windows; magenta). The time windows with the lowest amplitudes are classified as SOM class 4 (55.47%; orange) and higher amplitudes are classified as SOM class 2 (38.15%; light green). With the domain knowledge from the previous manual analysis it can be expected that SOM class 1 discriminates time windows with high wind conditions and that SOM classes 2 and 4 discriminate at least daytime and nighttime. The time period around time window 400 in Figure 5.12 with comparable low amplitudes and classified continuously as SOM class 4 corresponds to the Romanian Christmas vacation. Time windows classified as SOM classes 3 (1.4%; yellow) and 5 (2.4%; dark blue) are rare and hardly identifiable in Figure 5.12. They are discussed together with the other SOM classes in more detail below.

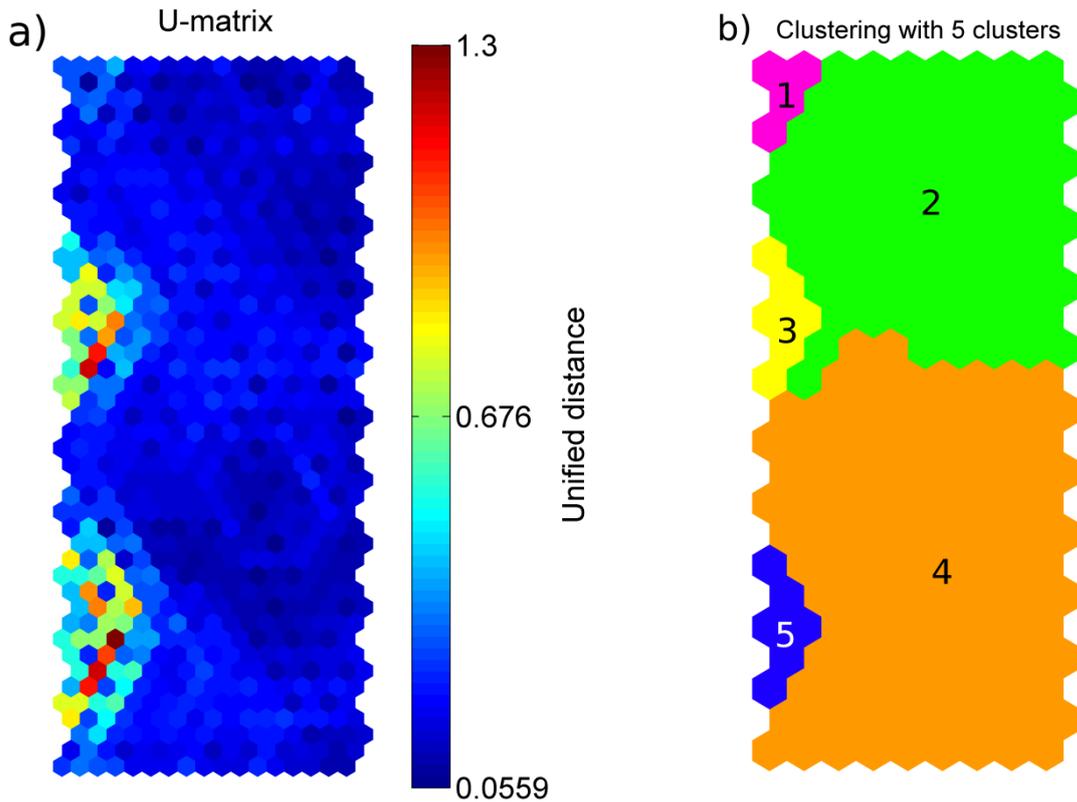


Figure 5.11: The U-matrix with a possible clustering of the SOM for the vert.-comp. USN (0.6-1Hz). In a) the U-matrix of the SOM trained with the 1709 data vectors (31 amplitude components, 31 peakfactor components) of the vertical component USN (0.6-1 Hz) is shown. The clustering of the U-matrix in a) with 5 clusters is shown in b).

In Figure 5.13 the relation between the average wind velocity in the analysed time windows and the SOM class affiliation of the time windows is demonstrated with the help of histograms. For each of the five SOM classes a histogram over the average wind velocity is plotted. The bars of the histograms show the number of the time windows (in percent) in the corresponding SOM class for a given average wind velocity.

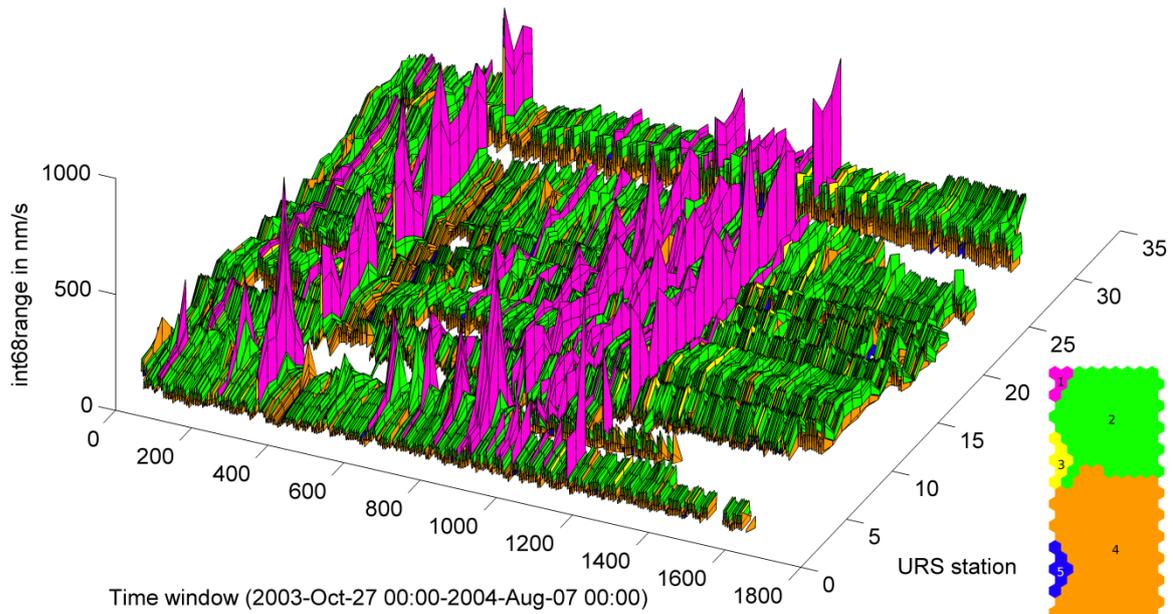


Figure 5.12: Amplitude components of the USN (0.6-1 Hz) data vectors after the SOM classification. The 68% amplitude interval ranges of the vertical-component USN (0.6-1 Hz) determined in the 1709 time windows are plotted for the 31 URS stations. The SOM classification of the time windows is indicated by the colouring in correspondence to the clustering shown in Figure 5.11a.

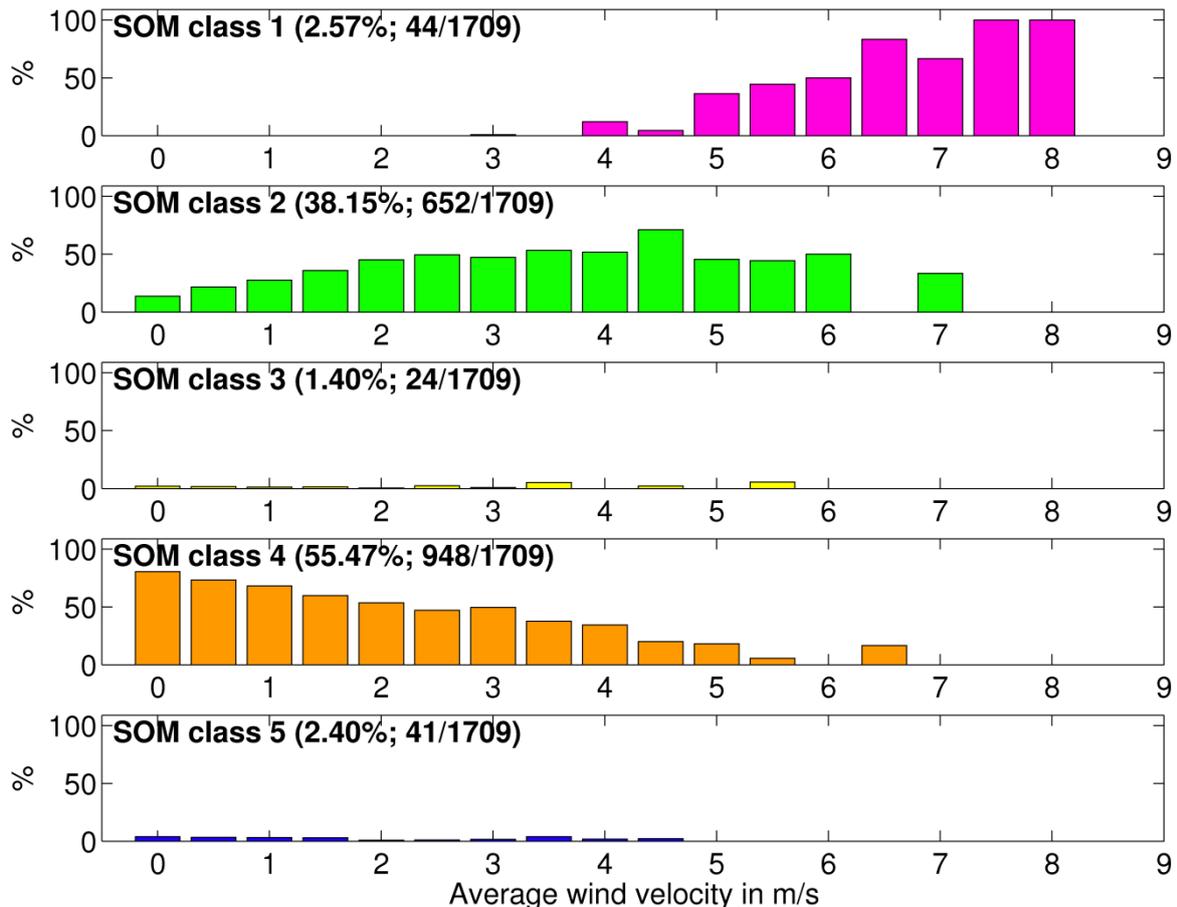


Figure 5.13: Relation between the average wind velocity and the SOM classification. The bars of the five histograms (one for each SOM class) show the number of time windows (in percent) for given average wind velocities. Example: 100% of the time windows with an average wind velocity larger than 7 m/s are classified as SOM class 1.

As an example: 100% of the time windows with an average wind velocity of ~8 m/s are classified as SOM class 1. The classification of a time window as SOM class 1 indicates reliably wind velocities which are significantly larger than the overall average wind velocity of ~2 m/s between October 2003 and August 2004. Time windows with average wind velocities smaller than the long term average of 2 m/s are increasingly classified as SOM class 4 whereas time windows with larger average wind velocities are classified increasingly as SOM class 2 (3-5 m/s) and then SOM class 1 (> 5 m/s). This corresponds well with the observation that the influence of wind induced seismic noise increases significantly for wind velocities larger than 4 m/s (see Figure 5.1c).

The relation between the weekday and daytime of the time windows with the SOM classification is also demonstrated with a histogram analysis which is shown in Figure 5.14. The occurrence (in percent) of the 42 time windows of a week (Monday to Friday, six daily time windows between 0 and 24 EET) is plotted for each of the five SOM classes. The SOM classes 2 and 4 with most members (93% of the 1709 time windows) image the daily human activity. The 38.15% percent of all time windows classified as SOM class 2 are mainly working hour time windows between 8-20 EET at working days. The most time windows classified as SOM class 4 are night, morning and evening hours at working days and time windows at Sundays. It is furthermore demonstrated by the classification that Saturdays appear as something between a working day and a Sunday. The time windows between 8-16 EET are found mainly in SOM class 2 together with the working day time windows. But the time window 16-20 EET is in general not classified as SOM class 2 as on working days but as SOM class 4 as on Sundays. This reflects very well the typical

weekend activity of most people in Christian dominated societies. The Saturday is a shopping and (home) working day for most people causing an increased traffic in the inner city area, whereas the Sunday is a day of rest with the family and therefore few 'seismic emissions'.

The time windows classified as SOM classes 3 and 5 discriminate the weekly time windows similar as SOM classes 2 and 4. In SOM class 3 mainly time windows in the working hours of working days are found in contrast to the time windows in class 5 which cover mainly night, evening, morning and the Sunday time windows. An analysis of the time series of the time windows of SOM classes 3 and 5 revealed that the USN (0.6-1 Hz) is classified as NC3 or NC4 at all station sites in these time windows. This is an unusual situation in the frequency band 0.6-1 Hz especially as the ranges of the 68% amplitude intervals are not unusual for the affected time windows (Figure 5.12). Systematically increased peakfactors in combination with only small changes of the 68% amplitude intervals at all station sites can be caused by short (in comparison to 4 hours) and large (in comparison to the 68% amplitude interval) transient signals. As all station sites are affected at the same time single strong sources such as earthquakes or explosions have to be considered as possible causes.

The body waves excited by regional and teleseismic earthquakes can be observed in the frequency range below 1 Hz depending on distance and magnitude (green stars in Figure 5.1b and c). The seismic waves excited by the earthquakes in the nearby (~150 km distance) Vrancea subduction zone with magnitudes typically smaller than 4 during the URS experiment are observed in the frequency range above 0.6 Hz (see red stars in Figure 5.1b and c). The 15 largest earthquakes and the 15 regional earthquakes closest to Bucharest which can be found in the NEIC catalogue for the period of the analysed time windows are summarised in Table 5.10 and Table 5.11. The five local earthquakes with magnitudes larger than four which occurred in the Vrancea subduction zone during the URS project are summarised in Table 5.12. The SOM classification of the corresponding time windows is given together with the estimated P-onset times, magnitudes, depths and distances of the earthquakes. The 35 earthquakes affect 34 time windows of which 28 are classified as SOM class 3 or 5. The SOM analysis based on the 68% amplitude intervals and the peakfactors has enough sensitivity to discriminate earthquake waves arriving in Bucharest during working hours at working days from earthquake waves arriving in time windows with less human activity. The remaining 37 time windows classified as SOM class 3 or 5 commonly cover arrival times of earthquakes with magnitudes above six in a distance range between 10° and 100° (not listed in the tables).

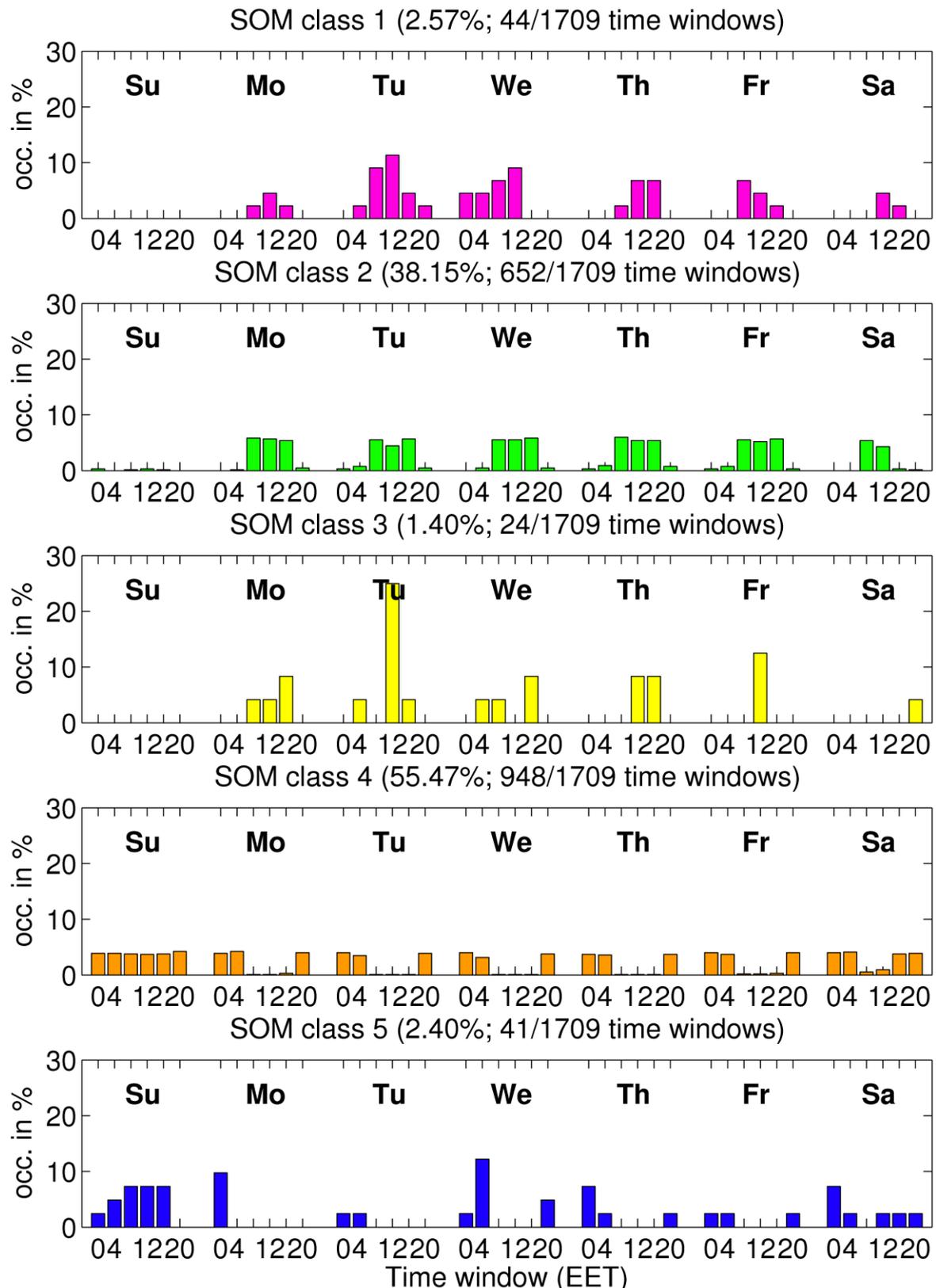


Figure 5.14: Relation between weekday and daytime of the time windows with the SOM classification. The occurrence (in percent) of the 42 time windows of a week (Monday to Friday, six daily time windows) is plotted for each of the five classes. The SOM classes 2 (38.15%) and 4 (55.47%) with most members reflect the daily human activity. SOM classes 3 and 5 discriminate time windows affected by earthquake waves during working hours at working days (class 3) and the remaining weekly time windows (class 5).

No relation is found between the P-onset times of local Vrancea earthquakes with magnitudes smaller than four and time windows classified as SOM class 3 and 5. The 76

seismic waves of these smaller local earthquakes are hardly observed in the frequency range below 1 Hz (see red stars in Figure 5.1b and c). For several of the time windows classified as class 3 or 5 no corresponding earthquakes can be found in the catalogues. Nevertheless, the NEIC catalogue is not complete for regional earthquakes in the magnitude range between four and five. The next step would be a careful analysis of the time series in the identified time windows to derive a magnitude-distance relationship for the detection threshold of the URS network. The detection of earthquake waves with this approach may be further improved by the selection of optimized frequency bands for the local, regional and teleseismic seismicity.

Concluding, the SOM analysis provides meaningful classes of time windows which are additional valuable information for the further manual analysis. Nevertheless, more work has to be done to establish SOMs as a standard tool for the analysis of a noise classification data set such as the URS data set. Especially the reliable and optimal clustering of the obtained SOM prototype vectors for this application of the SOM method needs further evaluation. It is assumed that more meaningful 'subclasses' of the SOM classes 2 and 4 can be found as these two classes comprise more than 93% of the 1709 time windows.

The 15 largest earthquakes (NEIC catalogue)						
Estimated P-onset EET	weekday	Time window EET	M _w	Depth /km	Distance /°	class
2003-11-17 08:55	Monday	8-12	7.8	33	82	3
2004-02-07 05:56	Saturday	4-8	7.5	10	106	5
2004-07-25 17:35	Sunday	16-20	7.3	582	83	5
2003-12-27 18:20	Saturday	16-20	7.3	10	143	4
2004-02-06 00:19	Friday	0-4	7.1	16	106	4
2004-01-03 18:43	Saturday	16-20	7.1	22	142	5
2004-07-15 07:47	Thursday	4-8	7.1	565	146	5
2003-10-31 03:18	Friday	0-4	7.0	10	80	5
2004-06-10 18:31	Thursday	16-20	6.9	188	73	3
2004-02-08 11:13	Sunday	8-12	6.9	25	106	4
2003-12-26 04:03	Friday	4-8	6.8	10	30	5
2003-12-10 06:50	Wednesday	4-8	6.8	10	78	5
2004-06-28 13:02	Monday	12-16	6.8	20	79	3
2003-12-26 23:45	Friday	20-24	6.8	10	143	4
2003-12-05 23:37	Friday	20-24	6.7	10	74	5

Table 5.10: The 15 largest earthquakes which occurred during the URS experiment. The table summarises the 15 largest earthquakes found in the NEIC catalogue. The estimated P-onset times are given together with the weekday, the affected time window, the magnitude and depth of the earthquakes as well as the distances of the epicentres to the city centre of Bucharest. The SOM classification of the affected time windows is given in the last column. Eleven of the 15 time windows affected by the largest earthquakes are classified as SOM class 3 and 5.

The 15 closest regional earthquakes (NEIC catalogue)						
Estimated P-onset EET	weekday	Time window EET	M _w	Depth /km	Distance /°	class
2004-06-15 15:03	Tuesday	12-16	5.2	12	4	3
2004-04-07 04:34	Wednesday	4-8	5.0	68	5.7	5
2004-05-23 18:21	Sunday	16-20	5.1	10	6.3	5
2004-08-03 16:13	Tuesday	16-20	5.2	2	7.5	3
2003-11-16 09:25	Sunday	8-12	5.2	8	7.5	5
2004-08-04 07:22	Wednesday	4-8	5.2	10	7.7	3
2004-08-04 06:03	Wednesday	4-8	5.6	10	7.7	3
2004-08-04 17:20	Wednesday	16-20	5.3	10	7.7	3
2004-03-01 02:38	Monday	0-4	5.6	9	7.9	5
2004-02-07 23:19	Saturday	20-24	5.2	25	8.4	5
2004-07-12 16:06	Monday	16-20	5.7	7	9	3
2004-03-17 07:23	Wednesday	4-8	6.1	24	10,1	5
2004-05-05 16:42	Wednesday	16-20	5.5	228	10,3	3
2004-03-25 21:33	Thursday	20-24	5.6	10	11,8	2
2004-03-28 06:53	Sunday	4-8	5.6	5	11,9	4

Table 5.11: The 15 closest regional earthquakes which occurred during the URS experiment. The table summarises the 15 closest regional earthquakes found in the NEIC catalogue. The table does not contain the local earthquakes in the nearby (~150 km distance) Vrancea subduction zone. The estimated P-onset times are given together with the weekday, the affected time window, the magnitude and depth of the earthquakes as well as the distances of the epicentres to the city centre of Bucharest. The SOM classification of the affected time windows is given in the last column. Thirteen of the 15 time windows affected by the 15 closest regional earthquakes are classified as SOM class 3 and 5.

Local earthquakes in the Vrancea zone with M > 4 (ROMPLUS catalogue)						
Estimated P-onset EET	weekday	Time window EET	M _w	Depth /km	Distance /°	class
2004-02-07 13:59	Saturday	12-16	4.4	143	1.28	5
2004-07-10 03:36	Saturday	00-04	4.3	150	1.29	5
2004-04-04 09:42	Sunday	08-12	4.3	150	1.23	5
2004-01-21 07:50	Wednesday	04-08	4.1	118	1.11	5
2004-03-18 01:43	Thursday	00-04	4.1	157	1.28	5

Table 5.12: The five largest local earthquakes which occurred in the Vrancea subduction zone. The table summarises the five local earthquakes with a magnitude larger than 4 which occurred in the near Vrancea subduction zone (ROMPLUS catalogue). The estimated P-onset times are given together with the weekday, the affected time window, the magnitude and depth of the earthquakes as well as the distances of the epicentres to the city centre of Bucharest. The SOM classification of the affected time windows is given in the last column. All of the five time windows affected by the largest local earthquakes are classified as SOM class 5. No correlation is found between the P-onset times of the Vrancea earthquakes with magnitudes smaller than four and the time windows classified as SOM classes 3 and 5.

5.6 Summary of chapter 5

A time-frequency analysis is used to determine 8 frequency ranges which represent the frequency-dependent influence of dominant natural and man-made sources on the urban seismic noise (USN). Below 0.6 Hz the USN is dominated by distant (0.04-0.18 Hz) and regional (0.18-0.6 Hz) ocean-generated microseism as well as seismic waves excited by earthquakes. In the frequency band 0.6-1 Hz the USN is dominated by man-made sources as well as wind induced seismic noise. The influence of the wind induced seismic noise is significantly increased for average wind velocities larger than 4-5 m/s. Above 1 Hz

the human activity is the dominating source of the USN. A human influence on the USN is observed also at very low frequencies especially below 0.09 Hz.

The time series classification introduced in chapter 4 proved to be capable to resolve the temporal and spatial variations of the USN and to provide additional information about the noise conditions in Bucharest. All typical noise sources identified by the time-frequency analysis are observed to cause typical variations of the statistical properties of the USN in the corresponding frequency bands. The deviations of the USN from the Gaussian distribution are typically larger at daytimes than at nighttimes due to man-made transient signals. This is not the case in the frequency band 0.6-1 Hz. The numerous man-made transient signals at daytime superimpose to a mainly Gaussian distributed USN in this frequency range especially during the working hours between 8-20 EET. Wind induced seismic noise causes higher noise amplitudes and reduces the influence of man-made transient signals on the statistical properties of the USN. Nevertheless, considering all frequency bands the deviations of the USN from the Gaussian distribution are increasing due to wind induced seismic noise.

The statistical properties of the horizontal components North-South and East-West are in general identical. Minor differences are observed at two station sites (URS19 and URS20) in the frequency range above 25 Hz due to periodic signals with a dominant direction of vibration. Most probable sources of these signals are electrical engines in direct vicinity of the sensors. The statistical properties of the vertical and the horizontal components show minor differences in the frequency range below 0.18 Hz and in the frequency range 1-25 Hz. In general a larger amount of transient signals is observed on the horizontal components than on the vertical component below 0.18 Hz. In the frequency band 1-25 Hz the amount of time windows classified as NC1 and NC2 is slightly increased on the horizontal components. The observed differences have to be assumed to be related to the sources of the seismic noise such as surface waves excited by earthquakes (below 0.18 Hz) or rotating machinery (especially above 25 Hz). This topic needs further research.

The average amplitude ratios between the vertical and the horizontal components (H/V ratio) obtained from the noise classification correspond in general well with the results of the spectral H/V ratio analysis of Ziehm (2006). The spectral H/V ratio is around or slightly below 1 for frequencies larger than 2 Hz and exhibits two peaks with average H/V ratios between 2 and 4 in the frequency ranges 0.18-0.25 Hz and 0.6-0.9 Hz. The frequency resolution of the presented time-domain H/V ratios is worse than the resolution of the spectral H/V ratio but can be improved by the selection of more suitable frequency bands.

The analysis of the vertical-component USN in the frequency band 0.6-1 Hz with the Self-Organizing Map (SOM) method provides five meaningful classes of time windows. The SOM classification discriminates reliably the working hours 8-20 EET on working days (SOM class 2, 38.1% of the 1709 analysed time windows) from the time windows at night, morning and evening as well as on Sundays (SOM class 4, 55.5%). The remaining 6.4% of the time windows are classified as SOM class 1 (2.6%), SOM class 3 (1.4%) and SOM class 5 (2.4%). The classification of a time window as SOM class 1 indicates reliably higher-than-average wind velocities. The time windows classified as SOM class 3 or 5 are found to be affected by seismic waves excited by local, regional and teleseismic earthquakes. The SOM classes 3 and 5 discriminate time windows affected by the global seismicity with high (SOM class 3) and low (SOM class 5) human activity identical to the SOM classes 2 and 4. In general, the feasibility study to analyse an urban seismic noise classification data set with the SOM method was very successful. Nevertheless, the clustering needs further evaluation as it is assumed to find more meaningful 'subclasses' of the SOM classes 2 and 4 which summarise more than 93% of the 1709 analysed time windows.

6 Improved calculation of seismic noise cross-correlation functions

The estimation of Green's functions based on seismic noise cross-correlation functions (CCFs) evolved to an important and widely used technique in seismology (see chapter 2). It enables seismology to provide high-resolution tomography studies from local (e.g. Bussat & Kugler, 2009) to continental (e.g. Shapiro et al., 2005) scale and independent from earthquake seismicity or active seismic sources. Nevertheless, practical experience shows, that one has to use long time series (months to years) and to apply extensive normalisation to the seismic noise time series to obtain CCFs which are suitable to estimate Green's functions (see section 2.3.3.2). The cross-correlation of non-normalised 'raw' seismic noise time series produces CCFs which are in general not suitable to estimate Green's functions due to disturbing dominant signals (e.g. earthquake waves, ocean-generated microseism) or instrumental irregularities. The important task of the processing is to provide an 'artificial equipartitioning' of the seismic noise wave field. An equalisation of the signals contributing to the seismic noise in the time and frequency domain is necessary to be able to estimate the Green's function from the finally obtained seismic noise cross-correlation function.

Concluding, the data processing applied to obtain suitable seismic noise CCFs is critical and underwent an evolution in the last years. A current status of ambient seismic noise data processing (see section 6.1) with a discussion of several proposed time and frequency domain normalisations to enhance the seismic noise was published by Bensen et al. in 2007 and is referenced in more than 40 follow-up publications. Thus, the recommendations by Bensen et al. (2007) are established as a processing standard. Another considered possibility to enhance seismic noise CCFs next to normalisation is the selection of appropriate input data and the exclusion of problematic data pieces like (teleseismic) earthquakes and instrumental irregularities. Pedersen et al. (2007) applied successfully a data selection approach based on global earthquake catalogues to exclude worldwide seismic events with magnitudes larger than 5 and a CCF amplitude threshold to exclude remaining small seismic events. The data 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.

The thesis suggests and evaluates methods of automated data selection to improve the calculation of seismic noise cross-correlation functions. The time series classification introduced in chapter 4 is capable to provide an automatic exclusion of time windows containing transient signals or corrupt data due to instrument regularities. Based on the time series classification the author developed a fully automated data selection approach independent of earthquake catalogues (see also Groos et al., 2010). The applied time window length is a critical parameter for the efficient automatic exclusion of transient signals such as teleseismic earthquake waves. The implementation of the new data selection approach required therefore a comprehensive investigation of the performance of the different established normalisation methods in dependence of the time window length which is discussed in this chapter. The author developed furthermore a second automated data selection approach based on the waveform symmetry of the CCFs as well as a waveform preserving time domain normalisation of the CCFs in the context of this study.

A detailed overview about the applied data processing and the specific aspects of the data processing addressed by this thesis is given in the first section of this chapter. The significant influence of the time window length on the important frequency domain normalisation is discussed in section 6.2 prior to the introduction of the two fully automated data selection approaches in section 6.3. The developed method for the waveform preserving normalisation of the CCFs in the time domain is presented in section 6.4. The considered normalisation schemes are evaluated without data selection in

section 6.5 and with data selection in section 6.6. The chapter is completed by a short summary of the most important conclusions in section 6.7.

6.1 Data processing

In this section the data processing for the calculation of seismic noise cross-correlation functions is introduced with a discussion of the state-of-the-art as proposed by Bensen et al. (2007) after a brief discussion of the time series preprocessing. All aspects of the data processing which are addressed by this thesis are identified and illustrated by a schematic overview in section 6.1.2. All mainly technical but nevertheless important aspects of the data processing such as the digital cross-correlation, the calculation of describing data parameters as well as the influence of the time window length on stacked CCFs are briefly discussed in the following subsections.

6.1.1 Time series preprocessing

For every station of the GSN data set (see section 3.2) a 12-months time series of the ground motion velocity is processed from the raw data provided by the IRIS DMC. The 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 percent) 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 12-months time series from 01-January 2004 to 31-December-2004. Missing data is zero padded to obtain complete time series. The alternative to the zero padding is the cutting of individual time windows with differing lengths and for every station pair with the exclusion of every time window for which only data of one station is available. This approach causes high organisational and computational effort due to the individual time window selection for every station pair as well as the resulting varying time window length. It is therefore usual to use a fixed time window length and accept zero padding (e.g. Bensen et al., 2007). Working with the GSN data set the author observed also no negative influence of the zero padding on the obtained CCFs in comparison to the strict exclusion of time windows for which only data at one station is available. Time windows which consist mainly of padded zeros are reliably identified by the time series classification. The single time window CCFs obtained from such time windows are excluded automatically from the stacking. The 12-months time series are filtered again with the 0.003 Hz HP filter to remove remaining low frequency artefacts due to the instrument response deconvolution. With this preparation the ground motion velocity is obtained with a broad frequency range from the raw data. Prior to the cross-correlation the 12-months time series of both stations are finally band-pass filtered with a fourth order zero-phase filter in the period band of interest (here 7-150 s). These 12-months time series are produced to simplify the application of sliding time windows with different lengths for the following evaluation of the cross-correlation processing.

6.1.2 Data processing scheme

The data processing proposed by Bensen et al. (2007) is illustrated by the schematic processing flow in Figure 6.1. The time series preprocessing corresponds to the preprocessing discussed in the previous subsection. The cross-correlation of very long (months to years) time series of seismic noise at once is in general not reasonable for practical reasons. The length of the time series which can be processed at once is limited by the working memory of the used computer system. The cross-correlation processing is therefore applied to several shorter time windows (e.g. 24 hours) covering the total amount of data (here 12 months). The CCFs of the single time windows are afterwards stacked to obtain the complete CCF. This procedure is possible due to the linearity of the cross-correlation (Bensen et al., 2007). To apply the cross-correlation processing to daily time series (24 hr time window, Figure 6.1) is the widely used standard for the calculation of continental scale CCFs (e.g. Bensen et al., 2007; Yang & Ritzwoller, 2008). This time

window length is motivated by purely practical considerations as continuous seismic data is often stored in so called ‘dayfiles’, one for each Julian day of a year, containing each 24 hours of data.

After the time series preprocessing Bensen et al. (2007) propose to normalise every 24 hour time series first in the time and afterwards in the frequency domain. The established and wide spread non-linear time domain normalisations are the one-bit (1B) normalisation (e.g. Shapiro et al., 2005) and the running absolute mean (ram) normalisation (Bensen et al., 2007). The one-bit normalisation heavily distorts the original waveform by leaving only values of +1 for all positive and -1 for all negative amplitude values of the seismic noise time series. It is the extreme case of the running absolute mean normalisation with a time window length of one sample. The one-bit normalisation is known to negatively influence the frequency content of broadband time series (Pedersen et al., 2007; Sabra et al., 2005). Nevertheless, it is very effective by improving the emergence of signals in the seismic noise CCFs and is therefore still widely used. Bensen et al. (2007) state, that the running absolute mean normalisation with a time window length of half of the largest analysed period is smoother and produces better results in terms of a slightly higher signal-to-noise ratio. Bensen et al. (2007) evaluated also several other time domain normalisations which were not further considered as they were observed to be less effective and/or efficient as the running absolute mean or the one bit normalisation. A crucial point is that all established time domain normalisation methods significantly distort the waveforms of the original signals. A further discussion of the time domain normalisations can be found in section 6.4.

In the frequency domain, the spectral whitening is the established normalisation method. The implementations of this method differ slightly between different authors (e.g. Bensen et al., 2007; Brenguier et al., 2008a). Spectral whitening is applied to the time series (Time Series Spectral Whitening, TSSW) with the intention to improve the finally obtained CCFs in terms of being broad-band Green’s function estimates (Bensen et al., 2007). The spectral whitening is discussed in detail in section 6.2.

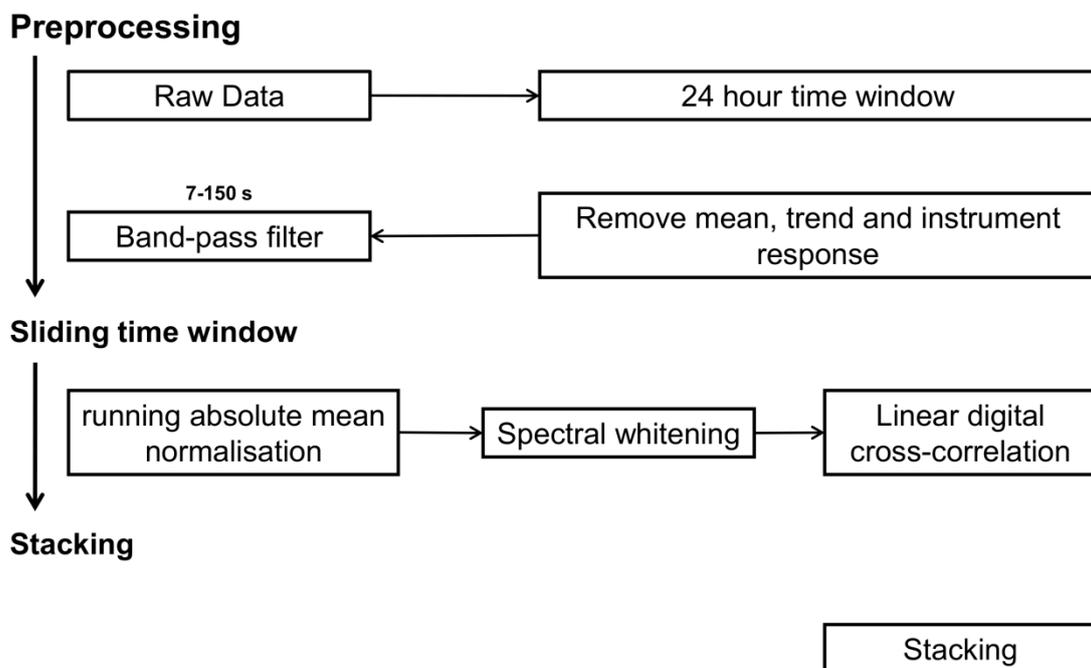


Figure 6.1: Processing scheme for the cross-correlation processing after Bensen et al. (2007). The whole data processing is applied to 24 hour long time series. The obtained CCFs are stacked afterwards to obtain a CCF which represents a longer time series (months to years). The time and the frequency domain normalisation are applied to the time series prior to the cross-correlation.

The processing scheme actually used for this thesis differs in some aspects from the scheme proposed by Bensen et al. (2007) and is illustrated in Figure 6.2. For the systematic evaluation of the data selection approaches and newly developed waveform preserving time domain normalisations several normalisation schemes were realised indicated by the optional fields in Figure 6.2. The 1B and ram time domain normalisations as well as the spectral whitening are considered as the established normalisation methods and compared with the new ones. The spectral whitening is not applied to the time series (TSSW) but to the CCFs (SW) for practical reasons as discussed in section 6.2. The data selection is realised by the exclusion of CCFs from the stacking as the CCFs are calculated for all single time windows at first.

The time window length of the sliding time window is varied between 1 hour and 1 year for this analysis to evaluate the performance of the different data selection and normalisation methods. A time window length of one hour is found to be the shortest reasonable time window length for continental scale CCFs as discussed in section 6.1.5. Time window lengths between one hour (Li et al., 2010) and more often the already discussed 24 hours were applied in published studies. Therefore, the analysis is focused on 47 different time window lengths between one hour and 24 hours by increasing the time window length step-wise by about 0.5 hr.

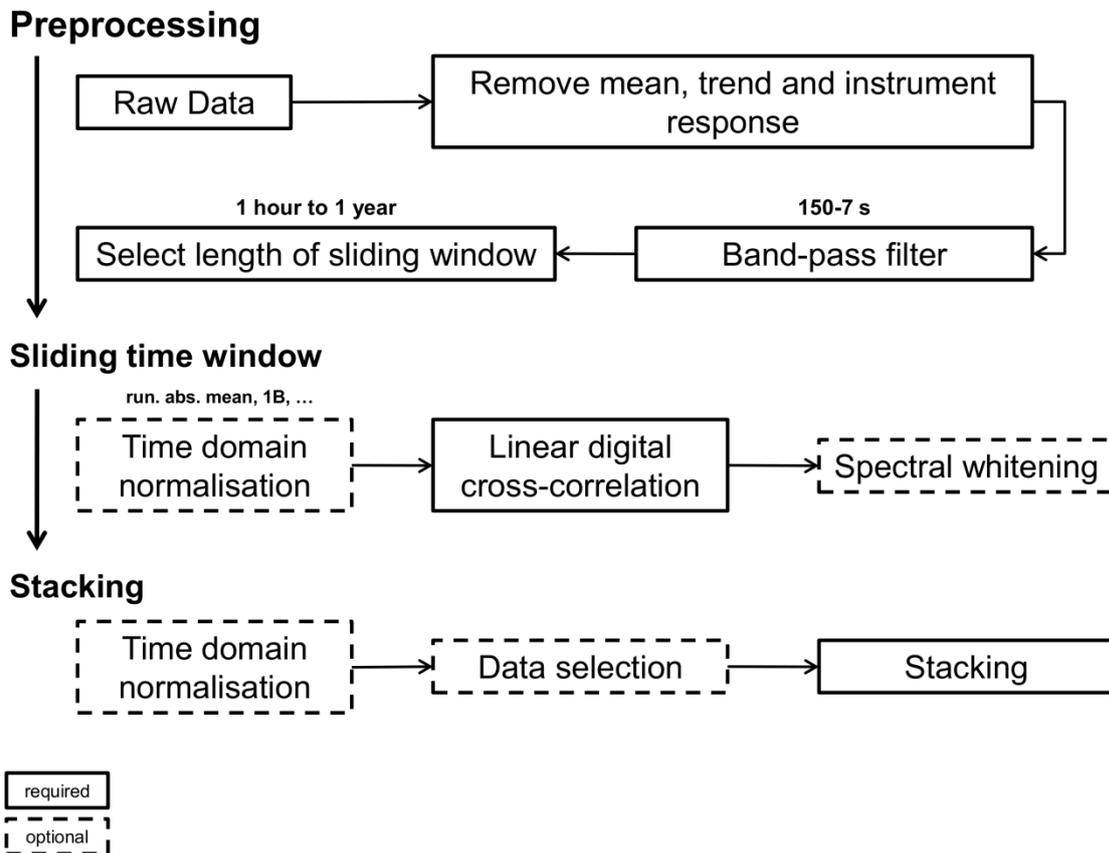


Figure 6.2: Processing scheme for the cross-correlation processing realised for this thesis. For the analysis presented in this thesis 12-months time series are preprocessed to simplify the application of sliding time windows with different lengths. Only the time domain normalisations of the time series are applied prior to the cross-correlation. The spectral whitening is applied without limitations after the cross-correlation for practical reasons as discussed in section 6.2. The data selection is realised by the exclusion of rejected CCFs (see the data selection approaches in section 6.3) from the stacking. The new developed waveform preserving time domain normalisation of the CCFs (wpcf normalisation) is introduced in section 6.4.

Furthermore eight very long time windows of 2 days (d), 4 d (96 hr), 8 d (192 hr), 16 d (384 hr), 32 d (768 hr), 64 d (1536 hr), 182 d (4368 hr) and 1 yr (8760 hr) are used to analyse the influence of the stacking on the obtained 12-months CCFs. In general the

stacked CCFs obtained with a 24 hour time window are used as the ‘reference CCFs’ in reference to the standard data processing proposed by Bensen et al. (2007).

All normalisation schemes realised for this thesis with the data processing scheme illustrated in Figure 6.2 are summarised in Table 6.1. The waveform preserving normalisations of the time series (wpts) and the CCFs (wpcf) are introduced in section 6.4. The numerical differences between the spectral whitening of the time series (TSSW) and the CCFs (SW) are discussed in section 6.2 together with the strategy to apply the one-bit normalisation after the spectral whitening of the time series (TSSW-1B). The digital cross-correlation (XC) is discussed in the following subsection.

Name	TSSW	1B	ram	wpts	TSSW	XC	SW	wpcf	STACK	Relevant figures (6.x)
raw						X			X	7a, 8a
1B		X				X			X	7c, 9, 10
ram			X			X			X	5a, 7b, 8b, 9-16
wpts				X		X			X	
wpcf						X		X	X	8c, 9-16
TSSW-1B	X	X				X			X	4
1B-TSSW		X			X	X			X	4
ram-TSSW			X		X	X			X	4
1B-SW		X				X	X		X	1a, 4, 9, 10
ram-SW			X			X	X		X	1a, 4, 5, 6, 9, 10, 12-15
SW						X	X		X	9, 10, 12-15

Table 6.1: Overview about the applied normalisation schemes and the relevant figures.

First column gives the abbreviation used in the text for the individual applied processing (processing flow from left to right). The abbreviations stand for: time series spectral whitening (TSSW; section 6.2), one-bit normalisation (1B; Shapiro et al., 2005), running absolute mean normalisation (ram; Bensen et al., 2007), waveform preserving time series normalisation (wpts; section 6.4), unbiased linear digital cross-correlation (XC; following subsection), spectral whitening of cross-correlation function (SW; section 6.2) and waveform preserving normalisation of cross-correlation function (wpcf; section 6.4).

6.1.3 The digital cross-correlation

The CCF is calculated as linear unbiased digital cross-correlation in the frequency domain (Bendat & Piersol, 1994). The more intuitive time domain representation of the linear unbiased CCF $R_{xy}(r)$ at lag r of the two random time series x and y with N samples is given as

$$R_{xy}(r) = \frac{1}{N - |r|} \sum_{n=1}^{N-r} x_{n+r} y_n^* \quad (6.1)$$

By this definition signal x is shifted against signal y . The samples of time series x which are shifted beyond the beginning or end of time series y are excluded from the summation. To calculate the linear digital CCF in the frequency domain both time series have to be zero padded to a length of $2N-1$ samples prior to the DFT (Bendat & Piersol, 1994). Without this zero padding a 'wrap-around', or 'circular', effect occurs and the end of time series x is effectively cross-correlated with the beginning of time series y and vice versa. Bendat & Piersol (1994) state, that this circular effect is no serious problem for maximum lag times up to 10% of the total signal length. By calculating the circular digital CCFs the minimum time window length would be limited to around 20,000 s (~6 hours) for the maximum lag times around 2000 s usually necessary for continental scale CCFs. The normalisation term $1/(N-|r|)$ is applied to account for the decreasing number of sample multiplications contributing to $R_{xy}(r)$ with increasing lag r . This normalisation avoids effectively an amplitude decrease (sometimes observed as ramp-like artefact) in the linear digital CCFs with increasing lag r .

The stacked linear cross-correlations are calculated with a sliding overlapping time window. The benefit of the stacked CCF calculation with overlapping time windows is discussed in section 6.1.5. The sliding time window overlap equals the maximum analysed lag time of the calculated CCFs (here 2000 s) to include all possible sample combinations for every lag time between 0 s and 2000 s. A standard maximum lag time of 2000 s is selected to include the signal time windows of all station pairs and to simplify data handling.

Concluding, the 'Total Number of Sample Multiplications' (TNSM) is introduced for the qualitative evaluation of the computational costs for stacked CCFs obtained with different time window lengths. The TNSM of a stacked CCF is calculated by summing up the number of sample multiplications (in fact the normalisation term described above) from the negative to the positive maximum analysed lag time for a single time window CCF and multiplying this sum with the number of stacked CCFs. The TNSM is not equal to the true number of arithmetic operations needed to calculate the stacked CCF! It is used only as a more intuitive measure to compare the computational costs caused by the different time window lengths in section 6.1.5.

6.1.4 Data parameters

Several parameters of the input time series and the obtained CCFs are determined to allow an automated selection of cross-correlations for the stacking and to evaluate the obtained CCFs. The determined parameters of the time series are the amplitude percentiles representing the 68%-, 95.45%- and 99.73%- amplitude intervals which are used for the noise classification introduced in chapter 4. For the obtained CCFs the causal, acausal and symmetric-component (stack of causal and acausal part of the CCF) signal-to-noise ratios (SNR) are determined (Figure 6.3). The SNR is calculated as the ratio between the peak signal in a signal time window and the 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 km/s to 4.8 km/s. Furthermore, the 'precursory' noise in a time window between lag time zero and the signal time window is used 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 earthquake-generated noise in the CCFs between lag time zero and the signal time window. 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. Furthermore, the potential artificial amplitude decrease towards larger lag times due to the linear digital cross-correlation may bias the calculation of the trailing noise SNR, although the CCFs are normalised as described above. Therefore the precursory noise SNR is used, because it is more suitable to evaluate the effectiveness of the different analysed normalisation methods by suppressing the negative influence of teleseismic earthquakes on continental scale seismic noise CCFs. The noise time window contains 60 % of the CCF and neglects the first and the last 20 % between lag time zero and the signal time window. The signal and noise time windows selected for the analysed station pairs are given in Table 6.2.

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 (parameter WSC in Figure 6.3). The symmetry of a CCF around lag time zero is the simplest criteria to evaluate the suitability of a CCF to estimate the Green's function. From the theoretical point of view one expects symmetric, in terms of waveform and amplitude, CCFs which are identical in their causal and acausal parts (Sabra et al., 2005; see section 2.3.3). Cross-correlations 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). Cross-correlations 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 in sections 6.5 and 6.6 to evaluate the CCFs obtained with the different processing schemes.

Station pair		Dist. /km	Signal time window /s			Noise time window /s		
PFO	HRV	4013	836	-	1672	167	-	669
PFO	ANMO	932	194		388	39		155
CCM	DWPF	1432	298	-	597	60	-	238
ANMO	DWPF	2487	518	-	1036	104	-	414
ANMO	CCM	1404	293	-	585	59	-	234
DWPF	HRV	1829	381	-	762	76	-	305

Table 6.2: Interstation distances with corresponding signal and noise time windows of several station pairs.

The signal and noise time windows are used to calculate the SNR of the causal, acausal and symmetric-component CCF. Furthermore, the linear correlation coefficient between the causal and the acausal signal time window is calculated to evaluate the waveform symmetry of the CCF. The signal time windows are selected to contain the Rayleigh waves with propagation speeds between 2.4 km/s and 4.8 km/s. The noise time window covers 60% of the CCF between lag time zero and the signal time window. The six station pairs shown here are highlighted in the analyses in sections 6.5 and 6.6 and are especially indicated in Figure 3.2.

In Figure 6.3 two 12-months stacked CCFs of the station pair PFO-HRV are shown which were obtained with two different time domain normalisations and consecutive spectral

whitening. The comparison of the SNR and WSC of the CCF obtained with the one-bit normalisation (Figure 6.3a) and the CCF obtained with the running absolute mean (ram) normalisation (Figure 6.3b) reveals measurable differences (Figure 6.3c) between both processing schemes. The higher waveform symmetry of the running absolute mean CCF is illustrated by the larger WSC of 0.35 in comparison to 0.22 of the one-bit CCF (Figure 6.3). This is caused mainly by differences in the low frequency content of the CCFs which can be observed directly from the waveforms especially at the beginning of the signal time window. The low frequency signals with low spectral amplitudes in the original time series are not adequately represented in the one-bit normalised time series. This 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 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) to reduce the computational costs and therefore processing time. Another solution to at least partly solve this problem is the application of the one-bit normalisation after the spectral whitening of the time series as done by Brenguier et al. (2008a). This less common approach is discussed together with other spectral whitening strategies in section 6.2. Nevertheless, it is not uncommon to calculate broad-band CCFs with one-bit normalised time series without precursory spectral whitening. Therefore, the one-bit normalisation of the original broad-band time series is included in this study none the less the known shortcomings.

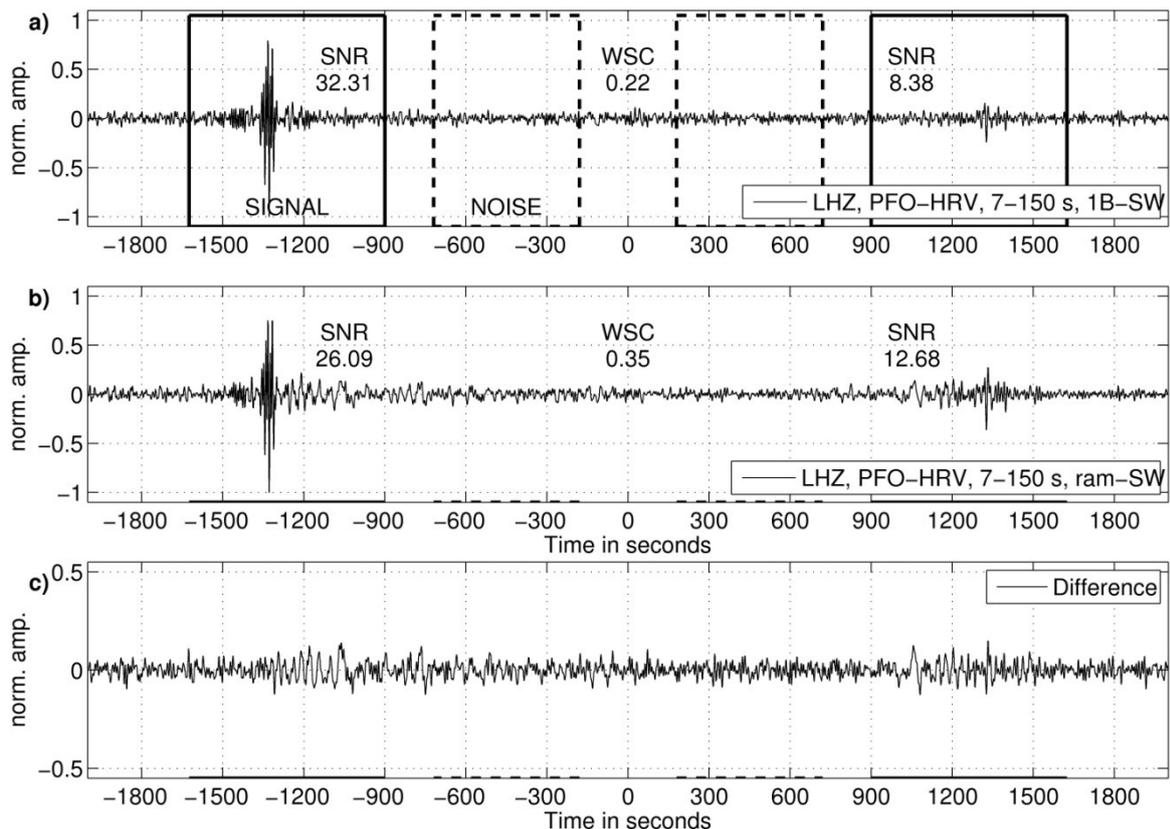


Figure 6.3: A comparison of CCFs obtained from the same data with different normalisation schemes. Two stacked 12-months CCFs for the station pair PFO-HRV obtained with (a) one-bit (1B) and (b) running absolute mean (ram) normalisation. In both cases the spectral whitening (sw) normalisation is applied after the cross-correlation prior to the stacking. The difference time series between both stacked CCFs is shown with a larger scale in (c). 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 (a) and bars (b + c). The SNR is given for the corresponding causal or acausal signal time window of the CCF. The WSC of the CCF is given at lag time zero.

6.1.5 Time window length

This subsection provides the discussion of an important numerical aspect of cross-correlation data processing. The time window length used to calculate stacked CCFs of very long time series. The time window length influences the result of the linear digital cross-correlation as well as the different normalisation methods and it should be chosen carefully. Furthermore, the number of time series samples which do not contribute to the stacked CCF increases with decreasing time window length if a non-overlapping sliding time window is used. It is demonstrated below that overlapping time windows avoid negative influences on the stacked CCFs especially for short time window lengths effectively.

One-bit normalised time series are used to analyse only the influence of the time window length. The one-bit normalisation is fully independent of the sliding time window and ensures the occurrence of clearly visible signals in the CCFs which can be compared to evaluate the influence of the time window length. The linear CCFs are calculated in sliding time windows with lengths between 40 minutes and 24 hours with and without an overlap of 2000 s (33 min 20 s) which corresponds to the largest analysed lag time. The single time window CCFs for each time window length are stacked to obtain the 12-months CCF. These stacked 12-months CCFs are compared against the 'unstacked' 12-months CCF calculated directly from the complete time series (1 yr time window). Only the signal time window of the symmetric-component CCF is considered. In this way, only variations of these signals are evaluated which are commonly derived from seismic noise cross-correlation functions for dispersion analyses (e.g. Bensen et al., 2007). The results for overlapping and non-overlapping time windows are shown in Figure 6.4. The linear correlation coefficient between the symmetric-component signal time windows of the stacked and the reference (time window length of one year) CCFs is larger than 0.998 for all used overlapping time window lengths between 40 minutes and 24 hours (Figure 6.4a). With non-overlapping time windows such high values (>0.998) are only obtained for time window lengths larger than 4 hours (Figure 6.4a). Towards shorter time window lengths, the values decrease rapidly to 0.985 at a time window length of 40 minutes. This effect is related to the decreasing amount of sample multiplications contributing to the linear digital cross-correlation with shorter non-overlapping time windows. This effect is illustrated also by the total number of multiplications in Figure 6.4b. Overlapping time windows can be used to avoid this artefact by the calculation of stacked CCFs with short time windows. Nevertheless, the computational costs properly explode for overlapping time windows shorter than one hour compared to the calculation of the unstacked reference CCF (Figure 6.4b).

Concluding, a minimum time window length in the range of four to six times the maximum analysed lag time is recommendable to obtain reliable results with non-overlapping time windows. The usage of overlapping time windows improves the stacked CCF as approximation of the complete unstacked reference CCF but increases the computational costs. With overlapping time windows even short time windows barely longer than the maximum analysed lag time can be used. The author decided to apply overlapping sliding time windows of one hour and longer to obtain reliably stacked CCFs and to evaluate the influences of the different normalisation methods in combination with different time window lengths with still acceptable computational costs.

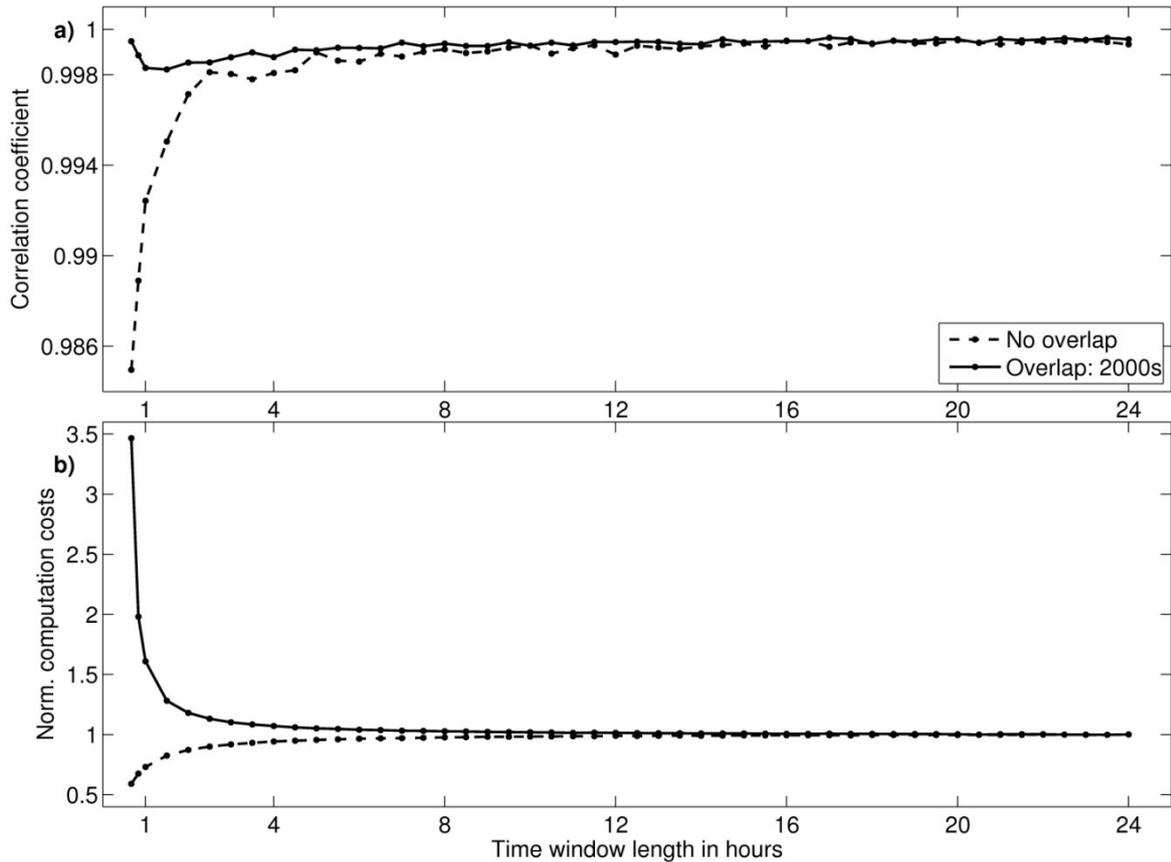


Figure 6.4: Comparison of overlapping and non-overlapping time windows and their computational costs. (a) Linear correlation-coefficients for the symmetric-component signal time windows of the stacked 12-months CCFs with a reference CCF. The reference CCF is the 'unstacked' CCF obtained without stacking by the cross-correlation of the complete 12-months one-bit normalised time series. (b) Normalised total number of multiplications (see section 6.1.3) necessary to derive the stacked CCFs against time window length. A value of one corresponds to the number of multiplications to derive the unstacked 12-months reference CCF.

6.1.6 Further processing of the CCFs

The single broad-band (7-150 s) CCFs are stacked in the time domain to obtain the 12-months CCF. Prior to the stacking the mean value of the single CCFs is removed. To analyse the broad-band CCF in a narrower frequency band a corresponding fourth-order zero-phase Butterworth band-pass filter is applied to the stacked 12-months CCF. The broad-band stacked CCFs are analysed also in the narrower frequency bands 7-14 s, 20-50 s and 70-150 s to evaluate the frequency-dependent influence of the different processing schemes (see sections 6.2.1 and 6.5).

6.2 Spectral whitening

The spectral whitening of the single time window time series or CCFs is a massive non-linear intervention into the signal averaging with the intention to improve the finally obtained CCFs in terms of being broad-band Green's function estimates (Bensen et al., 2007). In this section several common strategies are discussed how to include spectral whitening into the seismic noise cross-correlation processing. Furthermore the immense influence of the time window length on the CCFs obtained with spectral whitening is demonstrated and discussed.

Independent from the strategy the spectral whitening itself is applied as proposed by Brenguier et al. (2008a). The spectral whitening is done by normalising the complex spectrum of a time series or CCF to an absolute value of one in the period range of interest (here 7-150 s) and to zero outside this period range.

6.2.1 Strategies

The typically applied strategy is the spectral whitening of both input time series prior to the cross correlation after a precursory time domain normalisation of the input time series (Bensen et al., 2007). A less widespread second strategy is the spectral whitening of both input time series followed by a one-bit normalisation in the time domain (Brenguier et al., 2008a; Meier et al., 2010). A third strategy, also mentioned by Bensen et al. (2007), is the spectral whitening of the single CCFs after cross-correlation prior to the stacking.

The most important advantage of the third strategy is the preservation of the single CCFs without spectral whitening. It leaves the spectral whitening as an option and allows one to analyse the single CCFs and the stacked CCF with and without spectral whitening efficiently with low computational costs. Furthermore, it is not uncommon to shorten the CCFs directly after the cross-correlation to the largest necessary lag time range (here: -2000 s to +2000 s) to save disk space and to simplify data handling (e.g. Bensen et al., 2007). This shortening of the CCFs can be advantageous also in terms of computational costs of the spectral whitening depending on the length of the input time series. Nevertheless, the spectral whitening of the shortened CCFs is numerical not the same as the spectral whitening of the longer input time series or the original CCFs due to the different frequency resolutions. The different strategies are compared in the following.

The stacked 12-months CCF (2004) are calculated for the station pair PFO-HRV with time window lengths between one hour and one year and overlapping time windows. Thereby the one-bit (1B) as well as the running absolute mean (ram) time domain normalisations are used followed by a consecutive spectral whitening of the time series (1B/ram-TSSW) or the CCFs (1B/ram-SW). Furthermore, the stacked 12-months CCF are calculated by spectral whitening of the time series and consecutive one-bit normalisation (TSSW-1B). Again, the linear correlation coefficient CC of the symmetric-component signal time windows is used to compare the CCFs obtained from the same data but with different normalisation schemes with each other.

The first analysis shown in Figure 6.5a compares the CCFs obtained with the different time window lengths with their 24 hr reference CCF for every normalisation scheme/spectral whitening strategy. This analysis reveals in general a measurable and systematic influence of the time window length on the CCFs obtained with spectral whitening independent of the normalisation scheme. The variations are largest with the TSSW-1B approach and smallest with the ram-SW approach. With all approaches the differences increase towards time window lengths larger than 48 hr and shorter than 12 hr. Between 12 hr and 48 hr the largest CC values are observed around 0.96 (TSSW-1B), 0.97 (1B-TSSW), 0.98 (ram-TSSW, 1B-SW) and 0.99 (ram-SW).

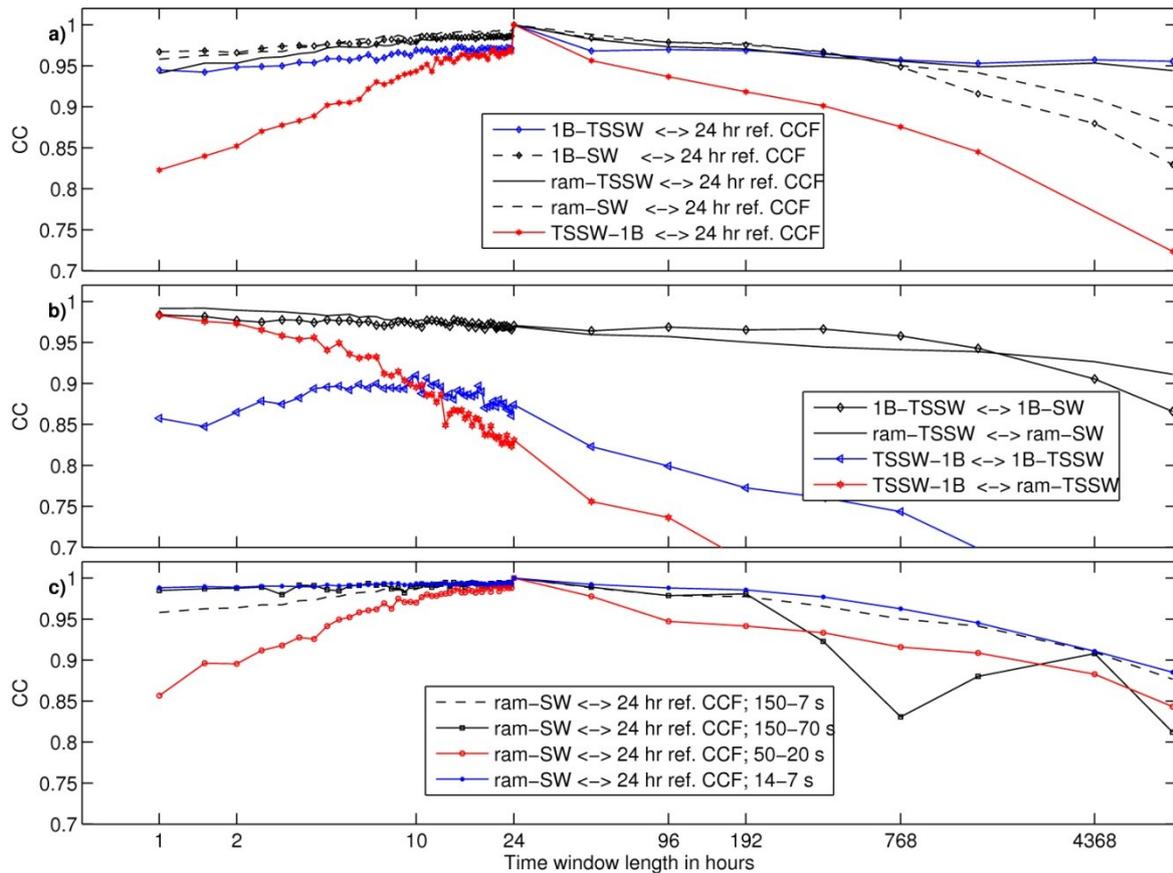


Figure 6.5: Influence of the time window length on CCFs obtained with different spectral whitening strategies. Comparison of CCFs (PFO-HRV, 7-150 s) obtained with different spectral whitening strategies and different lengths of overlapping time windows. Correlation coefficients (CC) between signal time windows of symmetric-component CCFs are shown against time window length. The CCFs are obtained with spectral whitening of the time series (TSSW) or the CCFs (SW) and with one-bit (1B) or running absolute mean (ram) time domain normalisation. The 1B-normalisation is applied before (1B-TSSW) or after (TSSW-1B) the spectral whitening. (a) Comparison of CCFs (7-150 s) with their identical normalised 24 hr reference CCF over time window length for different normalisation schemes. (b) Direct comparison of CCFs (7-150 s) obtained with different normalisation schemes over time window length. (c) Comparison of CCFs (ram-SW) with their 24 hr reference CCF over time window lengths for different frequency ranges. The dashed line in (c) equals the dashed line in (a).

In general, smaller variations are observed with the SW strategy which may be due to the unchanged frequency resolution of the spectral whitening with different time window lengths. Furthermore, the differences between the CCFs increasing with time window length are larger if the one-bit normalisation is applied after the spectral normalisation (TSSW-1B).

Figure 6.5b shows direct comparisons of CCFs obtained with different normalisation schemes. The differences between the CCFs obtained with spectral whitening of the time series (ram-TSSW; 1B-TSSW) and the CCFs obtained with spectral whitening of the CCFs (ram-SW; 1B-SW) decrease systematically with decreasing time window length (black lines). This effect is most probably caused by the converging frequency resolutions of both methods. The comparison of the TSSW-1B strategy with the ram-TSSW strategy (red line with asterisks) reveals decreasing differences with shorter time window lengths also between these two strategies. The large differences between the TSSW-1B and the 1B-TSSW strategy (blue line in Figure 6.5b) are caused mainly by the insufficient representation of the low frequency signals in the CCF if the one-bit normalisation is applied to the original broad-band time series prior to the spectral whitening. Therefore the TSSW-1B approach is more suitable than the 1B-TSSW/SW strategies to produce broad-band Green's function estimates.

6.2.2 Amplification of persistent monochromatic signals

The variations of the CCFs with decreasing time window lengths shorter than 12 hr observed with all normalisation schemes (Figure 6.5a) are a frequency dependent effect. A comparison of ram-SW CCFs with the 24 hr ram-SW CCF is shown for several frequency bands in Figure 6.5c. For this comparison the ram instead of the 1B normalisation is applied to the time series of stations PFO and HRV to minimise unwanted changes to the frequency content of the broad-band CCFs by the time domain normalisation. Furthermore the spectral whitening is applied to the CCFs (SW) to exclude a changing frequency resolution of the spectral whitening due to the different time window lengths as a cause.

The comparison with the broad-band CCFs (7-150 s) is shown in Figure 6.5c as in Figure 6.5a by the dashed line. The solid lines with markers (Figure 6.5c) show the results for the short period (7-14 s; asterisk), intermediate period (20-50 s; circle) and long period (150-70 s; square) CCFs. Increasing waveform differences towards shorter time windows are not evident at short and long periods but pronounced in the frequency band 20-50 s. This effect is caused by a persistent narrow-band microseism signal at 26 s which is discussed in detail below. In contrast, the increasing differences between time window lengths larger and shorter than 48 hr are observed in all frequency bands. Exactly the same behaviour is observed with all other analysed station pairs.

In Figure 6.6a the 12-months 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) and running absolute mean normalisation. The same CCFs obtained with additional spectral whitening are shown in Figure 6.6b. The waveform differences between the 2 hr-CCFs and the 24 hr-CCFs are given in Figure 6.6c and Figure 6.6d, respectively. The amplitude spectra of the CCFs are shown in Figure 6.6e and Figure 6.6f.

No significant differences between the 2 hr-CCF and the 24 hr-CCF are observed in the waveforms or the amplitude spectra of the CCFs if spectral whitening is not applied (left side of Figure 6.6). With spectral whitening a strong monochromatic signal with a period of approx. 26.33 s emerges in the causal part of the 2 hr-CCF (right side of Figure 6.6). A signal at the same frequency is also present in the other CCFs but with significantly smaller relative amplitudes (compare the amplitude spectra in Figure 6.6e and Figure 6.6f).

As aspired, the spectral whitening influences massively the signal averaging effect of the cross-correlation and produces a more broad-band Green's function estimate. Undesired is the amplification of a monochromatic signal not related to the Green's function. Furthermore, the results obtained with spectral whitening are significantly influenced by the time window length. Thus, the time window length is an important parameter of the seismic noise cross-correlation processing.

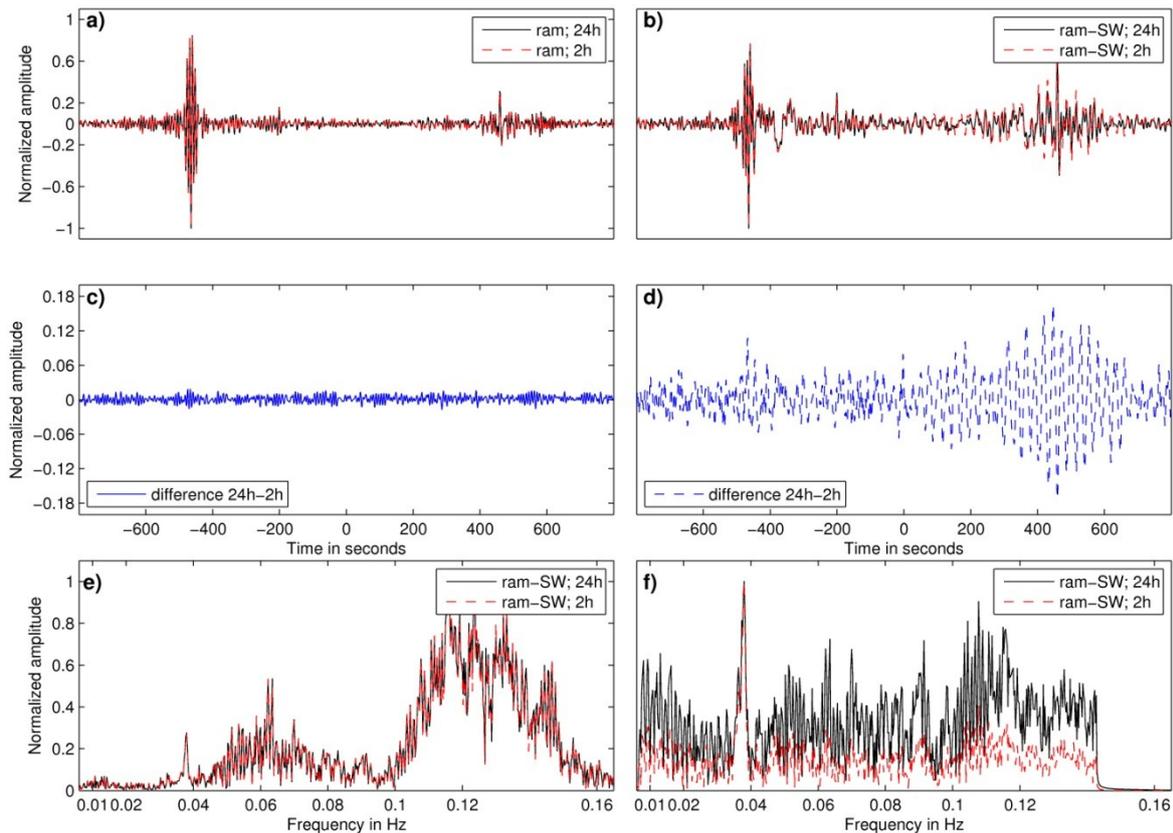


Figure 6.6: Influence of the time window length on CCFs obtained with spectral whitening. Comparison of 12-months CCFs (ANMO-CCM, 7-150 s) obtained with different normalisation schemes and time window lengths. (a) ram-CCFs obtained with a time window length of 24 hours (black line) and 2 hours (red dashed 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.

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) identify a monochromatic source of microseism located in the Gulf of Guinea to most probably cause the 26 s signal in the CCFs. They observe the 26 s signal to be very persistent in time with a seasonal amplitude variation with larger amplitudes in the northern hemisphere summer months. Nevertheless, the amplitude information is discarded by the spectral whitening and only the phase information is retained. This leads to an amplification of temporally stable persistent signals even with small relative amplitudes if CCFs are stacked after spectral whitening. Although the amplitude spectra of the single CCFs are flat after spectral whitening, the amplitude spectrum of the stacked 12-months CCF is not. The amplitude spectrum of the stacked CCF is related to the coherence of the single CCFs and is shaped by the stacking process.

A histogram based analysis of the phase information at the discrete frequencies of the complex spectra of the single CCFs is suitable to further investigate the emergence of the 26 s signal with decreasing time window length. For this analysis the single time window CCFs of the year 2004 and station pair ANMO-CCM are used. The complex values at a given discrete frequency obtained from all single CCFs are located on the unit circle and should group in a limited angular range of the unit circle if a stable signal is present in the single CCFs. Therefore, the unit circle is divided in 12 angular bins of equal size between $-\pi$ and π to calculate histograms of the phase angles of the single CCFs at every discrete frequency (Figure 6.7). If the single CCFs contain no dominant signal the complex values can be expected to be uniformly distributed over the unit circle with 8.3 % of the complex values in every of the 12 bins.

In Figure 6.7a, the histogram of phase angles at the peak period 26.33 s is shown for the 24 hr CCFs. The phase angles are not uniformly distributed and group at angles between $-\pi/2$ and $-\pi$. In general, the same is observed for the 2 hr-CCFs (Figure 6.7b). These histograms are obtained for all discrete frequencies to determine the standard deviation of the occurrence. The standard deviations are plotted over frequency in Figure 6.7c for the 24 hr-CCFs and in Figure 6.7d for the 2 hr-CCFs. The standard deviation peaks also at the period of 26.33 s indicating the stability of the 26 s signal also in the shorter 2 hr-CCFs. In general, the shape of the standard deviation over frequency equals the shape of the amplitude spectra of the stacked CCFs obtained with 24 hr and 2 hr time windows, respectively. The standard deviation of the 2 hr-CCFs (Figure 6.7d) is considerably smaller than for the 24 hr-CCFs (Figure 6.7c) indicating that the phase angles are more uniformly distributed at all frequencies. With the 2 hr time window the standard deviation at 26.33 s is smaller in absolute value but more pronounced in comparison to the other frequencies than with the 24 hr time window. This is also observed with a smaller sub-sample of 374 randomly selected 2 hr-CCFs.

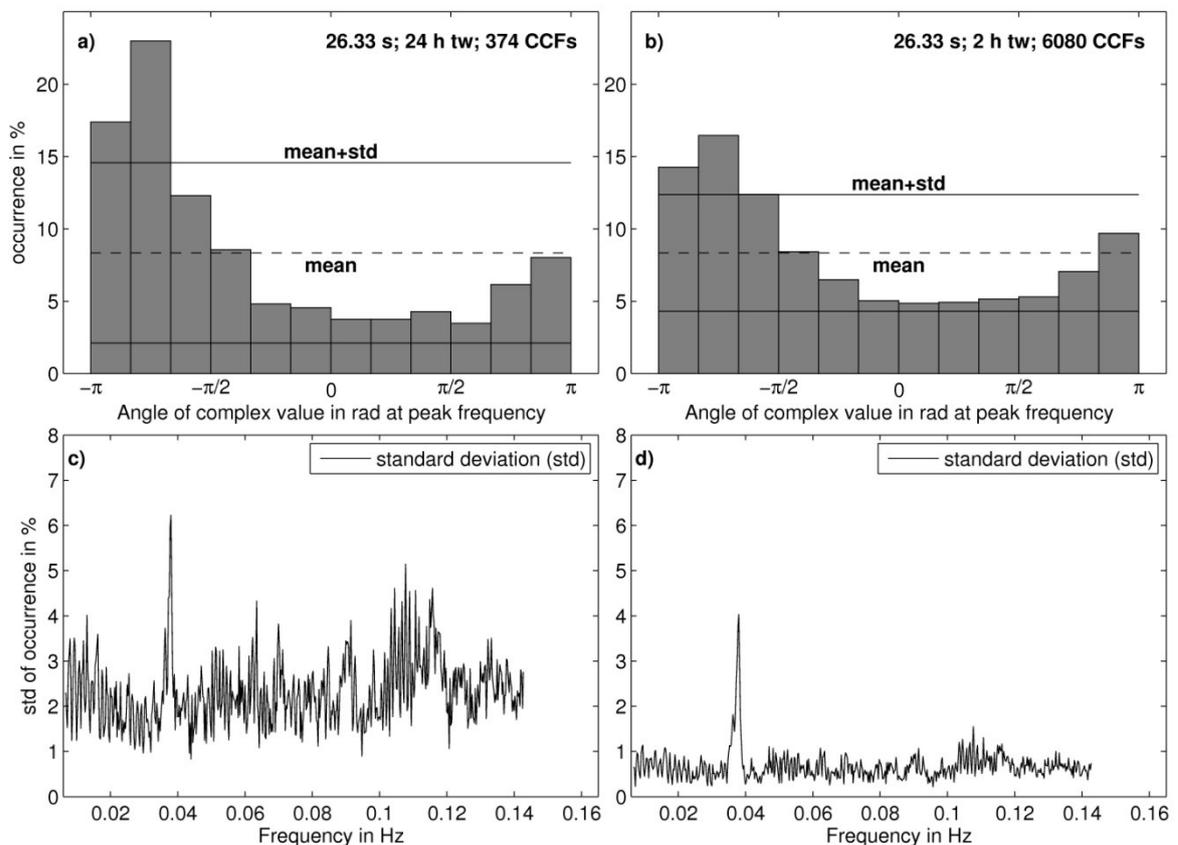


Figure 6.7: Histogram-based spectral analysis of single time window CCFs.

The analysis illustrates the phase information in the complex spectra of single 24 hr and 2 hr time window CCFs of the station pair ANMO-CCM. In (a) the histogram of the phase angles at the period of the 26 s microseism signal for all 24 hr time window CCFs (overlapping time windows) of the year 2004 is shown. In (b) the corresponding histogram for the 2 hr time window CCFs is shown. The mean of the histograms are shown by the dashed lines in the histograms. The solid lines in (a) and (b) indicate the range of the standard deviation around the mean. The standard deviations of the phase angle histograms of all discrete frequencies are plotted in (c) for the 24 hr CCFs and in (d) for the 2 hr time window.

6.2.3 Discussion

The influence of time window length on the results obtained with spectral whitening needs further research. Therefore, only a first interpretation of the observations is given below.

Regarding the increasing dominance of the 26 s signal with decreasing time window length the author assumes two underlying aspects. On the one hand, the temporally very persistent signal surely takes profit from the higher summation order with the shorter time

window length. On the other hand, this is only possible because the signal emerges reliably in the single CCFs, even if a short time window length of 2 hr is used. As the 26 s signal is excited continuously by a localised 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 at once, the averaging of all contributing signals is obtained by the cross-correlation itself. By dividing the long time series in shorter time windows, the averaging is increasingly transferred from the cross-correlation to the stacking process and therefore increasingly affected by the spectral whitening. This would amplify especially temporally persistent signals in the CCFs like the 26 s microseism which efficiently produce signals in the single CCFs. Therefore, the time window length should be carefully selected.

The time window length should be as long as possible to enhance the emergence of the Green's function in the single CCFs independent of the stacking. Nevertheless, a stacking of CCFs after spectral whitening is recommendable. The stacking ensures the cancellation of incoherent noise in the single CCFs with flat amplitude spectra which is also amplified by the spectral whitening. This effect is illustrated by the shaping of the amplitude spectrum during the stacking as shown in Figure 6.7.

Following these implications and observations, the time window lengths shorter than 10 hours (see also Figure 6.5a and c) are increasingly not suitable to apply spectral whitening to the data used for this study. Furthermore, the very long time windows of several days to weeks seem to be too long to ensure an effective cancellation of the noise introduced by the spectral whitening with one year of data. Nevertheless, the results obtained with the different normalisation schemes with and without spectral whitening are discussed in the following sections including the short time windows to evaluate the performance of the data selection approaches and the waveform preserving time domain normalisations.

A possible solution which would allow us to use short time windows (e.g. to introduce an automated data selection) and to avoid the amplification of monochromatic persistent signals by the spectral whitening of the short time windows may be a stacking process with two stages. In the first stage several CCFs obtained from short time windows can be stacked without spectral whitening to obtain a 'pre-stacked' CCF representing a longer time window to enhance the emergence of the Green's function. These pre-stacked CCFs could be spectral whitened and stacked afterwards to obtain the final stacked CCF to estimate the broad-band Green's function.

Concluding, the author recommends and applies the spectral whitening of the CCFs after the cross-correlation and prior to the stacking. This strategy has two advantages. First, the obtained CCFs can be analysed with and without spectral whitening at low computational costs. Second, this strategy would allow a stacking process with two stages which may be suitable to combine very short time windows reasonably with spectral whitening. This topic needs further research and is not covered by this thesis. The strategy TSSW-1B which is proposed by Brenguier et al. (2008a) is not further considered for two reasons. On the one hand different time domain normalisations are combined with spectral whitening and on the other hand the spectral whitening is applied after the cross-correlation for the discussed practical reasons.

6.3 Data selection

The improvement of the stacked CCFs by an automated data selection is tested with two approaches. The first approach considers the characteristics of the two broad-band time series which are cross-correlated and is based on the noise classification introduced in chapter 4. This approach is called time series approach (TSA) in the following. The

second approach considers the waveform symmetry of the single CCFs and uses a criterion based on the waveform symmetry measure WSC discussed in 6.1.4. This approach is called waveform symmetry approach (WSA) in the following.

Time windows with contemporary transient signals at the two recording sites are problematic to obtain suitable seismic noise CCFs due to their amplitude 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. With the time series approach time windows are excluded if at least one of the two time series is classified as corrupt (no data, extreme transients, step in the time series; classified as NC10-NC13) or if both time series are dominated by transient signals (like contemporaneously arriving surface waves; classified as NC3 or NC4).

The aim of the waveform symmetry approach is to exclude strongly asymmetric CCFs before stacking the CCFs of the single time windows to obtain the complete CCF. Therefore, the intention is to reject all CCFs with a negative linear correlation between the causal and acausal signal time window which is represented by a negative WSC value. In Figure 6.8 the histogram of the WSC values is shown for the 24 hr CCFs of the year 2004 obtained with the running absolute mean normalisation for all 10 station pairs. The fraction of non-significant WSC values is shown in red, the fraction of significant WSC values is shown in blue. A linear correlation coefficient is considered to be significant, if the probability to observe a value as large as the observed one for two uncorrelated random time series by chance is smaller than 5 percent. The significance test is done as proposed by Bendat & Piersol (1994). WSC values with a value between -0.07 and 0.07 are observed to be in general not significant in the case of the analysed CCFs. Therefore, a minimum WSC value of 0.07 is defined for the waveform symmetry approach to exclude CCFs with a negative correlation from the stacking.

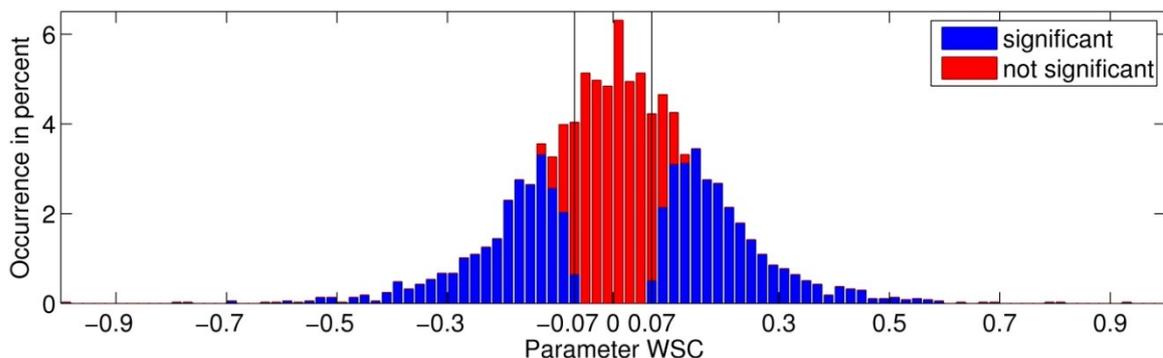


Figure 6.8: Significance of the calculated waveform symmetry coefficient (WSC) values
Histogram of the WSC values of the 24 hr CCFs (7-150 s) of the year 2004 obtained with the running absolute mean normalisation for all 10 station pairs. The fraction of the non-significant WSC values is shown by the red colour. A WSC value is significant (blue) if the probability is smaller than 5 percent to observe a WSC value as large as the observed one for an uncorrelated random time series by chance.

To obtain a band-pass filtered stacked CCF with the waveform symmetry approach, the single CCFs should be filtered in the desired frequency band prior to the determination of the WSC value and the stacking. This is a drawback of the waveform symmetry approach in terms of computational costs. Furthermore, the waveform symmetry approach is observed to be less effective if the spectral whitening was already applied. The single time window CCFs are significantly biased by incoherent noise which is also amplified by the spectral whitening. This effect leads in general to decreased WSC values of the single time window CCFs. Therefore, the selection of the CCFs by their WSC values is done prior to the spectral whitening. The performance of both data selection approaches is discussed in detail in section 6.6.

6.4 Waveform preserving time domain normalisation

As mentioned above, all established normalisation methods intend to equalise the signals contributing to seismic noise in order 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 from seismic noise CCFs is disturbed due to the impacts of dominating seismic signals such as transient earthquakes or monochromatic signals of ocean-generated microseism, even if very long time series are used.

The two most widespread time domain normalisations (1B and ram) are briefly discussed prior to the introduction of the two new waveform preserving normalisation methods wpts and wpcf. In Figure 6.9 stacked 12-months CCFs are shown in the broad period band 7-150 s obtained with an overlapping sliding time window (6 hours) of the station pair PFO-HRV without any normalisation ('raw', Figure 6.9a) and with the running absolute mean (Figure 6.9b) as well as the one-bit (Figure 6.9c) time domain normalisations. No spectral whitening is applied to the shown time series and no time windows with data gaps or spurious signals were excluded. In Figure 6.9a the raw stacked CCF is shown with the largest amplitudes near lag time zero within the 'noise time window'. The SNRs of the causal and acausal parts are therefore very small (< 2). The CCF is strongly asymmetric as represented by the negative WSC value. In Figure 6.9b and Figure 6.9c the CCFs obtained with the running absolute mean normalisation and the one-bit normalisation, respectively, are shown. The waveform difference between the CCFs obtained with the two non-linear time domain normalisations is displayed in Figure 6.9d. Both normalisations work well by improving the emergence of a signal in the stacked CCFs and to improve the symmetry also with the complete data including gaps and instrument irregularities. Unfortunately, the obtained CCFs are still asymmetric in waveform and amplitude and it is therefore arguable if these CCFs should be used to estimate broadband Green's functions at all. Nevertheless, no obvious differences of the broad-band CCFs obtained with the one-bit and the running absolute mean normalisation can be observed from the waveforms without spectral whitening (compare with Figure 6.3) due to the very small relative amplitudes of the low frequency signal content (Figure 6.9e).

The approaches to improve the CCFs with a waveform preserving normalisation benefit from the calculation of stacked CCFs with a sliding time window. The idea is to equalise the amplitude differences between the single CCFs prior to the stacking to improve the emergence of the Green's function. To do so, two methods are tested. The first one is to normalise both time series before the cross-correlation and the second one is to normalise the single CCFs after the cross-correlation prior to the stacking. Both normalisations are done by dividing the waveforms (CCFs or time series, respectively) by an amplitude value and are discussed in detail below.

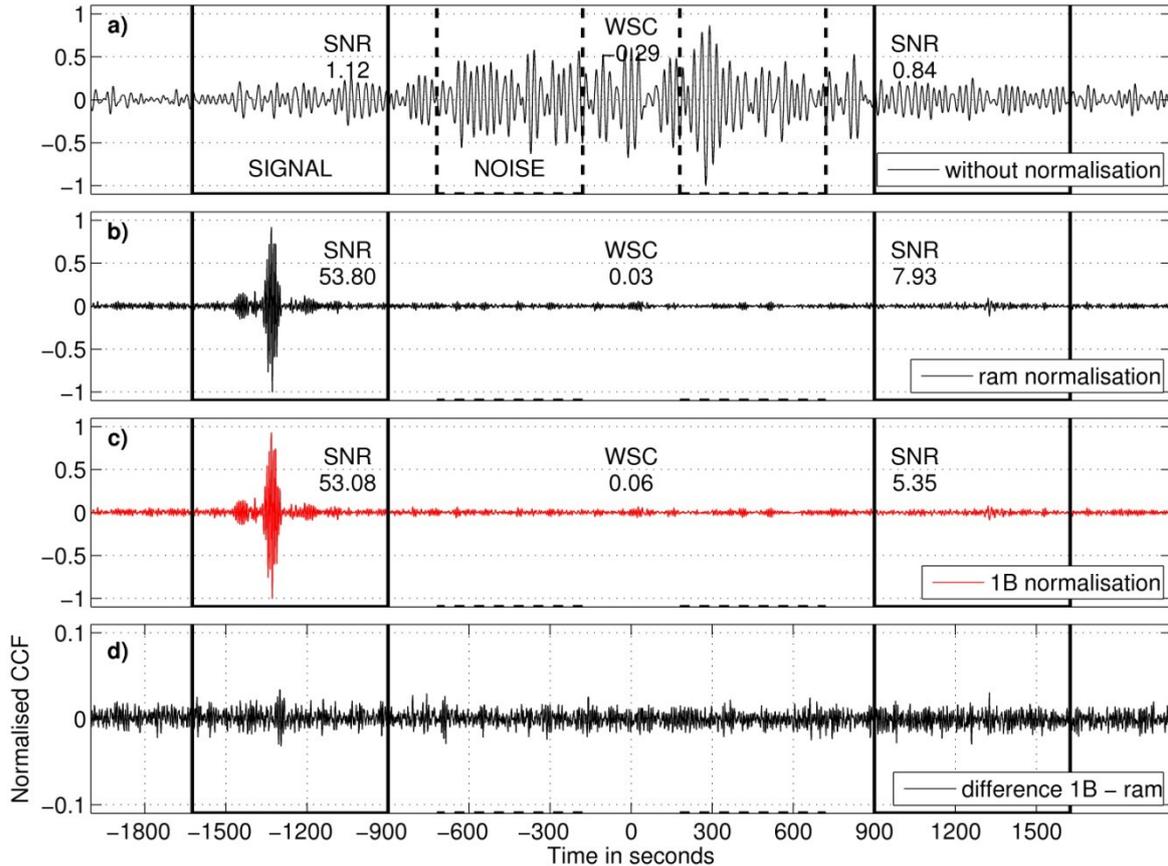


Figure 6.9: Comparison of 12-months CCFs obtained with different time domain normalisations (ram+one bit). The stacked 12-months CCFs (PFO-HRV, 7-150 s) are obtained with (a) none, (b) running absolute mean and (c) one-bit normalisation. No spectral whitening is applied prior to the stacking. The waveform difference between the CCFs obtained with the one-bit (c) and the running absolute mean (b) normalisation is shown in (d) at a larger scale.

6.4.1 Normalisation of the time series before cross-correlation

The waveform preserving normalisation of the two time series (wpts-normalisation) before the cross-correlation is a simple linear normalisation with an amplitude value. The time series are divided by the value of their 68% amplitude interval range which is also used for the time domain classification necessary for the TSA data selection. By this approach 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 single time windows with very large amplitudes is reduced. Nevertheless, single CCFs originating from time windows with strong transient signals at both sites, such as earthquake waves, dominate still the stacked CCF due to their in general larger amplitudes.

6.4.2 Normalisation of the CCFs after cross-correlation

The waveform preserving normalisation of the single CCFs (wpcf-normalisation) after cross-correlation and before the stacking is similar to the method used by Campillo & Paul (2003) who normalise the amplitudes of each CCF with their absolute maximum. This normalisation method is extended with an additional simple procedure to weight the single CCFs prior to the stacking. The root mean square (rms) value of the entire waveform is used to normalise a CCF. To introduce a weighting to the CCF normalisation it is assumed that single 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 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 2 are normalised 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. The upper SNR boundary of 13 is selected based on the observations with the analysed station pairs. No single CCFs with a SNR larger than 13 are observed which are 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.

6.4.3 Discussion

A detailed discussion of the performance of both waveform preserving normalisations is given in sections 6.5 and 6.6. Only a first discussion of some general characteristics of both normalisations is given here.

It is foreseeable that both normalisations are influenced by the time window length and that the best results may be expected with a time window length in the same order of magnitude as the duration of the dominating signals (e.g. teleseismic surface waves). The performance of both normalisations in respect to the time window length is therefore discussed in detail in sections 6.5 and 6.6.

The wpcf-normalisation with weighting proved to be more robust and powerful than the wpts-normalisation. In Figure 6.10 the stacked 12-months CCFs obtained without normalisation (a) as well as obtained with the ram normalisation (b) and the wpcf-normalisation (c) are shown. The waveform difference between the CCFs in (b) and (c) is shown in (d). The wpcf-normalisation (Figure 6.10c) works well by improving the emergence of a signal (SNR >50) in the stacked CCF even without the exclusion of problematic data. The obtained CCF exhibits minor waveform differences (Figure 6.10d) especially in the middle part of the acausal signal time window in comparison to the ram-CCF. At the same time the WSC value of 0.1 is slightly higher than the WSC value of the ram-CCF (0.03) but still very low. This first example shows that the waveform preserving time domain normalisation is in principal capable to determine CCFs which are suitable to estimate Green's functions and which are of the same quality as the CCFs obtained with the running absolute mean or the one-bit normalisation. A further advantage of the wpcf-normalisation is that the normalisation is applied to the CCFs after cross-correlation. That means that no changes to the original data occur prior to the cross-correlation.

In contrast, the wpts-normalisation is observed to fail without data selection due to the dominance of few spurious transient signals with comparably large amplitudes. Most of these spurious signals can be assumed to be 'non seismic' signals such as calibration pulses, instrument irregularities or signals excited in the direct vicinity of the sensor. Therefore, the wpts-normalisation is excluded from the first detailed evaluation in section 6.5 (without data selection) and discussed only in section 6.6 (with data selection).

The last significant aspect of both waveform preserving normalisations is their interaction with the spectral whitening. The wpts-normalisation prior to the spectral whitening is without any effect on the finally obtained stacked CCF. This is also the case for the wpcf-normalisation, if the CCFs are normalised before the spectral whitening. Nevertheless, the wpcf-normalisation could be applied after the spectral whitening as the last processing step prior to the stacking. In this case, the wpcf-normalisation could have an effect on the CCFs obtained with spectral whitening. However, the application of the wpcf-normalisation after the spectral whitening is observed to produce altogether the same results as the spectral whitening without any time domain normalisation. The spectral whitening is already very effective by suppressing extreme transient signals in CCFs and by compensating the overall amplitude differences between the CCFs prior to the stacking.

That means that the waveform preserving time domain normalisations are both neither necessary nor effective if spectral whitening is used. Consequential, the waveform preserving time domain normalisations are not combined with the spectral whitening for the comparison of the normalisation schemes in sections 6.5 and 6.6. Instead, the normalisation of the CCFs only with spectral whitening is included as normalisation scheme. Nevertheless, the waveform preserving time domain normalisations can be included without any difficulty to the cross-correlation processing although if an optional spectral whitening of the CCFs may be applied later on. Furthermore, a reasonable combination of the wpcf normalisation with the spectral whitening can be realised by a stacking process with two stages.

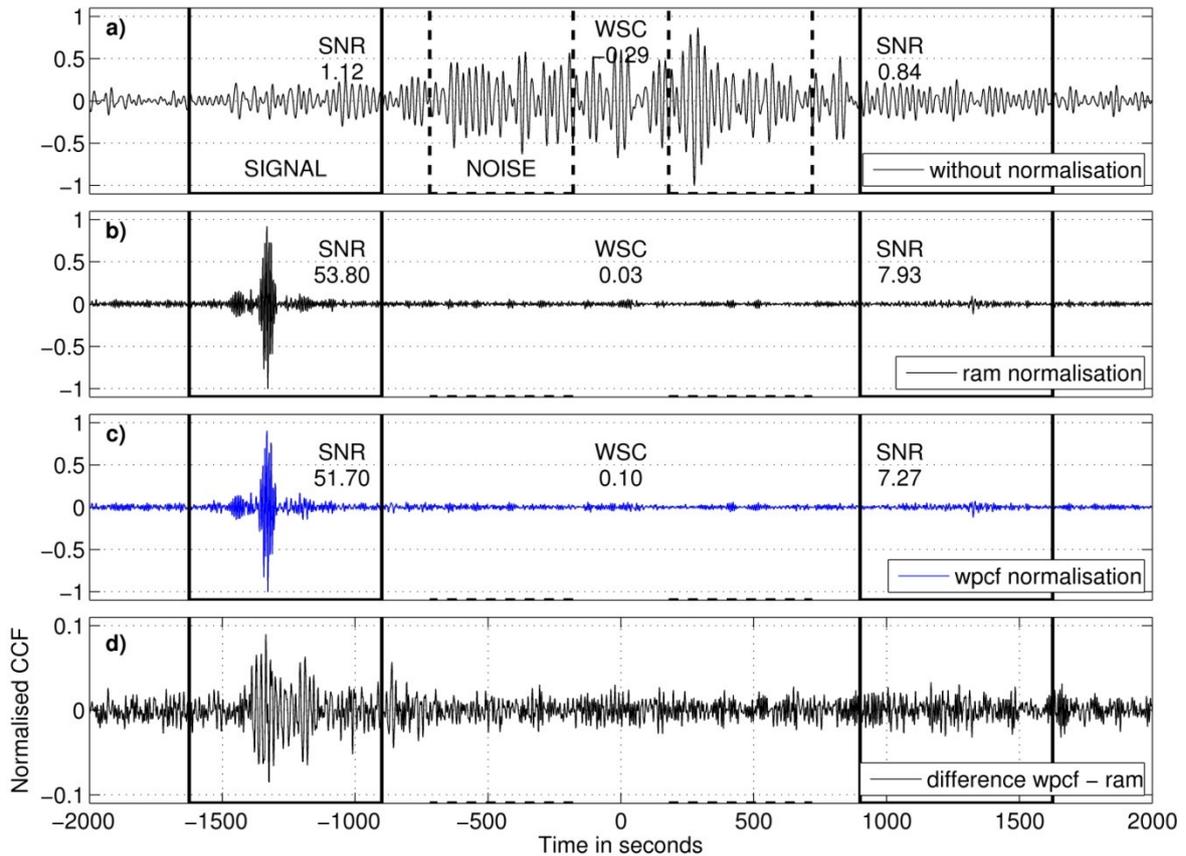


Figure 6.10: Comparison of 12-months CCFs obtained with different time domain normalisations (ram+wpcf). The stacked CCFs (PFO-HRV, 7-150 s) are obtained with (a) none, (b) running absolute mean and (c) waveform preserving CCF normalisation. No spectral whitening is applied prior to the stacking. The waveform difference between the CCFs obtained with the wpcf (c) and the ram (b) normalisation is shown in (d) at a larger scale.

6.5 Evaluation of the different normalisation schemes

The evaluation of the different normalisation schemes is conducted by a comparison of the obtained symmetric-component CCFs with a reference CCF (parameter CC), the waveform symmetry (parameter WSC) and the SNR of the symmetric-component CCFs (parameter SNR). The running absolute mean normalisation is used as the reference time domain normalisation and a 24 hr time window is used as the reference time window length following the recommendations of Bensen et al. (2007). The CCFs obtained with running absolute mean (ram), one-bit (1B) as well as the waveform preserving normalisation of the CCFs (wpcf) are compared with each other. The ram- and the 1B-normalisation are evaluated with and without spectral whitening of the CCFs (ram-SW/1B-SW). Also the stand-alone application of spectral whitening (SW) is analysed. All applied normalisation schemes are summarised in Table 6.1. The complete data set of the year 2004 is used for this first evaluation. Only time windows containing no data at all at one or

both of the stations are excluded. The observations are first described and discussed afterwards.

6.5.1 Observations

The observations are illustrated in the following with the help of the six station pairs indicated by the connection lines in Figure 3.2. The comparisons of the broad-band symmetric-component CCFs (parameter CC) in the period band 7-150s obtained by the different normalisation schemes with the corresponding reference CCFs (ram and ram-SW, 24 hr) are shown for the station pairs PFO-HRV (Figure 6.11a), PFO-ANMO (Figure 6.11b), ANMO-DWPF (Figure 6.11c), DWPF-HR (Figure 6.11d), ANMO-CCM (Figure 6.12a) and DWPF-HRV (Figure 6.12b). The comparisons are shown for the CCFs obtained without (solid lines) and with (dashed lines) spectral whitening. The comparisons for the different station pairs with distances between 932 km and 4013 km and with four different azimuths (Figure 3.2) show similar results.

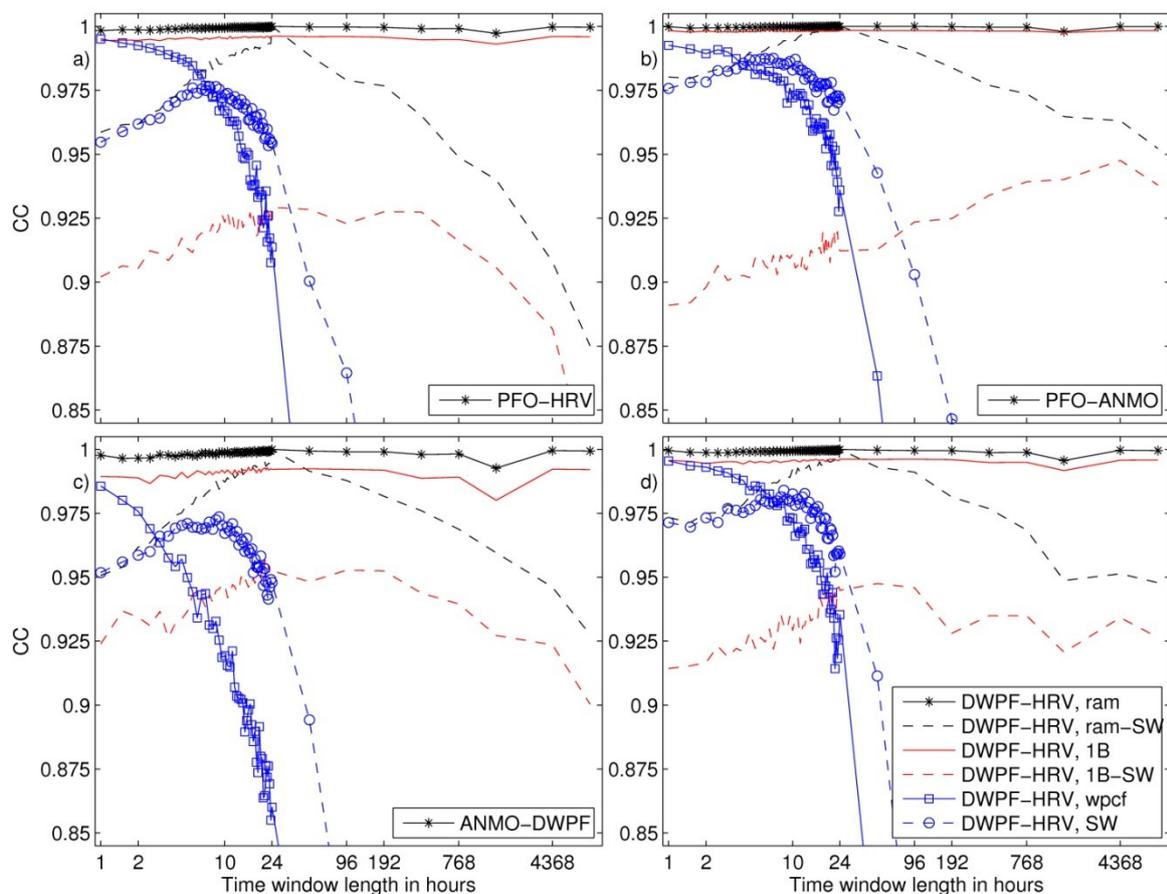


Figure 6.11: Influence of the time window length on the CCFs obtained with different normalisation schemes. A comparison of broad-band (period band 7-150 s) symmetric-component CCFs with a reference CCF (parameter CC) is shown for different normalisation schemes and several station pairs. Only time windows containing no data are excluded from the stacking. The running absolute mean (ram) normalisation is used as reference time domain normalisation and a reference time window length of 24 hours. The comparison is shown for the station pairs PFO-HRV (a), PFO-ANMO (b), ANMO-DWPF (c) and DWPF-HRV (d).

The CCFs obtained with the ram (solid black line with asterisk) or 1B (solid red line) time domain normalisation do not vary with time window length (Figure 6.11). Small waveform differences between the ram- and the 1B-CCFs are observed systematically independent from time window length and for all station pairs (e.g. Figure 6.9). The comparison of both time domain normalisations in combination with spectral whitening (corresponding dashed lines) shows the already discussed time window length dependency of the spectral whitening as well as the already discussed significant waveform differences between the

ram-SW and the 1B-SW CCFs due to the variations of the frequency content caused by the one-bit normalisation (Figure 6.11).

The differences between the wpcf-CCFs and the reference CCF (solid blue line with squares) decrease rapidly with decreasing time window length. The comparison of the SW-CCFs (blue dashed line with circles) as well as the ram-SW-CCFs (black dashed line) with the reference CCF indicates also decreasing waveform differences between these two normalisation schemes with decreasing time window length (Figure 6.11). This is confirmed by a direct comparison of the ram-SW with the SW-CCFs for all time window lengths between one hour and one year (not shown here). In general, no significant waveform differences (correlation coefficients larger than 0.99) between the signal time windows of the SW- and the ram-SW-CCFs can be observed for time window lengths between 1 hour and 6 hours.

In the following the influence of the different normalisation schemes and time window lengths on the waveform symmetry (Figure 6.12c and d) and signal-to-noise ratio (Figure 6.12e and f) of the CCFs is demonstrated. Different absolute WSC and SNR values are observed for the different station pairs but the variations with time window length and the differences of these variations between the normalisation schemes are highly systematic and similar for the different station pairs. Therefore the discussion is constrained to the two representative stations pairs ANMO-CCM and CCM-DWPF with nearly perpendicular azimuths shown in Figure 6.12.

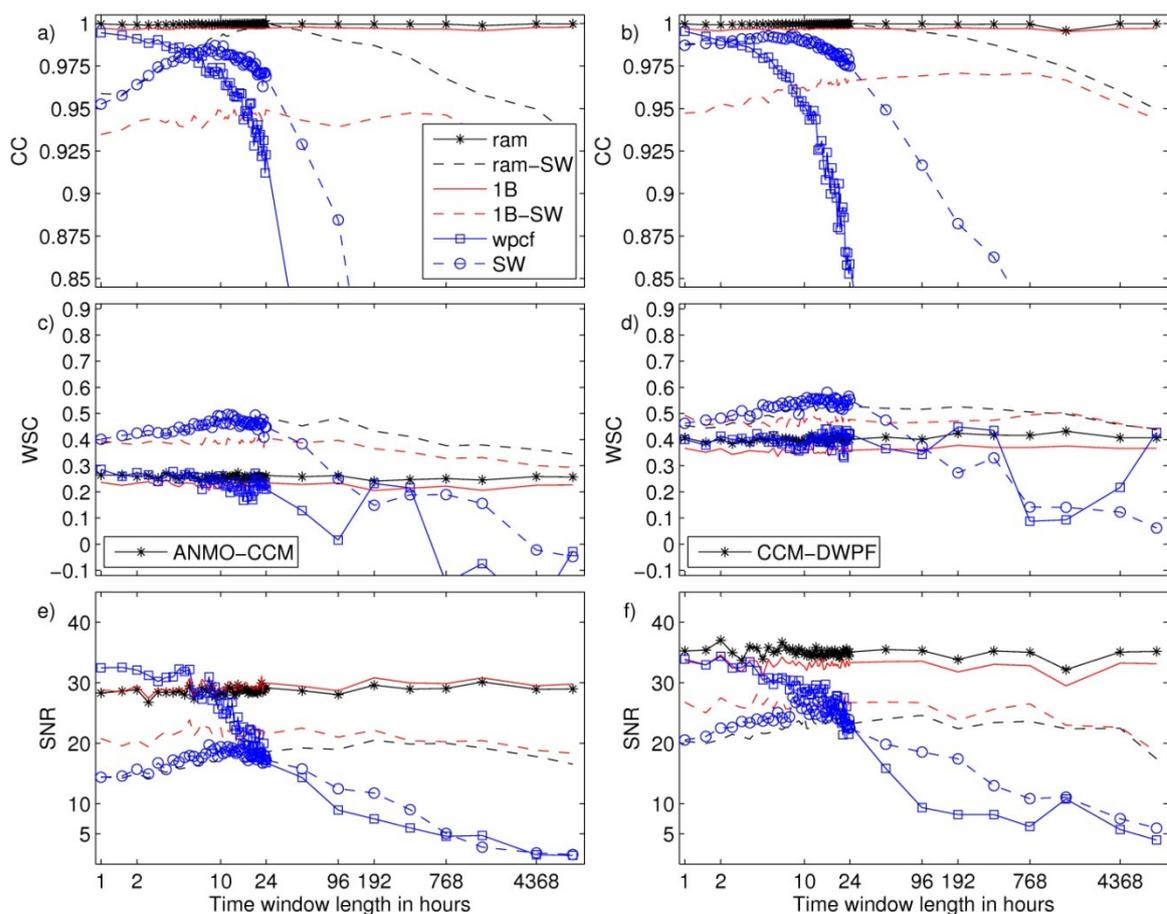


Figure 6.12: Influence of normalisation and time window length on the obtained CCFs (without data selection). A detailed comparison of broad-band CCFs (7-150 s) obtained with different normalisation schemes and time window lengths is shown for the station pairs ANMO-CCM (left side) and CCM-DWPF (right 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.

In general higher waveform symmetries, and therefore higher WSC values, are observed for CCFs obtained with spectral whitening. This is mainly due to the better representation of the low frequency signals in the CCFs after spectral whitening and the observation that the CCFs are very symmetric (WSC values > 0.7) in the long period band 70-150 s (see also the example in Figure 6.3). The WSC values of the broad-band ram-SW- and SW-CCFs (7-150 s) are in general very similar for time window lengths shorter than 24 hr and decrease towards shorter time window lengths. This decrease is directly related to the emergence of the asymmetric 26 s signal in these CCFs. The emergence of the 26 s signal with decreasing time window length is also observed in the 1B-SW-CCFs but less pronounced and with lower relative amplitudes. This effect leads to smaller variations of the in general also lower WSC values of the 1B-SW-CCFs. Without spectral whitening comparable WSC values are observed for the ram- and the wpcf-CCFs which are typically slightly larger than the WSC values of the 1B-CCFs.

The SNR values are also significantly influenced by the spectral whitening (Figure 6.12e and f). Without spectral whitening significantly larger SNR values are observed. Nevertheless, these higher SNRs indicate not mandatory a higher signal quality. The very high SNR values obtained without spectral whitening are caused by very high amplitudes of signals in a narrow frequency band. These large amplitude differences over frequency are intentionally balanced by the spectral whitening as can be seen by a direct comparison of Figure 6.3b and Figure 6.9b. The increasing SNR of the wpcf-CCFs with decreasing time window length is observed to be caused more by decreasing amplitudes in the noise time window than by waveform variations in the signal time window.

6.5.2 Discussion

The differences between the 1B- and the ram-CCFs are rather small. Nevertheless and as already discussed above, the one-bit normalisation is observed to be not adequate to obtain CCFs to estimate broad-band Green's functions due to disadvantageous variations of the signal's frequency content. This is especially true and obvious if spectral whitening is introduced to the normalisation scheme. Two possible solutions to reduce the impact of this problem are a narrowband pre-filtering of the seismic noise time series prior to the cross-correlation (Pedersen et al., 2007) or the application of the one-bit normalisation after the spectral whitening of the seismic noise time series (Brennguier et al., 2008a) as discussed in section 6.2.1.

The waveform preserving time domain normalisation of the CCFs (wpcf) is as capable as the non-linear running absolute mean normalisation to suppress transient signals if a suitable time window length (shorter than 6 hr in this case) is chosen. This is shown by the decreasing waveform differences in comparison to the reference CCF with decreasing time window length (Figure 6.11). This decrease can be explained by the better correlation of the time window length with the length of the typically occurring transient signals such as teleseismic earthquake waves. Loosely speaking, the aiming accuracy of the wpcf-normalisation increases and the typically occurring transient signals are more effectively suppressed.

The spectral whitening is capable and important to improve the CCFs in terms of being broad-band Green's function estimates notwithstanding the decreased SNR. Spectral whitening is in principal also capable to efficiently suppress the influence of transient signals like earthquakes if a suitable time window length is chosen. This is shown by the small differences between the ram-SW- and the SW-CCFs for time window lengths shorter than 6 hours (Figure 6.11). Unfortunately, the application of spectral whitening with such a short time window leads to the problematic amplification of persistent monochromatic periodic signals like the observed 26 s signal (see section 6.2.2). Therefore, a combination of spectral whitening with a time domain normalisation like the running absolute mean normalisation is necessary, if comparably short transient signals should be suppressed. Another possibility would be the combination of the wpcf-

normalisation with spectral whitening in a stacking process with two stages, as discussed above.

6.6 Evaluation of the data selection approaches

The evaluation of the two data selection approaches is realised like the evaluation without data selection in the previous section. The CCFs obtained with data selection are also compared with the identical reference CCFs used above. The evaluation of the broadband CCFs obtained with one-bit-normalisation is excluded from the following evaluation to improve the clarity of the figures. First, the amount of data selected by the two approaches is discussed in dependency of the time window length. Afterwards, the data selection with the time series approach (TSA) is evaluated in section 6.6.1 and with the waveform symmetry approach (WSA) in section 6.6.2.

The total amount of used data in percent is shown over time window length for the different data selection approaches in Figure 6.13. With the TSA data selection (time windows without contemporaneous transients at both stations) the amount of used data increases with decreasing time window length (red line in Figure 6.13). This is caused by the better concurrence of a short time window length (< 4 hr) with the length of the occurring transient signals (teleseismic surface waves).

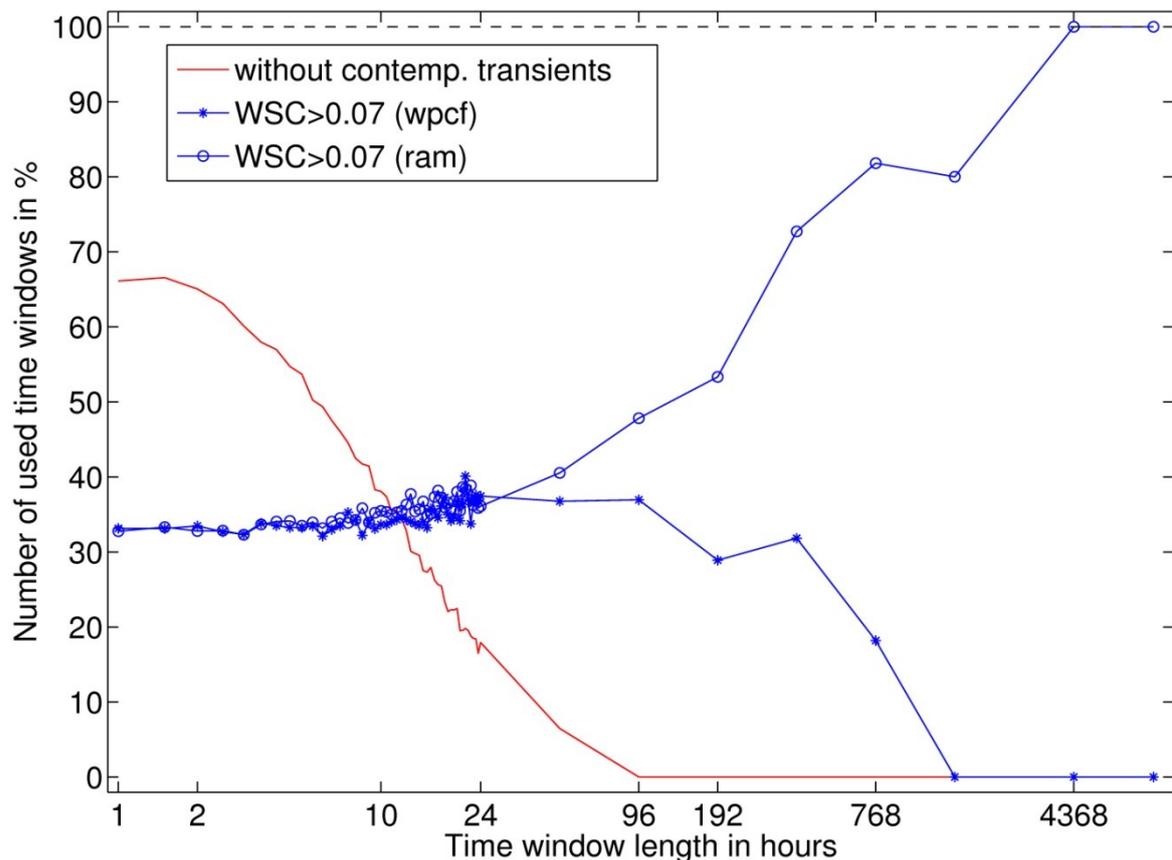


Figure 6.13: Number of time windows in percent selected by the WSA and TSA data selection approaches. The number of time windows in percent of the year 2004 (ANMO-CCM, 7-150 s) is shown for the different data selection approaches over time window length. Time windows with contemporaneous transients at both stations are excluded (red line). The waveform symmetry approach (WSA) excludes CCFs with a WSC value smaller than 0.07. The WSA data selection is influenced by the normalisation scheme. Therefore, the number of time windows selected by the WSA data selection is shown for the wpcf-normalisation (blue line with asterisks) as well as the ram-normalisation (blue line with circles).

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 exclusion of all time

windows longer than 48 hr. Due to the continuous global seismicity such long time windows are correctly identified to contain strong transient signals. The total amount of used data increases to approx. 68 % with a time window length of 1.5 hr.

The amount of used data with the WSA data selection based on the waveform symmetry of the CCFs depends on the applied normalisation. With the running absolute mean normalisation (blue line with circles in Figure 6.13) the amount of used data decreases with decreasing time window length to a level of ~35%. This illustrates, that the waveform symmetry of the single time window CCFs decreases with decreasing time window length in general. With the wpcf-normalisation (blue line with asterisks in Figure 6.13) the amount of used data increases towards the same level of ~35 % for a time window length shorter than 48 hr. With longer time windows (> 48 hr) the wpcf-normalisation fails because it is not capable to suppress the negative influence of problematic signals significantly shorter than 48 hr on the obtained CCFs.

6.6.1 Time series approach (TSA)

The comparison of the CCFs (7-150 s) with the reference CCFs is shown for the station pairs ANMO-CCM and CCM-DWPF in Figure 6.14.

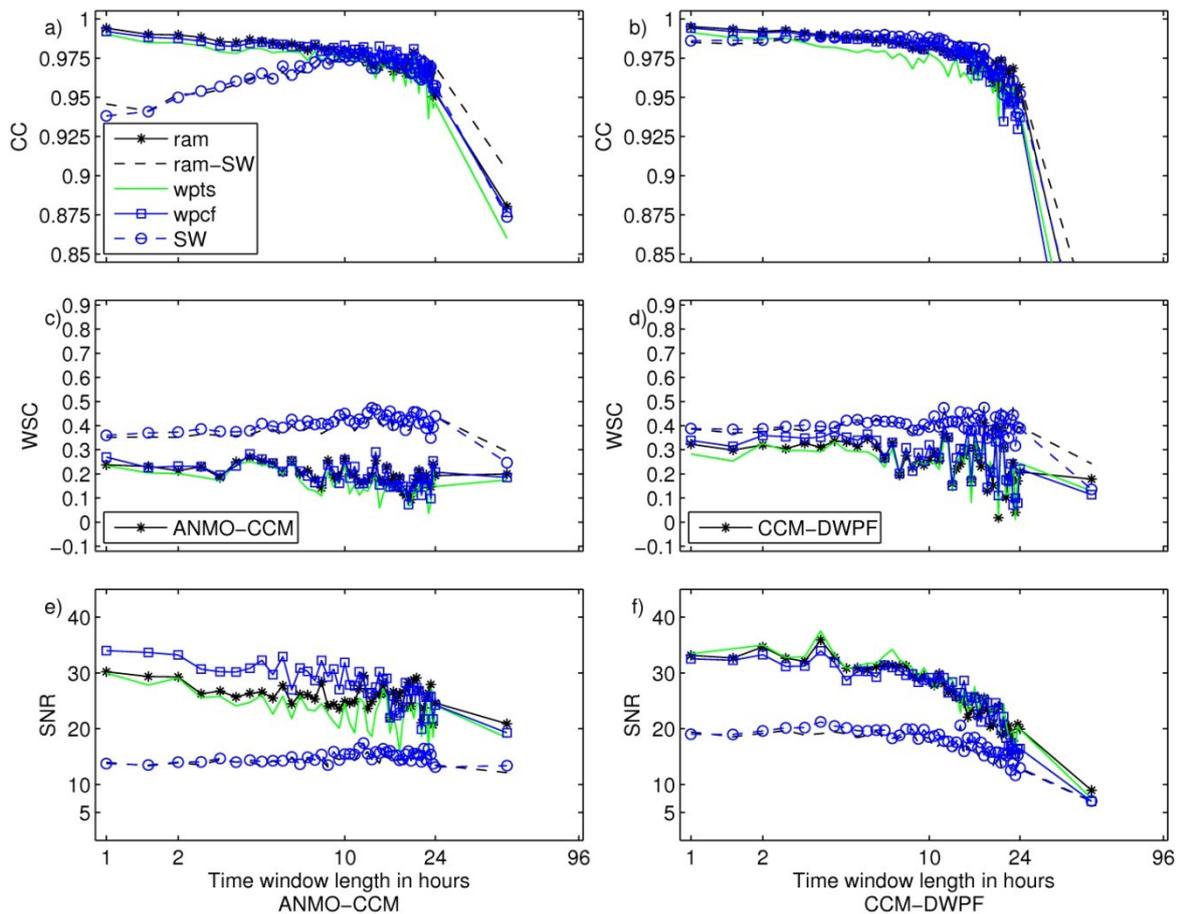


Figure 6.14: Influence of the TSA data selection approach on the obtained CCFs.

A detailed comparison of broad-band CCFs (7-150 s) obtained with different normalisation schemes and time window lengths is shown for the station pairs ANMO-CCM (left side) and CCM-DWPF (right side) with the parameters CC (top; a+b), WSC (middle; c+d) and SNR (bottom; e+f). Time windows containing no data or with transient signals observed contemporaneous at both sites (TSA data selection) are excluded from the stacking. All time windows longer than 48 hours contain dominant transient signals due to the continuous global seismicity. Therefore, no results are obtained with such long time windows in combination with the TSA data selection. The axis of the time window length is reduced in this figure to the range 1-100 hr to improve readability.

With the exclusion of time windows containing contemporaneous transients at both stations the simple waveform preserving normalisation of the time series (wpts, green lines in Figure 6.14) produce CCFs comparable with the reference CCF. In general, the differences to the reference CCF are of the same size as the time domain normalisations ram, wpcf and wpts. A direct comparison of the ram- and the wpcf-CCFs shows correlation coefficients larger than 0.99 for time window lengths shorter than 20 hours (not shown here). Furthermore, no significant differences between the ram-SW- and the SW-CCFs can be observed. The waveform symmetry and SNR are comparable between the different normalisation schemes and mainly influenced by the time window length and therefore the total amount of used data (Figure 6.14b and c).

The differences between the WSC values of the CCFs obtained with data selection and the reference CCFs (ram/ram-SW, 24 hr, only excluding data gaps) are shown as function of time window length in Figure 6.15. In general slightly lower WSC values are observed for the CCFs obtained with data selection. The WSC differences are the same for the ram- and wpts-normalisation as well as the ram-SW- and SW-normalisations. The WSC values of the wpts-CCFs are in general slightly lower, but comparable for time window lengths shorter than 6 hours.

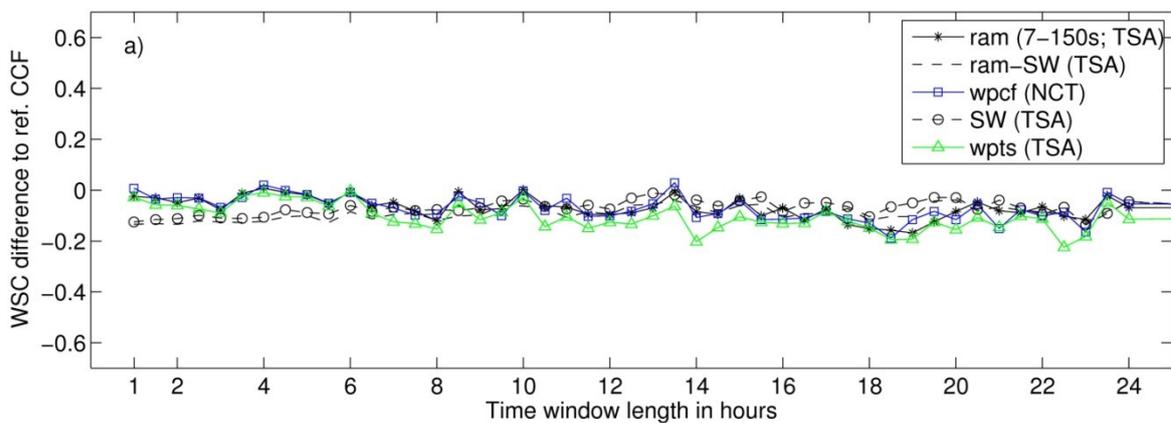


Figure 6.15: Influence of the TSA data selection approach on the CCF waveform symmetry. The differences between the WSC values of the CCFs (7-150 s) obtained by the TSA data selection and the WSC value of the reference (ram/ram-SW, 24 hr, no data selection) CCFs are shown for different normalisation schemes.

The comparable results for the ram-, wpcf- and wpts-normalisation show the effective exclusion of transient signals from the data by the selection approach based on the time series classification. With this data selection, the non-linear running absolute mean normalisation yields no better CCFs in terms of waveform symmetry or SNR. The fully automated exclusion of transient signals is operational and effective by replacing a non-linear time domain normalisation. Nevertheless, the CCFs obtained with this data selection approach are not improved in comparison to CCFs obtained without data selection but an effective normalisation of transient signals. The author concludes that an appropriate normalisation of contemporaneous transient signals should be preferred instead of excluding them from the cross-correlation processing.

6.6.2 Waveform symmetry approach (WSA)

In this subsection the performance of the WSA data selection is discussed in combination with the different normalisation schemes and time window lengths. The results of the data selection based on the waveform symmetry of the single CCFs are shown in Figure 6.16. The obtained CCFs are comparable with the reference CCFs in terms of the symmetric-component signal time windows and the SNR, although only ~35 % of the data are used (see also Figure 6.18 discussed below). Only the wpts-normalisation fails with this data selection approach. The behaviour of the different normalisation schemes regarding the

time window length is also the same as without data selection (compare Figure 6.12 a+b with Figure 6.16a+b).

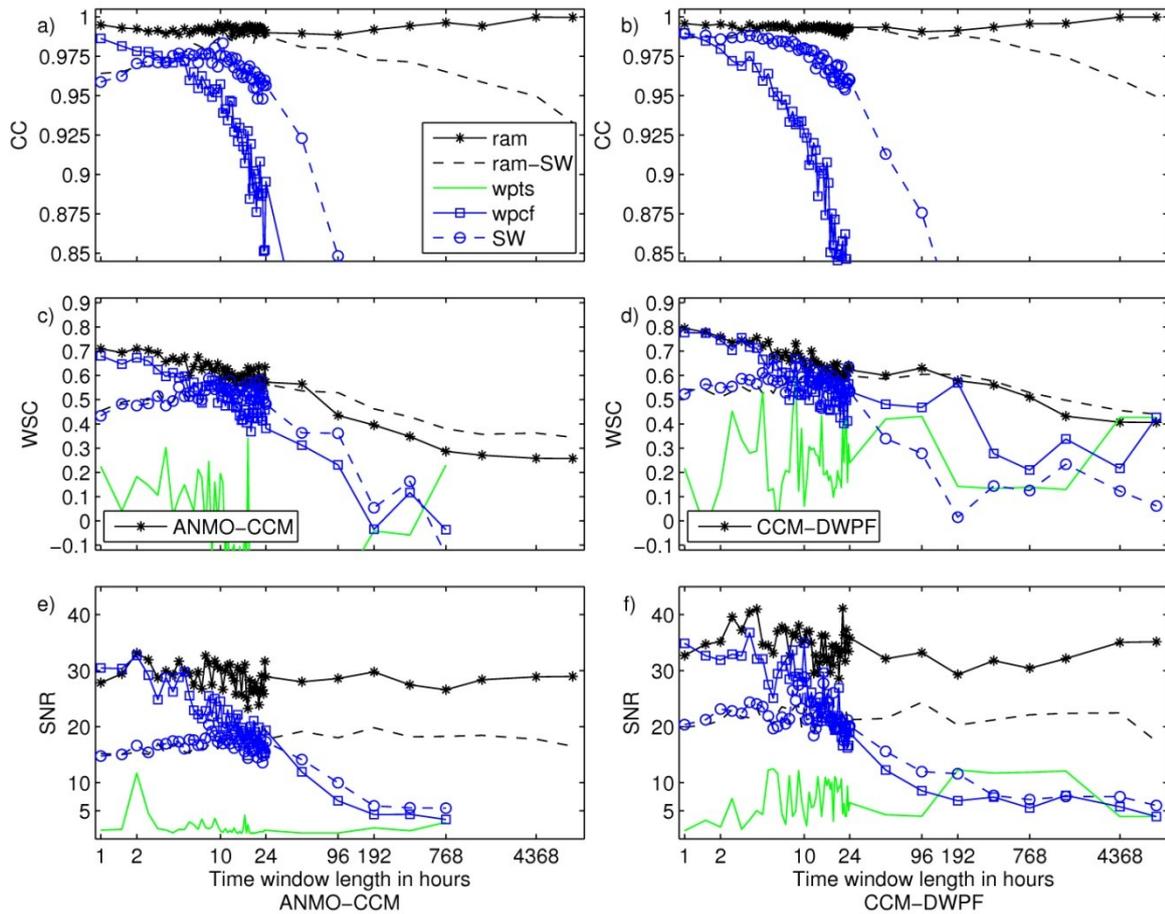


Figure 6.16: Influence of the WSA data selection approach on the obtained CCFs. A detailed comparison of broad-band CCFs (7-150 s) obtained with different normalisation schemes and time window lengths is shown for the station pairs ANMO-CCM (left side) and CCM-DWPF (right side) with the parameters CC (top; a+b), WSC (middle; c+d) and SNR (bottom; e+f). Time windows with a WSC value smaller than 0.07 are excluded from the stacking (WSA data selection).

Remarkable improvements of the waveform symmetry of CCFs obtained with the ram- and wpcf-normalisation but without spectral whitening are observed in several frequency bands. The differences of the WSC values in comparison to the reference CCFs (ram/ram-SW, no data selection) are shown in Figure 6.17a. The increase of the WSC value in comparison to the reference CCF is larger with shorter time windows and larger with the ram-normalisation. With spectral whitening no significant improvement of the waveform symmetry is observed (Figure 6.17a). A comparison of the WSC values of the ram-CCFs is given for different period bands in Figure 6.17b. An improvement of waveform symmetry is observed especially for the intermediate period band 20-50 s and the short period band 7-14 s. The very high waveform symmetry in the period band 70-150 s (~0.9) is not improved. A more detailed waveform analysis (not shown here) reveals that in general that part (causal or acausal) of the two-sided CCFs with the weaker signal is changing in waveform by getting more similar to the stronger signal part.

This is also illustrated by the waveform sections with the CCFs of all 10 station pairs shown in Figure 6.18. The CCFs obtained with the running absolute mean normalisation, a 24 hr time window and all data (no data selection) are shown in Figure 6.18a. In Figure 6.18b the CCFs obtained with the waveform preserving wpcf-normalisation, a 2 hr time window and the WSA data selection (~35% data usage) are shown. No significant differences in the stronger acausal parts between the CCFs obtained with the different

normalisation and data selection approaches are observed. The acausal parts of the CCFs correspond to a propagation of the seismic waves from the western station towards the eastern station. The variations of the causal parts can actually be seen in the coarsely resolved waveform sections (Figure 6.18a and b). As the symmetric-component CCFs are dominated by that part of the CCFs with the larger amplitudes the changes to the symmetric-component are rather small, although the waveform symmetry of the two-sided CCF is improved.

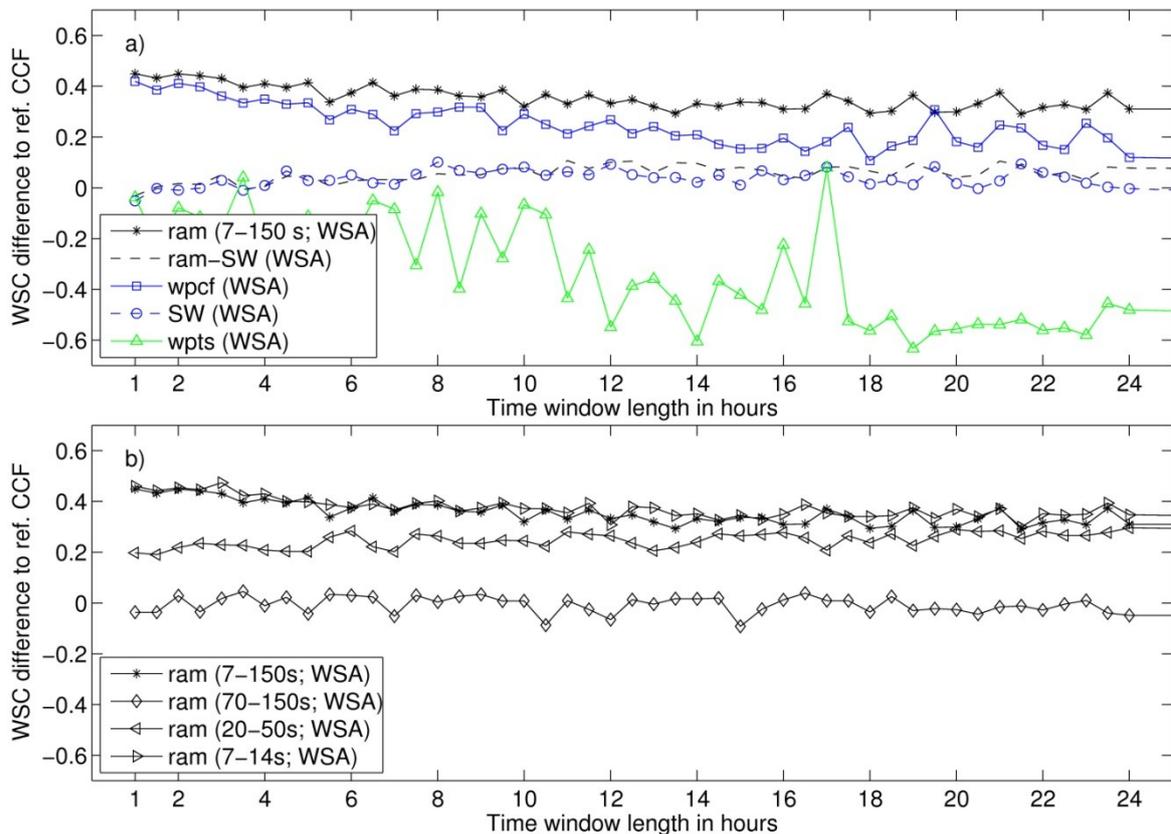


Figure 6.17: Influence of the WSA data selection approach on the CCF waveform symmetry. The differences between the WSC values of the CCFs (7-150 s) obtained by the exclusion of CCFs with a WSC value smaller than 0.07 (WSA data selection) and the WSC value of the reference (ram/ram-SW, 24 hr, no data selection) CCFs are shown for different normalisation schemes in (a). These differences are shown in (b) for the ram-CCFs in different period bands.

The failure of the wpts-normalisation indicates that time windows with large transient signals remain in the selected data as the symmetry of the signal time windows is the controlling parameter. This parameter is not mandatorily influenced by strong transients in the CCFs near lag time zero. Therefore, an effective normalisation of the transient signals is important in combination with this symmetry based selection approach. The author regards this selection approach as successful because CCFs with an improved waveform symmetry in comparison to the reference CCF are obtained from only ~35% of the data. The improvement of the waveform symmetry is nevertheless based on the improvement in some frequency ranges with high relative amplitudes, especially 7-14 s in this case. The CCFs obtained with spectral whitening benefit not significantly from the selection approach, as the relative amplitudes are balanced. The symmetry of the CCFs obtained with spectral whitening is significantly supported by the low frequency signals in the period band 70-150 s which play a secondary role in the CCFs obtained without spectral whitening due to their small relative amplitudes. The low relative amplitudes and the already high waveform symmetry of the low frequency signals explain also that they do not benefit from this data selection approach.

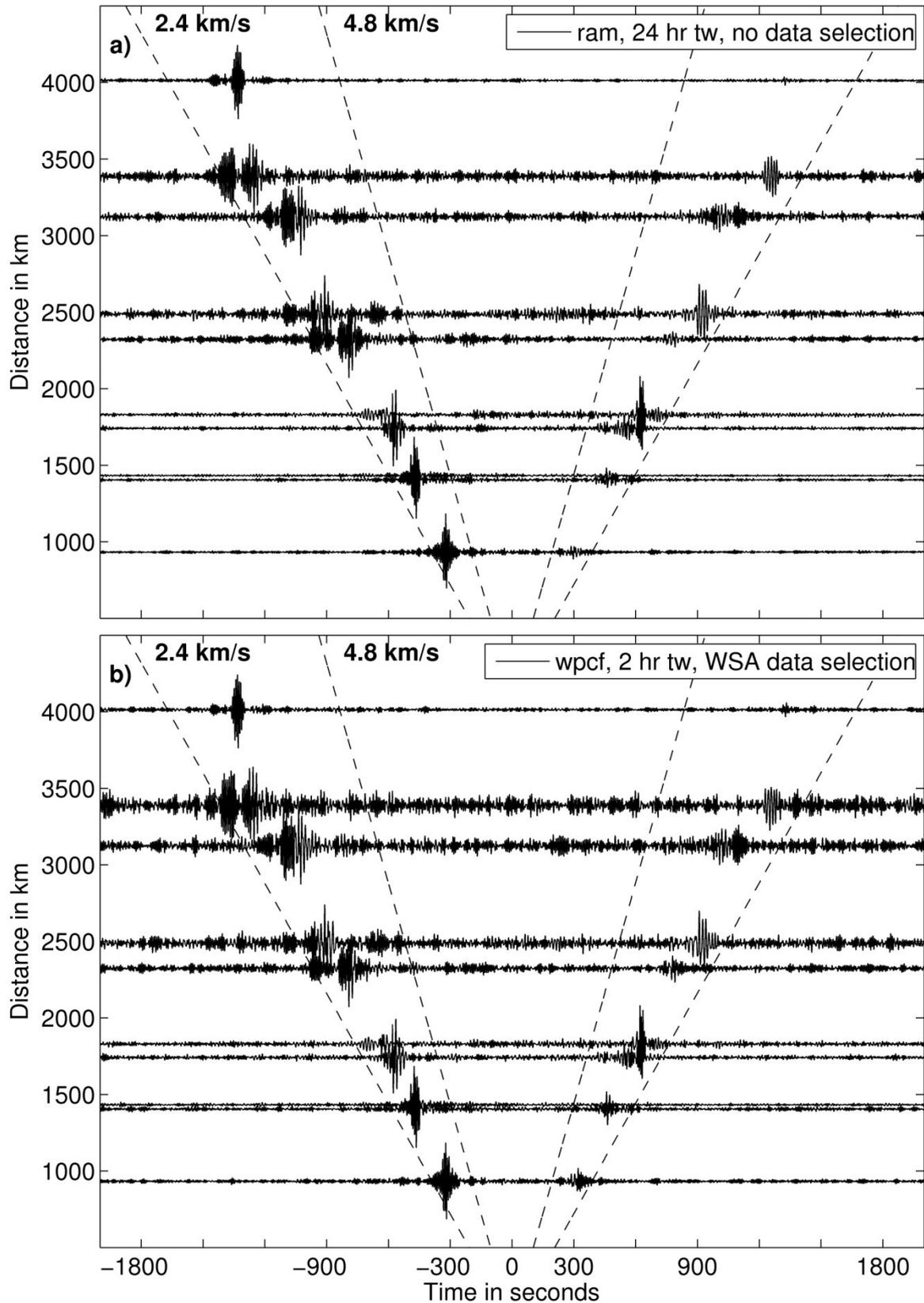


Figure 6.18: Distance-dependent illustration of the CCFs (7-150 s) of all ten station pairs. The signal time windows are indicated by the dashed lines. In (a) the CCFs obtained with running absolute mean normalisation, 24 hr time window and one year of data (without data selection) are shown. In (b) the CCFs obtained with the waveform preserving wpcf-normalisation, 2 hr time window and with the WSA data selection are shown. Only 35% of the data (1 year) are used to obtain the CCFs shown in (b).

The observations that the stronger part of the CCFs changes less than the weaker part indicates that the stronger parts of asymmetric CCFs are useful to estimate the Green's function. It furthermore seems questionable if the stacking of the causal and acausal part to a common symmetric-component is recommendable in the case of strongly asymmetric CCFs to improve the Green's function estimation. This topic needs further research.

6.7 Summary of chapter 6

The author recommends the application of the unbiased linear digital cross-correlation to avoid the unwanted influence of the circular effect. This is especially true if a time window length shorter than ten times the maximum analysed lag time is used. Furthermore, it is recommended to apply overlapping time windows with a minimum time window length in the range of four to six times the maximum analysed lag time to improve the estimation of the complete CCF by the stacking of shorter CCFs.

The proposed waveform-preserving normalisation of the CCFs (wpcf normalisation) prior to the stacking is capable to replace a non-linear time domain normalisation. It is necessary to select a suitable time window length which is similar to the length of typically occurring coherent transient signals such as seismic waves excited by earthquakes.

The spectral whitening normalisation is observed to be significantly influenced by the time window length. The application of spectral whitening on very short time windows of a few hours undesirably amplifies temporally persistent coherent signals like the 26 s microseism excited in the Gulf of Guinea in continental scale seismic noise CCFs. The time window length used with spectral whitening should be as long as possible to assure the emergence of signals of the Green's function in the CCFs obtained from the single time windows. A possibility to apply a time window dependent time domain normalisation like the wpcf normalisation in combination with spectral whitening is a stacking process with two stages.

The TSA data selection, which removes time windows with contemporaneous transient signals from the data, is capable to replace a non-linear time domain normalisation. But it is necessary to select a suitable time window length comparable to the length of the transient signals. Nevertheless, no improvement of the obtained CCFs in terms of waveform symmetry is observed in comparison to the CCFs obtained without data selection. It is the opinion of the author that a suitable normalisation of transient signals is preferable in contrast to a strict exclusion.

The WSA data selection based on the waveform symmetry of the single time window CCFs proved to be very effective by producing high-quality CCFs in comparison to the full data CCFs using only ~35% of the data. Furthermore, an improvement of the waveform symmetry of the stacked CCFs is observed without spectral whitening. The author considers data selection approaches based on the characteristics of the single time window CCFs instead of the time series as most promising to improve stacked seismic noise CCFs.

7 Summary and Conclusions

In this thesis the temporal and spatial variations of the statistical properties of the seismic noise have been analysed and classified. A new statistical time series classification was developed which is capable to distinguish between corrupt and non-corrupt time series as well as to classify non-corrupt time series in six meaningful noise classes (chapter 4). The time series classification was used successfully to provide a comprehensive analysis of the spatial and temporal variations of the seismic noise in the metropolitan area of Bucharest (chapter 5). This analysis improves the understanding of the statistical properties of the urban seismic noise due to temporally and spatially varying noise sources in general. The combination of the time series classification with an unsupervised neural network technique, the Self-Organizing Map method, proved as a promising approach to enhance the analysis of complex urban seismic noise data sets. The time series classification was furthermore used to realise a data selection approach for the calculation of seismic noise cross-correlation functions for seismic interferometry. The implementation of this data selection approach involved a comprehensive evaluation of the common cross-correlation processing (chapter 6). Based on this evaluation a more flexible processing scheme was realised and critical parameters of the processing such as the time window length were identified. Furthermore, a wave form preserving time domain normalisation and a second data selection approach were developed in this context to improve the calculation of seismic noise cross-correlation functions. The main conclusions and results of this thesis are summarised in the following with reference to the corresponding chapters.

Chapter 2 provides a brief review of the state of knowledge about seismic noise and summarises the most important applications of seismic noise. A focus is laid on seismic interferometry as the corresponding data processing is addressed in chapter 6.

In chapter 3 the data sets used for this thesis are introduced. Furthermore several data format inconsistencies which became apparent to the author while working with the SEED data from international data centres are discussed together with their practical relevance. The SEED format is the common standard for the distribution of digital seismological time series. It provides the time series together with the complete metadata of the seismic measuring system. An inconsistency found in the metadata of GSN station HRV affects the description of FIR filters which are commonly used for the decimation stages of digitisers. It turned out, that the metadata of dozens of seismic stations world-wide were (and partly still are) affected by one or several inconsistencies in the description of their FIR filters. The filter coefficients of asymmetric FIR filters are given sometimes in reverse order but should be given in forward order following the SEED definition. Another common inconsistency affects the value for the corrected filter delay time of a FIR filter stage. The corrected filter delay time has to be positive in every case following the SEED definition but is often found as a negative value or to be zero. The amount of the time shift introduced by the removal of an incorrect frequency response function depends on the FIR configuration of the seismic channel and can be up to 2 s. This time shift is not acceptable for many seismological applications such as seismic tomography.

The time series classification introduced in chapter 4 uses ratios between time series percentiles to identify deviations of a time series distribution from the Gaussian distribution with high sensitivity on statistical data properties. The ratio between the 99.73% amplitude interval and the 95.45% amplitude interval is introduced as the quantity *peakfactor*. The peakfactor (pf) equals 1.5 in the case of a Gaussian distributed time series and increases/decreases with increasing/decreasing kurtosis of the time series.

The time series classification is capable to identify several types of corrupt time series (e.g. technical problems with the sensor). Regarding the non-corrupt time series six noise classes are introduced to classify the typically observed deviations of seismic noise time

series from the Gaussian distribution. Gaussian distributed time series are classified as noise class 1 (NC1). Non-Gaussian but symmetric time series are classified as NC2-NC5. Time series which exhibit determinable but rather small and unspecific deviations from the Gaussian distribution ($pf \ 1.5 \pm 0.1$) are classified as noise class 2 (NC2). Time series with a gentle peaked histogram in comparison to the Gaussian distribution ($1.6 < pf \leq 2$) 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 of the time series as noise class 4 (NC4). Symmetric time series with a flattened histogram in comparison to the Gaussian distribution ($pf < 1.4$) are classified as noise class 5 (NC5). All time series which are not identified as symmetric time series are classified as noise class 6 (NC6). The influence of the applied band pass filters and the time window length on the time series classification was evaluated with a data set of synthetic time series in section 4.5. A time window length of at least 200 times the longest contained period is recommended by the author to ensure an unbiased statistical time domain analysis.

In chapter 5 the urban seismic noise in the metropolitan area of Bucharest is analysed with a time-frequency analysis and the time series classification introduced in chapter 4. The time-frequency analysis discussed in section 5.1 is used to determine 8 frequency ranges which represent the frequency-dependent influence of dominant natural and man-made sources on the urban seismic noise (USN).

The time series classification provides useful information in addition to the noise amplitudes to analyse spatial and temporal variations of the USN conditions. Time series of USN with a duration of 4 hours exhibit predominantly (~90%) bell-shaped distributions, but are not predominantly Gaussian distributed. Deviations from the Gaussian distribution occur due to strong transient or periodic signals. In fact, the influence of human activity can be observed across the whole analysed frequency range of 0.008-45 Hz by changing statistical properties from day- to nighttime. Especially in the frequency range 1-25 Hz transient signals characterise the USN in Bucharest. At higher frequencies between 25 Hz and 45 Hz periodic signals contribute to the USN, which are most probably excited by rotating machinery. In the frequency range 0.6-1 Hz urban seismic noise in Bucharest is dominated by man-made ground motion as long as the wind velocity is below 3 m/s. With wind velocities exceeding 3 m/s the USN is increasingly dominated by wind-induced ground motion. Both, the influence of human activity and wind, can be observed by temporal and spatial changes of noise amplitudes and statistical properties.

Towards lower frequencies between 8 mHz and 0.6 Hz natural sources of seismic energy, like ocean-generated microseism and earthquakes, dominate the USN in Bucharest. Due to the larger amplitudes of the naturally induced seismic waves (e.g. ocean-generated microseism), the influence of human activity is not as obvious as at higher frequencies from the variation of noise amplitudes between day- and nighttime. Nevertheless, the influence of human activity can be observed by changes of the statistical properties with time. Human influence can cause significant deviations from the Gaussian distribution also at frequencies below 0.18 Hz. Concerning the USN amplitudes in the frequency range 0.09-0.25 Hz a spatial dependency is observed. The noise amplitudes increase from the southern part of the metropolitan area towards the North. The same effect of ground motion amplification was observed by an amplitude and site effect study utilising earthquake signals and is related to resonance effects in the unconsolidated sediments above the dipping Neogene-Cretaceous boundary (Mandrescu et al., 2004; Sudhaus & Ritter, 2009). The similar observation for the noise amplitudes demonstrates the potential of noise amplitude mapping to complement information for site effect studies.

The statistical properties of the USN on the horizontal North-South component are in general identical to these of the USN on the East-West component. Minor differences are observed at two station sites (URS19 and URS20) in the frequency range above 25 Hz due to periodic signals with a dominant direction of vibration. Most probable sources of

these signals are electrical engines in direct vicinity of the sensors. The statistical properties of the vertical and the horizontal components of the USN show minor differences in the frequency range below 0.18 Hz and in the frequency range 1-25 Hz. In general a larger amount of transient signals is observed on the horizontal components than on the vertical component below 0.18 Hz. In the frequency band 1-25 Hz the amount of time windows classified as NC1 and NC2 is slightly increased on the horizontal components. The observed differences have to be assumed to be related to the sources of the seismic noise such as surface waves excited by earthquakes (below 0.18 Hz) or rotating machinery (especially above 25 Hz). This topic needs further research including a careful analysis of individual signals contributing to the USN. The average amplitude ratios between the vertical and the horizontal components (H/V ratio), obtained from the noise classification results, correspond well with the results of the spectral H/V ratio analysis of Ziehm (2006). The spectral H/V ratio is around or slightly below 1 for frequencies larger than 2 Hz and exhibit two peaks with average H/V ratios between 2 and 4 in the frequency ranges 0.18-0.25 Hz and 0.6-0.9 Hz. The frequency resolution of the presented time-domain H/V ratio is worse than the resolution of the spectral H/V ratio but can be improved by the selection of more suitable frequency bands and the implementation of a data selection procedure similar to the data selection used for the calculation of the spectral H/V ratio.

Concluding, the changes of the statistical properties and noise amplitudes in USN, like the ones presented in chapter 5, are linked to the variability of their generating processes and (changing) underground conditions. The time series classification proved to be capable to derive information about noise amplitudes and statistical properties automatically from a large broadband seismological data set. At present this information can be used to select time windows of data not influenced by deterministic transient or periodic signals for sensitive methods utilising seismic noise like H/V or seismic interferometry. As a future step, this information can be used as input for well established techniques of pattern recognition and knowledge discovery to resolve up to now unrecognized temporal or spatial interrelations. The discovery of unrecognized interrelations is furthermore a starting point to identify and finally better understand the different physical processes contributing to USN.

A feasibility study to demonstrate the potential of this approach was the analysis of the vertical-component USN in the frequency band 0.6-1 Hz with the Self-Organizing Map (SOM) method (section 5.5). The analysis provided five meaningful classes of time windows. The SOM classification discriminates reliably the working hours 8-20 EET on working days (SOM class 2, 38.1% of 1709 analysed time windows) from the time windows at night, morning and evening as well as on Sundays (SOM class 4, 55.5%). The classification of a time window as SOM class 1 (2.6%) indicates reliably higher-than-average wind velocities. The time windows classified as SOM class 3 (1.4%) or 5 (2.4%) are found to be affected by seismic waves excited by local, regional and teleseismic earthquakes. The SOM classes 3 and 5 discriminate time windows affected by the global seismicity with high (SOM class 3) and low (SOM class 5) human activity identical to the SOM classes 2 and 4. In general, the feasibility study to analyse a seismic noise classification data set with the SOM method was very successful. Nevertheless, the clustering of the SOM needs further evaluation as it is assumed to find more meaningful 'subclasses' of the SOM classes 2 and 4 which summarise more than 93% of the 1709 analysed time windows.

Chapter 6 provides a discussion of all important aspects of the seismic noise cross-correlation processing for the estimation of Green's functions from seismic noise including new data normalisation and selection methods. Several established normalisation methods were analysed such as the running absolute mean (Bensen et al., 2007) and the one bit normalisation (Shapiro et al., 2005) in the time domain as well as the spectral whitening normalisation (Bensen et al., 2007) in the frequency domain. Furthermore two

new approaches of waveform preserving time domain normalisation before and after the cross-correlation were developed. All normalisation methods were evaluated with one year of data (GSN data set, section 3.2) which was fragmented with different time window lengths between one hour and one year. As addition to the normalisation of the data two fully automated data selection approaches were realised and evaluated. The Time Series Approach (TSA) is based on the time series classification and excludes time windows containing contemporaneous transient signals at both sites (e.g. teleseismic earthquakes). The second data selection approach, the Waveform Symmetry Approach (WSA), is based on the waveform symmetry of the single time window Cross-Correlation Functions (CCFs). The main conclusions and recommendations of chapter 6 are summarised in the following.

The author recommends the application of the unbiased linear digital cross-correlation instead of the circular digital cross-correlation to avoid the unwanted influence of the circular effect. This is especially true if a time window length shorter than ten times the maximum analysed lag time is used. Furthermore, it is recommended to apply overlapping time windows with a minimum time window length in the range of four to six times the maximum analysed lag time to improve the estimation of the complete CCF by the stacking of shorter CCFs.

The proposed waveform-preserving normalisation of the CCFs (wpcf normalisation, section 6.4.2) prior to the stacking is capable to replace a non-linear time domain normalisation such as the running absolute mean or the one-bit normalisation. It is necessary to select a suitable time window length which is similar to the length of typically occurring coherent transient signals such as seismic waves excited by earthquakes.

The spectral whitening normalisation is observed to be significantly influenced by the time window length (section 6.2). The application of spectral whitening on very short time windows of a few hours undesirably amplifies temporally persistent coherent signals such as the 26 s microseism excited in the Gulf of Guinea in continental scale seismic noise CCFs. The time window length used with spectral whitening should be as long as possible to assure the emergence of signals of the Green's function in the CCFs obtained from the single time windows. A possibility to apply a time window dependent time domain normalisation like the wpcf normalisation in combination with spectral whitening is a stacking process with two stages.

The TSA data selection, which removes effectively time windows with contemporaneous transient signals at both sites from the data, is capable to replace a non-linear time domain normalisation. It is necessary to select a suitable time window length comparable to the length of the transient signals. Nevertheless, no improvement of the obtained CCFs in terms of waveform symmetry is observed in comparison to the CCFs obtained without data selection. It is the opinion of the author that a suitable normalisation of transient signals is preferable in contrast to a strict exclusion.

The WSA data selection based on the waveform symmetry of the single time window CCFs proved to be very effective by producing high-quality CCFs in comparison to the full data CCFs using only ~35% of the data. Furthermore, an improvement of the waveform symmetry of the stacked CCFs is observed without spectral whitening in several frequency bands. The author considers data selection approaches based on the characteristics of the single time window CCFs instead of the time series as most promising to further improve stacked seismic noise CCFs.

Concluding, an effective time series classification for seismic noise time series is proposed in this thesis. It is demonstrated that the time series classification can be used to obtain new insights into the temporal and spatial variations of (urban) seismic noise. The time series classification provides furthermore valuable data selection capabilities for all methods utilising seismic noise.

Appendix A: Karlsruhe Seismology Processing Toolbox

The Karlsruhe Seismology Processing (KaSP) toolbox for MATLAB provides basic as well as specialised functionalities for seismological data processing. The KaSP toolbox is an ongoing development in the seismology working group at the Geophysical Institute (GPI) of the Karlsruhe Institute of Technology (KIT). The development of the KaSP toolbox was initiated by the author in 2009 to improve the collaborative development of seismological MATLAB code at the GPI. The KaSP toolbox is a further development of the “SeisNoise toolbox” created by the author in 2007. The toolbox is now used and further extended by several graduate and undergraduate students for the development and implementation of seismological processing techniques at the GPI. Main contributing authors up to now are Britta Wawerzinek, Tobias Horstmann, Daniel Armbruster and Tobias Baumann. Detailed references to the authors are given in the source code.

The KaSP toolbox is free software under the GNU license and can be obtained from the staff of the seismological working group at the GPI (www.gpi.kit.edu) or directly from the author.

The most important functionalities of the KaSP toolbox are summarised in the following. A detailed technical documentation of the toolbox is not reasonable here due to the large extent and the ongoing development of the toolbox.

Data handling functionalities

The toolbox provides own functions to read seismic data from established data formats such as Q files (*SeismicHandler* data format) and binary SAC files (*Seismic Analysis Code* data format). The reading of seismic data from fullSEED volumes into MATLAB can be done automatically if the software *rdseed* from IRIS is installed. An interface function to the C++ library *DATREAD++* from Thomas Forbriger (<http://www.rz.uni-karlsruhe.de/~bi77/txt/cxx/libdatreadxx/html/index.html>) was written to be able to read data from files with the formats miniSEED, Stuttgarter File Format (SFF), GSE and others to MATLAB. The KaSP toolbox is based on a MATLAB struct with specified fields (the ‘time series struct’) which is used by all KaSP processing functions. The toolbox provides the possibility to write seismic data in a KaSP time series struct to a *SeismicHandler* Q file.

The time series struct provides furthermore a history of all processing steps applied to the data. The transfer function of the corresponding instrument and further information on the seismic station can be stored with the time series. The toolbox contains furthermore a station information file and the possibility to store additional instrument response files. The preprocessing of seismic time series including the removal of the instrument response can be fully automated with this station metadata inventory.

The toolbox provides furthermore functions for the import of transfer functions (seismometer and FIR filter stages) from SEED response files generated with *rdseed* from fullSEED volumes. The imported transfer functions can be automatically checked for common inconsistencies in the SEED metadata (see chapter 3) and used directly for the removal of the frequency response function. Several functions for plotting and analysis of frequency response functions are implemented in the toolbox.

Data requests via the e-mail based request tool *breq_fast* (http://www.iris.washington.edu/manuals/breq_fast.htm) can be generated interactively and sent directly from MATLAB to a data centre.

Basic processing functionalities

The toolbox provides all important functionalities of basic seismological (pre-)processing such as:

- Removal of means and linear trends
- Tapering
- Filtering with Butterworth filter in the time domain (TD) and frequency domain (FD)
- Filtering with Gaussian low-pass filter (FD) [Britta Wawerzinek]
- Removal of the instrument response (TD/FD)
- Integration of time series (TD/FD) [Britta Wawerzinek]
- Resampling of time series [Britta Wawerzinek]
- Rotation of components Z, N, E to L, Q, T as well as Z, R, T [Britta Wawerzinek]
- Cutting and merging of time series
- Picking algorithms [Tobias Horstmann, Tobias Baumann]

Seismic noise classification

The seismic noise classification (see chapter 4) is fully implemented in the KaSP toolbox. Time series can be classified at once or analysed with a sliding time window classification. The classification of entire data sets (see chapter 5) can be automated and several tools for the further analysis of the classification results exist.

Spectral analysis

The spectral time-frequency analysis of time series (see chapter 5) is implemented in two ways. On the one hand short time series (less than one week of data, data in the working memory) can be analysed with a very high temporal resolution (seconds to minutes). On the other hand very long time series (several weeks to months, data saved as a fragmented time series in several files) can be analysed with a more coarse temporal resolution (minutes to hours).

Array processing

A realisation of the FK technique (see chapter 2) is implemented which can be applied also with a sliding time window (written by Tobias Baumann).

Seismic noise cross-correlation processing

The seismic noise cross-correlation processing scheme as described in chapter 6 is fully implemented in the toolbox. Several tools (e.g. distance plots, CCF comparison, dispersion curve estimation, ...) for the analysis of seismic noise cross-correlation functions exist. Tobias Horstmann implemented a migration analysis for the localisation of point sources within a seismic network such as the URS network (see chapter 3) with seismic noise cross-correlation functions.

S wave receiver functions

The processing scheme to obtain and analyse S wave receiver functions is implemented by Britta Wawerzinek.

KABBA quality control

The KaSP toolbox provides tools for the evaluation of the KARlsruhe BroadBand Array with routine huddle tests between seismological experiments. Next to the spectral time-frequency analysis and the noise classification also cross-correlation techniques are applied to identify technical problems of sensors as well as data loggers. The three-channel correlation analysis (Sleeman et al., 2006) to measure the instrumental noise of data loggers and sensors was implemented by Daniel Armbruster.

Appendix B: Metadata of seismic stations

This appendix provides the relevant station metadata of the URS data set (section 3.1) and the GSN data set (section 3.2) used for this thesis. Next to the station coordinates the transfer functions of the seismometers (ground motion velocity to ground motion velocity) are given by their pole-zero representations in the s-plane (see section 3.3.3).

The station list of the URS data set providing the station name, location, sensor type as well as starting and closing date is given in Table B.1. The transfer functions are given in Table B.2. The amplitude and phase frequency response functions are plotted in Figure B.1.

The station list of the GSN data set providing the station name, location, sensor type and the transfer functions is given in Table B.3. The amplitude and phase frequency response functions are plotted in Figure B.2.

URS data set

Name	Sensor	Location	Latitude degree	Longitude degree	Starting mm.dd.yyyy	Closing mm.dd.yyyy
URS01	STS-2	Zoological garden	44.51853° N	26.10449° E	03.11.2003	27.07.2004
URS02	STS-2	Voluntari	44.48668° N	26.16914° E	27.10.2003	05.08.2004
URS03	STS-2	Acroconstucto	44.44505° N	26.02311° E	29.10.2003	17.06.2004
URS04	STS-2	Sere Horticola	44.42220° N	25.99993° E	28.10.2003	01.04.2004
URS05	STS-2	Jimbolia	44.48975° N	26.03197° E	24.10.2003	09.06.2004
URS06	STS-2	Sulari	44.67768° N	26.25261° E	26.10.2003	02.08.2004
URS07	STS-2	Jilava	44.32940° N	26.08646° E	27.10.2003	18.05.2004
URS08	STS-2	Agropole	44.35540° N	26.20337° E	24.10.2003	17.07.2004
URS09	STS-2	Geotec	44.43712° N	26.10683° E	24.10.2003	27.07.2004
URS10	STS-2	Otopeni	44.55055° N	26.07512° E	25.10.2003	27.07.2004
URS11	STS-2	geol. Museum	44.45526° N	26.08514° E	24.10.2003	27.07.2004
URS12	STS-2	Curtea Veche	44.43004° N	26.10170° E	24.10.2003	27.07.2004
URS13	STS-2	Scoala Steri	44.41503° N	26.17772° E	29.10.2003	27.07.2004
URS14	STS-2	Tei	44.46977° N	26.11258° E	29.10.2003	27.07.2004
URS15	STS-2	Stefanesti	44.53223° N	26.21334° E	25.10.2003	27.07.2004
URS16	STS-2	Tineretul Park	44.40589° N	26.11897° E	25.10.2003	03.08.2004
URS17	STS-2	Ciorogarla	44.44885° N	25.87995° E	25.10.2003	04.08.2004
URS18	STS-2	Herestrau Park	44.47748° N	26.08085° E	24.10.2003	09.08.2004
URS19	STS-2	Buftea	44.56491° N	25.94173° E	26.10.2003	06.08.2004
URS20	STS-2	INCERC	44.44091° N	26.16236° E	28.10.2003	03.08.2004
URS21	STS-2	Magurele	44.34796° N	26.02812° E	22.10.2003	07.08.2004
URS22	STS-2	Geotec Moreni	44.47040° N	26.15510° E	27.10.2003	27.07.2004
URS23	LE-3D/5s	Berser	44.36515° N	26.12239° E	27.10.2003	03.08.2004
URS24	G40T	Imperatui	44.39434° N	26.10826° E	28.10.2003	03.08.2004
URS25	G40T	Hofigal	44.39791° N	26.15939° E	24.10.2003	16.06.2004
URS26	G3ESP	Cutitul Argint	44.41088° N	26.09424° E	25.10.2003	27.07.2004
URS27	KS-2000	Centrul Scolar	44.43577° N	26.12380° E	28.10.2003	03.12.2003
URS28	KS-2000	Scola 128	44.42353° N	26.08211° E	27.10.2003	04.08.2004
URS29	KS-2000	Academy	44.44655° N	26.08997° E	30.10.2003	03.08.2004
URS31	KS-2000	C.S./BST	44.44573° N	26.09839° E	25.10.2003	27.07.2004
URS32	KS-2000	Botanical garden	44.43674° N	26.06561° E	29.10.2003	27.07.2004
URS33	KS-2000	Avram Iancu	44.43737° N	26.11828° E	06.12.2003	27.07.2004
URS34	STS-2	Physics building basement	44.35011° N	26.03121° E	14.05.2004	05.08.2004
URS35	LE-3D/5s	Physics building 11 th story	44.34990° N	26.03118° E	14.05.2004	05.08.2004

Table B.1: Table of all stations of the URS project in the metropolitan area of Bucharest (see Figure 3.1) with locations in decimal degrees, description of the locations as well as the starting and closing dates.

Seismometer type	Gain (A_0) rad/s	Poles rad/s	Zeros rad/s
Streckeisen STS-2	$3.474 \cdot 10^{17}$	$-0.037+i0.037$	0
		$-0.037-i0.037$	0
		$-1.33 \cdot 10^4$	$(-4.631+i4.305) \cdot 10^2$
		$(-1.053+i1.005) \cdot 10^4$	$(-4.631-i4.305) \cdot 10^2$
		$(-1.053-i1.005) \cdot 10^4$	$-1.766 \cdot 10^2$
		$-5.203 \cdot 10^2$	-15.15
		$-3.748 \cdot 10^2$	
		$(-0.973+i4.007) \cdot 10^2$	
		$(-0.973-i4.007) \cdot 10^2$	
		-15.64	
Geotech KS-2000	$9.857 \cdot 10^4$	$-0.044+i0.044$	0
		$-0.044-i0.044$	0
		$(-2.22+i2.22) \cdot 10^2$	
		$(-2.22-i2.22) \cdot 10^2$	
Lennartz LE-3D/5s	1	$-0.888+i0.888$	0
		$-0.888-i0.888$	0
		-0.22	0
Güralp G40T/G3ESP	$5.716 \cdot 10^8$	$-0.148+i0.148$	0
		$-0.148-i0.148$	0
		$-5.026 \cdot 10^2$	
		$-1.005 \cdot 10^3$	
		$-1.131 \cdot 10^3$	

Table B.2: Transfer functions of the URS data set seismometer types described by the gain (A_0) as well as the poles and zeros. The given transfer functions are from velocity to velocity.

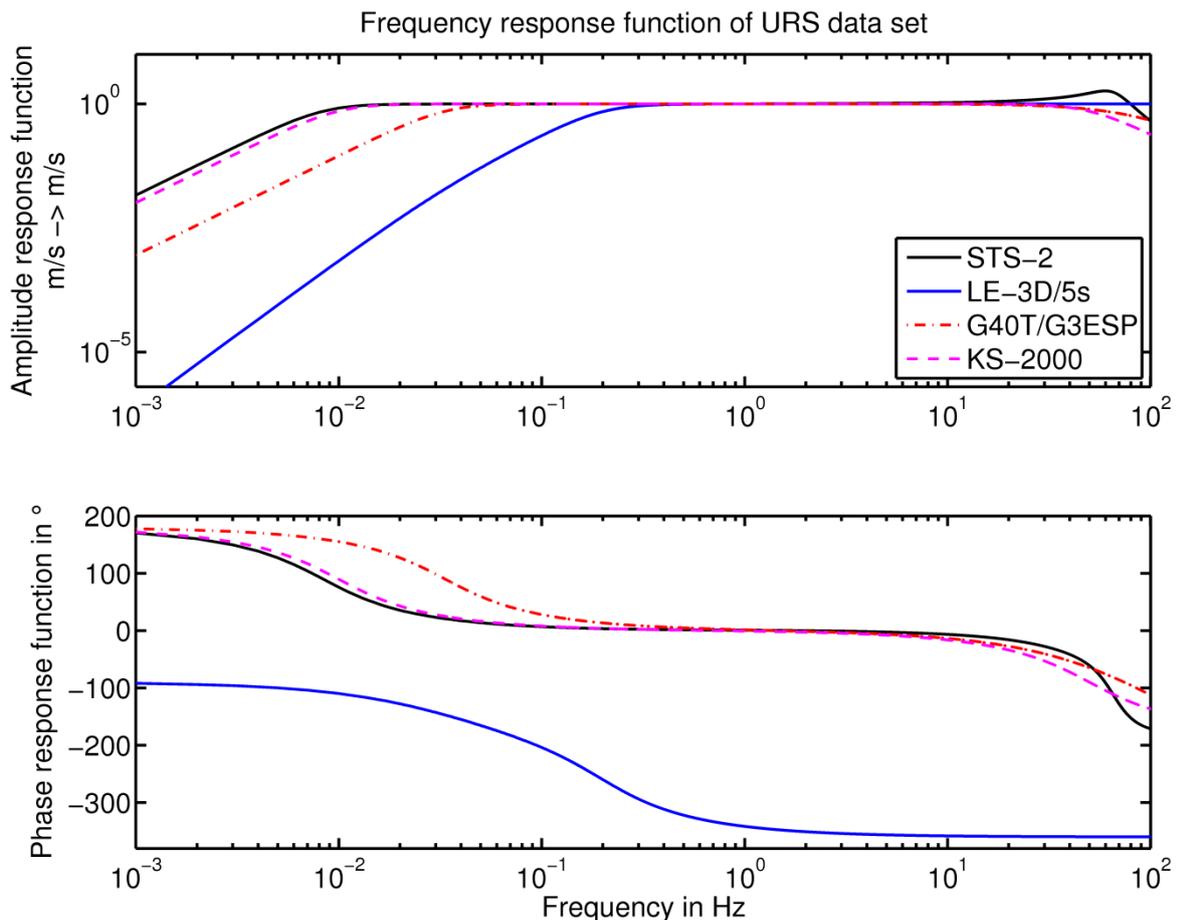


Figure B.1: Amplitude (top) and phase (bottom) frequency response spectra of the seismometer types of the URS data set (see Table B.2)

GSN data set (see section 3.2)

Station / Seismometer	Latitude	Longitude	Gain (A_0) rad/s	Poles rad/s	Zeros rad/s
ANMO / KS-54000	34.9460 N	-106.4571 E	86083	-22.7121+i27.1065	0
				-22.7121-i27.1065	0
				-59.4313	
				-0.0048	
				-0.0731	
CCM / STS-1 VBB	38.0557 N	-91.2446 E	3948.58	-0.0123+i0.0123	0
				-0.0123-i0.0123	0
				-39.18+i49.12	
				-39.18-i49.12	
DWPF / KS-54000	28.1103 N	-81.4327 E	86083	-22.7121+i27.1065	0
				-22.7121-i27.1065	0
				-59.4313	
				-0.0048	
				-0.0731	
HRV / STS-1 VBB	42.5064 N	-71.5583 E	3948.58	-0.0123+i0.0123	0
				-0.0123-i0.0123	0
				-39.18+i49.12	
				-39.18-i49.12	
PFO / STS-1 VBB	33.6092 N	-116.4553 E	50.316	-0.0123+i0.0123	0
				-0.0123-i0.0123	0
				-40.0568+i48.4075	-78.5398
				-40.0568-i48.4075	-0.1525
				-0.1566	-0.1525

Table B.3: Locations and transfer functions of the GSN data set stations described by the gain (A_0) as well as the poles and zeros. The given transfer functions are from velocity to velocity.

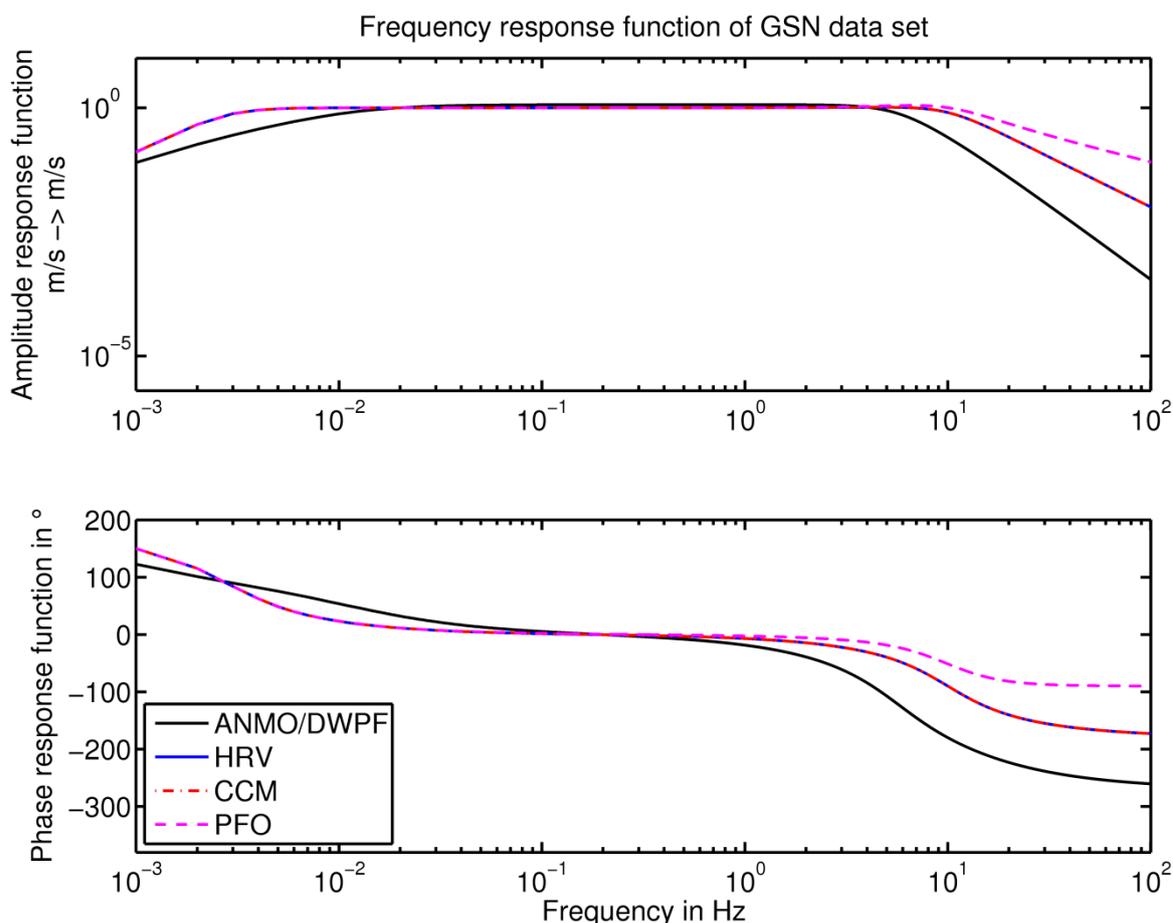


Figure B.2: Amplitude (top) and phase (bottom) frequency response spectra of the seismometer types of the GSN data set (see Table B.3)

Appendix C: Noise classification of the URS data set

This appendix provides the results of the complete classification of the URS data set with consecutive 4 hour long time windows. Each day between 27th October 2003 and 7th August 2004 is covered by six 4 hour time windows starting at 0, 4, 8, 12, 16 and 20 o'clock Eastern European Time (EET). The time window 0-4 EET on 28th March 2004 is excluded from the analysis due to the change from winter time to summer time in this time window. The seismic noise time series (vertical component Z and horizontal components N and E) measured by the URS stations (see Appendix B and Figure 3.1) in these time windows are classified in the eight frequency ranges discussed in section 5.2. The classification results (exactly: the noise class distributions) of all analysed time windows are given in Table C2 (vertical component), C3 (North-South component) and C4 (East-West Component).

Next to the noise class distributions of all time windows also the distributions for several selected data sets are given in Tables C4 to C13. An overview about the selected data sets with the corresponding selection criteria is given in Table C1.

Data set name	Comp.	weekdays	Mean wind velocity	Noise classes	Table
All_Days_Z	Z	all	all	1-6; 11-13	C2
All_Days_N	N	all	all	1-6; 11-13	C3
All_Days_E	E	all	all	1-6; 11-13	C4
Working_Days_LowWind_Z	Z	Mo.-Fri.	< 3 m/s	1-6	C5
Working_Days_LowWind_N	N	Mo.-Fri.	< 3 m/s	1-6	C6
Working_Days_LowWind_E	E	Mo.-Fri.	< 3 m/s	1-6	C7
Working_Days_HighWind_Z	Z	Mo.-Fri.	> 6 m/s	1-6	C8
Working_Days_HighWind_N	N	Mo.-Fri.	> 6 m/s	1-6	C9
Working_Days_HighWind_E	E	Mo.-Fri.	> 6 m/s	1-6	C10
Sundays_LowWind_Z	Z	Sunday	< 3 m/s	1-6	C11
Sundays_LowWind_N	N	Sunday	< 3 m/s	1-6	C12
Sundays_LowWind_E	E	Sunday	< 3 m/s	1-6	C13

Table C1: Specific noise classification data sets selected from the complete URS data set with selection criteria.

Data set: All_Days_Z

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6	NC11-13
				%	%	%	%	%	%	%	
Z	0.008-0.04	0-4	5050	0.06	4.30	4.36	13.98	21.74	0.42	56.44	3.07
Z	0.008-0.04	4-8	5033	0.06	3.72	3.78	24.00	13.97	0.76	55.12	2.38
Z	0.008-0.04	8-12	4995	0.26	8.19	8.45	19.34	17.38	0.46	50.93	3.44
Z	0.008-0.04	12-16	4960	0.24	8.93	9.17	19.92	17.26	0.34	50.52	2.78
Z	0.008-0.04	16-20	5027	0.10	9.53	9.63	19.36	15.99	0.50	51.98	2.55
Z	0.008-0.04	20-24	5052	0.12	4.32	4.43	24.03	15.42	0.69	53.54	1.88
Z	0.04-0.09	0-4	5836	6.44	17.37	23.82	24.01	47.67	0.02	2.47	2.02
Z	0.04-0.09	4-8	5816	3.49	11.64	15.13	30.19	46.41	0.02	4.64	3.61
Z	0.04-0.09	8-12	5771	1.80	8.46	10.26	28.68	53.27	0.00	5.11	2.69
Z	0.04-0.09	12-16	5731	1.76	10.03	11.80	32.26	48.79	0.02	4.80	2.34
Z	0.04-0.09	16-20	5810	2.94	11.26	14.20	34.17	44.72	0.02	4.99	1.91
Z	0.04-0.09	20-24	5839	4.18	13.99	18.17	32.09	43.67	0.03	4.42	1.61
Z	0.09-0.18	0-4	7732	45.65	27.50	73.15	14.34	11.38	0.03	0.57	0.53
Z	0.09-0.18	4-8	7702	41.08	25.03	66.11	14.76	17.50	0.06	0.66	0.90
Z	0.09-0.18	8-12	7643	33.89	22.45	56.34	19.93	22.20	0.07	0.41	1.06
Z	0.09-0.18	12-16	7593	35.11	22.92	58.03	20.15	20.33	0.00	0.34	1.15
Z	0.09-0.18	16-20	7699	38.77	24.09	62.87	16.83	18.29	0.04	0.48	1.49
Z	0.09-0.18	20-24	7733	45.04	25.36	70.40	14.24	14.47	0.05	0.69	0.16
Z	0.18-0.25	0-4	7732	62.67	25.83	88.50	6.39	4.99	0.04	0.00	0.08
Z	0.18-0.25	4-8	7702	59.41	25.23	84.64	8.61	6.19	0.03	0.01	0.52
Z	0.18-0.25	8-12	7642	56.36	26.30	82.66	10.17	6.19	0.04	0.01	0.93
Z	0.18-0.25	12-16	7593	57.21	25.50	82.71	10.13	6.27	0.05	0.00	0.84
Z	0.18-0.25	16-20	7698	58.66	26.12	84.79	7.72	6.77	0.03	0.00	0.70
Z	0.18-0.25	20-24	7733	62.18	25.99	88.17	7.40	4.31	0.09	0.00	0.04
Z	0.25-0.6	0-4	8082	71.13	12.00	83.14	7.52	8.57	0.00	0.01	0.75
Z	0.25-0.6	4-8	8053	68.40	13.90	82.29	8.15	8.98	0.00	0.06	0.52
Z	0.25-0.6	8-12	7988	63.85	15.52	79.37	11.87	7.71	0.00	0.06	0.99
Z	0.25-0.6	12-16	7940	65.28	15.48	80.76	9.56	8.51	0.01	0.01	1.15
Z	0.25-0.6	16-20	8047	70.11	13.84	83.96	7.62	7.54	0.00	0.01	0.87
Z	0.25-0.6	20-24	8082	73.50	12.57	86.07	6.51	7.34	0.00	0.06	0.02
Z	0.6-1	0-4	8082	14.88	19.98	34.87	46.19	18.32	0.05	0.00	0.57
Z	0.6-1	4-8	8053	3.04	13.56	16.60	72.35	10.89	0.00	0.00	0.16
Z	0.6-1	8-12	7988	43.40	22.43	65.84	25.39	8.10	0.00	0.00	0.68
Z	0.6-1	12-16	7940	41.10	22.18	63.27	24.58	11.27	0.00	0.00	0.87
Z	0.6-1	16-20	8047	33.04	35.42	68.46	22.48	8.35	0.00	0.00	0.71
Z	0.6-1	20-24	8082	7.68	49.31	56.99	33.56	9.39	0.01	0.01	0.04
Z	1-25	0-4	8082	0.40	1.08	1.47	12.60	83.59	1.35	0.87	0.12
Z	1-25	4-8	8053	0.06	0.70	0.76	17.53	81.26	0.26	0.07	0.11
Z	1-25	8-12	7988	0.09	0.55	0.64	36.02	61.79	0.29	0.39	0.88
Z	1-25	12-16	7940	0.28	0.69	0.97	35.08	62.88	0.21	0.20	0.65
Z	1-25	16-20	8046	0.31	0.60	0.91	33.22	65.26	0.09	0.25	0.27
Z	1-25	20-24	8082	0.19	0.88	1.06	23.86	74.70	0.19	0.15	0.05
Z	25-45	0-4	8082	1.74	5.09	6.83	19.54	67.10	3.85	0.78	1.91
Z	25-45	4-8	8053	0.10	0.75	0.84	10.74	84.78	1.71	0.96	0.97
Z	25-45	8-12	7988	0.01	0.39	0.40	11.88	83.16	0.98	2.42	1.16
Z	25-45	12-16	7940	0.00	0.31	0.31	11.41	84.14	0.65	2.30	1.17
Z	25-45	16-20	8044	0.07	0.55	0.62	15.17	81.85	0.97	1.02	0.37
Z	25-45	20-24	8082	0.27	1.48	1.76	16.13	79.81	2.03	0.12	0.15

Table C2: Noise class distributions of the complete URS data set (vertical component). The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all time windows.

Data set: All_Days_N

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6	NC11-13
				%	%	%	%	%	%	%	
N	0.008-0.04	0-4	5024	0.02	2.29	2.31	16.26	20.18	0.04	59.08	2.13
N	0.008-0.04	4-8	5000	0.04	1.94	1.98	13.92	20.28	0.24	59.28	4.30
N	0.008-0.04	8-12	4957	0.04	1.25	1.29	14.02	21.53	0.14	58.00	5.02
N	0.008-0.04	12-16	4931	0.00	1.66	1.66	15.57	19.69	0.08	58.20	4.79
N	0.008-0.04	16-20	5000	0.00	1.02	1.02	13.38	20.74	0.10	60.48	4.28
N	0.008-0.04	20-24	5029	0.04	2.13	2.17	14.75	19.21	0.06	61.50	2.31
N	0.04-0.09	0-4	5812	1.14	4.28	5.42	22.51	61.73	0.00	6.93	3.41
N	0.04-0.09	4-8	5782	0.76	2.44	3.20	17.78	60.74	0.00	10.62	7.66
N	0.04-0.09	8-12	5732	0.17	0.96	1.13	19.61	61.24	0.00	11.17	6.86
N	0.04-0.09	12-16	5704	0.16	1.00	1.16	23.81	58.75	0.02	9.36	6.91
N	0.04-0.09	16-20	5783	0.29	1.59	1.88	18.38	62.87	0.02	11.69	5.15
N	0.04-0.09	20-24	5817	0.79	3.25	4.04	18.26	62.71	0.00	11.84	3.15
N	0.09-0.18	0-4	7650	41.37	26.27	67.65	16.50	15.03	0.09	0.10	0.63
N	0.09-0.18	4-8	7614	35.09	23.01	58.10	19.25	21.13	0.04	0.25	1.22
N	0.09-0.18	8-12	7545	25.43	18.38	43.82	22.16	31.99	0.03	0.13	1.87
N	0.09-0.18	12-16	7509	26.37	18.16	44.53	22.89	30.20	0.01	0.20	2.16
N	0.09-0.18	16-20	7613	29.75	20.29	50.05	20.87	26.69	0.04	0.29	2.06
N	0.09-0.18	20-24	7653	38.04	25.41	63.45	18.45	17.39	0.05	0.03	0.63
N	0.18-0.25	0-4	7650	61.02	26.31	87.33	6.78	5.44	0.03	0.00	0.42
N	0.18-0.25	4-8	7614	57.92	26.29	84.21	8.60	6.34	0.08	0.03	0.74
N	0.18-0.25	8-12	7544	53.50	25.46	78.96	11.25	8.54	0.12	0.01	1.11
N	0.18-0.25	12-16	7508	53.18	24.65	77.84	12.28	8.62	0.03	0.01	1.23
N	0.18-0.25	16-20	7613	55.72	25.63	81.35	9.89	7.50	0.03	0.00	1.23
N	0.18-0.25	20-24	7653	62.35	26.21	88.57	6.49	4.53	0.05	0.00	0.35
N	0.25-0.6	0-4	8001	71.73	12.06	83.79	7.21	8.11	0.00	0.01	0.87
N	0.25-0.6	4-8	7965	61.73	17.60	79.33	11.22	8.66	0.00	0.03	0.75
N	0.25-0.6	8-12	7890	61.70	15.96	77.66	13.59	7.62	0.00	0.09	1.05
N	0.25-0.6	12-16	7856	63.25	14.36	77.61	11.24	9.62	0.00	0.04	1.49
N	0.25-0.6	16-20	7961	64.93	15.54	80.47	10.26	8.10	0.00	0.00	1.17
N	0.25-0.6	20-24	8002	72.08	14.53	86.62	6.74	6.32	0.00	0.00	0.32
N	0.6-1	0-4	8001	17.20	23.01	40.21	42.19	16.79	0.05	0.01	0.75
N	0.6-1	4-8	7965	4.12	9.44	13.56	74.63	11.36	0.00	0.00	0.45
N	0.6-1	8-12	7890	44.64	22.81	67.45	24.58	7.10	0.00	0.01	0.86
N	0.6-1	12-16	7854	44.39	22.17	66.55	22.74	9.60	0.00	0.00	1.11
N	0.6-1	16-20	7961	25.55	38.41	63.96	28.06	7.13	0.01	0.00	0.83
N	0.6-1	20-24	8002	7.61	39.93	47.54	45.85	6.27	0.02	0.00	0.31
N	1-25	0-4	8001	1.09	2.70	3.79	26.82	65.08	1.59	2.19	0.54
N	1-25	4-8	7965	0.25	1.47	1.72	33.41	62.81	1.02	0.53	0.51
N	1-25	8-12	7890	0.52	6.64	7.16	42.98	48.09	0.54	0.19	1.04
N	1-25	12-16	7853	0.69	6.48	7.17	42.85	48.44	0.47	0.15	0.92
N	1-25	16-20	7961	0.53	4.70	5.23	43.32	50.48	0.39	0.16	0.41
N	1-25	20-24	8002	0.52	1.91	2.44	38.78	57.14	0.86	0.47	0.31
N	25-45	0-4	8001	1.91	5.30	7.21	15.65	73.44	2.41	0.26	1.02
N	25-45	4-8	7965	0.31	1.19	1.51	7.57	88.42	1.02	0.30	1.18
N	25-45	8-12	7890	0.09	0.42	0.51	12.14	83.68	1.74	0.60	1.34
N	25-45	12-16	7853	0.09	0.20	0.29	10.68	86.02	1.30	0.55	1.16
N	25-45	16-20	7961	0.19	0.82	1.00	12.46	84.94	0.65	0.30	0.64
N	25-45	20-24	8002	0.90	2.17	3.07	12.63	82.48	1.16	0.15	0.50

Table C3: Noise class distributions of the complete URS data set. (North-South component)
The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all time windows.

Data set: All_Days_E

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6	NC11-13
				%	%	%	%	%	%	%	
E	0.008-0.04	0-4	4975	0.10	2.31	2.41	17.13	20.38	0.18	58.83	1.07
E	0.008-0.04	4-8	4955	0.02	1.74	1.76	14.13	20.85	0.34	60.42	2.50
E	0.008-0.04	8-12	4918	0.00	1.24	1.24	15.56	20.07	0.14	58.30	4.70
E	0.008-0.04	12-16	4887	0.00	1.56	1.56	16.55	18.38	0.20	58.48	4.83
E	0.008-0.04	16-20	4954	0.02	1.29	1.31	14.82	20.00	0.28	59.93	3.65
E	0.008-0.04	20-24	4979	0.02	1.83	1.85	15.81	19.50	0.38	61.22	1.25
E	0.04-0.09	0-4	5763	1.06	4.93	5.99	23.56	61.37	0.00	6.35	2.72
E	0.04-0.09	4-8	5737	0.44	2.82	3.26	17.83	63.24	0.02	10.41	5.25
E	0.04-0.09	8-12	5694	0.12	0.74	0.86	20.23	60.94	0.00	11.35	6.62
E	0.04-0.09	12-16	5659	0.14	0.94	1.08	23.18	59.98	0.00	9.44	6.33
E	0.04-0.09	16-20	5737	0.33	1.73	2.06	19.70	62.89	0.00	10.65	4.71
E	0.04-0.09	20-24	5766	0.80	3.42	4.21	19.22	63.32	0.00	10.70	2.55
E	0.09-0.18	0-4	7667	40.13	26.87	67.00	16.55	15.87	0.07	0.05	0.46
E	0.09-0.18	4-8	7636	34.83	24.31	59.14	19.05	20.94	0.03	0.05	0.79
E	0.09-0.18	8-12	7573	25.43	19.48	44.91	21.35	32.11	0.07	0.11	1.45
E	0.09-0.18	12-16	7532	25.86	20.14	46.00	21.57	30.60	0.03	0.16	1.63
E	0.09-0.18	16-20	7635	30.40	20.88	51.28	20.08	26.35	0.07	0.18	2.04
E	0.09-0.18	20-24	7672	39.23	24.50	63.74	17.45	18.34	0.05	0.09	0.33
E	0.18-0.25	0-4	7667	61.13	25.62	86.75	7.50	5.60	0.03	0.00	0.13
E	0.18-0.25	4-8	7636	58.54	26.55	85.08	8.50	5.89	0.04	0.03	0.46
E	0.18-0.25	8-12	7573	54.58	26.04	80.62	10.43	8.24	0.04	0.00	0.67
E	0.18-0.25	12-16	7532	53.49	25.56	79.05	11.70	7.95	0.07	0.00	1.23
E	0.18-0.25	16-20	7635	55.77	26.21	81.98	10.09	6.57	0.03	0.00	1.34
E	0.18-0.25	20-24	7672	60.87	27.50	88.37	7.35	4.14	0.07	0.00	0.07
E	0.25-0.6	0-4	8018	72.71	11.49	84.20	7.65	7.48	0.00	0.01	0.66
E	0.25-0.6	4-8	7987	62.88	18.42	81.29	10.24	7.95	0.00	0.01	0.50
E	0.25-0.6	8-12	7919	63.08	16.34	79.42	13.23	6.65	0.00	0.03	0.67
E	0.25-0.6	12-16	7879	63.74	15.05	78.79	10.57	9.33	0.00	0.00	1.31
E	0.25-0.6	16-20	7982	66.71	15.07	81.78	9.45	7.83	0.01	0.00	0.93
E	0.25-0.6	20-24	8021	72.53	15.30	87.83	6.35	5.77	0.00	0.00	0.05
E	0.6-1	0-4	8018	17.83	22.70	40.53	42.03	16.92	0.00	0.00	0.51
E	0.6-1	4-8	7987	3.66	8.94	12.60	78.24	9.03	0.00	0.00	0.14
E	0.6-1	8-12	7919	43.33	28.31	71.64	22.79	5.03	0.00	0.00	0.54
E	0.6-1	12-16	7879	41.96	28.05	70.01	21.23	7.87	0.00	0.00	0.89
E	0.6-1	16-20	7982	21.94	44.49	66.42	26.95	5.99	0.00	0.00	0.64
E	0.6-1	20-24	8021	7.58	35.72	43.30	50.62	5.95	0.09	0.00	0.05
E	1-25	0-4	8018	1.62	4.22	5.84	27.28	61.04	1.72	3.70	0.42
E	1-25	4-8	7987	0.34	1.60	1.94	32.04	63.98	0.74	1.10	0.20
E	1-25	8-12	7919	0.58	6.65	7.24	40.95	50.27	0.48	0.34	0.72
E	1-25	12-16	7879	0.95	5.80	6.75	41.13	50.81	0.44	0.34	0.52
E	1-25	16-20	7982	0.66	5.11	5.78	41.48	51.70	0.35	0.56	0.13
E	1-25	20-24	8021	0.79	2.26	3.04	39.43	54.91	0.85	1.72	0.05
E	25-45	0-4	8017	3.09	2.63	5.73	16.27	69.69	4.53	2.52	1.27
E	25-45	4-8	7987	1.06	1.67	2.73	8.03	83.27	1.75	3.21	1.01
E	25-45	8-12	7918	0.32	0.78	1.10	9.98	80.50	1.89	5.58	0.95
E	25-45	12-16	7879	0.61	0.75	1.36	9.61	81.51	1.35	5.27	0.91
E	25-45	16-20	7982	1.40	1.40	2.81	11.90	79.92	1.20	3.80	0.38
E	25-45	20-24	8021	2.57	2.04	4.61	10.51	79.92	2.13	2.67	0.16

Table C4: Noise class distributions of the complete URS data set. (East-West component)
The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all time windows.

Data set: WorkingDays_LowWind_Z

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
Z	0.008-0.04	0-4	3057	0.00	4.32	4.32	13.25	23.81	0.39	58.23
Z	0.008-0.04	4-8	2937	0.07	3.44	3.51	26.12	14.95	0.95	54.48
Z	0.008-0.04	8-12	2557	0.23	8.49	8.72	18.54	17.83	0.55	54.36
Z	0.008-0.04	12-16	1962	0.20	9.02	9.23	17.38	21.46	0.31	51.63
Z	0.008-0.04	16-20	2276	0.09	10.15	10.24	17.97	16.83	0.44	54.53
Z	0.008-0.04	20-24	2928	0.03	4.23	4.27	23.57	17.96	0.55	53.65
Z	0.04-0.09	0-4	3596	6.37	17.85	24.22	22.41	50.92	0.03	2.42
Z	0.04-0.09	4-8	3355	3.37	11.06	14.43	30.01	50.73	0.03	4.80
Z	0.04-0.09	8-12	2976	1.85	8.90	10.75	26.61	57.26	0.00	5.38
Z	0.04-0.09	12-16	2262	2.30	11.98	14.28	26.66	54.33	0.04	4.69
Z	0.04-0.09	16-20	2672	3.67	13.44	17.10	32.63	45.17	0.04	5.05
Z	0.04-0.09	20-24	3408	4.05	13.26	17.31	32.10	45.63	0.06	4.90
Z	0.09-0.18	0-4	4804	47.50	27.35	74.85	13.36	11.16	0.02	0.60
Z	0.09-0.18	4-8	4575	42.19	24.11	66.30	14.73	18.14	0.09	0.74
Z	0.09-0.18	8-12	4028	35.10	22.47	57.57	18.77	23.39	0.05	0.22
Z	0.09-0.18	12-16	3059	34.32	23.24	57.57	18.76	23.34	0.00	0.33
Z	0.09-0.18	16-20	3541	40.27	25.28	65.55	16.38	17.74	0.00	0.34
Z	0.09-0.18	20-24	4591	44.02	25.55	69.57	15.16	14.62	0.02	0.63
Z	0.18-0.25	0-4	4819	62.67	25.79	88.46	6.41	5.08	0.04	0.00
Z	0.18-0.25	4-8	4600	59.74	25.93	85.67	8.17	6.11	0.02	0.02
Z	0.18-0.25	8-12	4026	57.85	24.54	82.39	10.61	6.98	0.00	0.02
Z	0.18-0.25	12-16	3077	55.80	24.57	80.37	11.41	8.16	0.06	0.00
Z	0.18-0.25	16-20	3577	60.33	26.47	86.80	7.77	5.42	0.00	0.00
Z	0.18-0.25	20-24	4594	63.21	26.27	89.49	7.60	2.90	0.02	0.00
Z	0.25-0.6	0-4	4989	72.56	11.65	84.21	6.69	9.08	0.00	0.02
Z	0.25-0.6	4-8	4812	68.33	13.90	82.23	8.40	9.31	0.00	0.06
Z	0.25-0.6	8-12	4211	66.16	16.08	82.24	10.73	6.96	0.00	0.07
Z	0.25-0.6	12-16	3200	65.19	15.16	80.34	9.16	10.50	0.00	0.00
Z	0.25-0.6	16-20	3747	73.05	14.30	87.35	6.73	5.92	0.00	0.00
Z	0.25-0.6	20-24	4807	74.72	12.04	86.77	6.47	6.70	0.00	0.06
Z	0.6-1	0-4	5006	14.64	20.06	34.70	48.64	16.60	0.06	0.00
Z	0.6-1	4-8	4842	2.15	8.76	10.90	78.87	10.22	0.00	0.00
Z	0.6-1	8-12	4237	48.45	21.86	70.31	23.27	6.42	0.00	0.00
Z	0.6-1	12-16	3231	44.26	20.64	64.90	22.75	12.35	0.00	0.00
Z	0.6-1	16-20	3754	34.07	41.45	75.52	19.42	5.06	0.00	0.00
Z	0.6-1	20-24	4807	6.07	49.22	55.29	35.59	9.09	0.02	0.00
Z	1-25	0-4	5041	0.36	0.81	1.17	10.47	86.21	1.25	0.89
Z	1-25	4-8	4847	0.02	0.70	0.72	19.25	79.88	0.10	0.04
Z	1-25	8-12	4230	0.12	0.59	0.71	39.24	59.31	0.31	0.43
Z	1-25	12-16	3247	0.31	0.71	1.02	36.86	61.66	0.25	0.22
Z	1-25	16-20	3752	0.37	0.37	0.75	35.10	63.83	0.05	0.27
Z	1-25	20-24	4807	0.12	0.83	0.96	23.40	75.43	0.10	0.10
Z	25-45	0-4	4955	1.63	5.33	6.96	18.79	69.28	4.12	0.85
Z	25-45	4-8	4785	0.06	0.65	0.71	9.68	86.81	1.61	1.19
Z	25-45	8-12	4211	0.02	0.43	0.45	12.51	83.02	1.23	2.78
Z	25-45	12-16	3234	0.00	0.40	0.40	11.13	84.76	0.71	3.00
Z	25-45	16-20	3745	0.05	0.61	0.67	16.29	80.61	0.91	1.52
Z	25-45	20-24	4801	0.33	1.54	1.87	15.46	80.63	1.96	0.08

Table C5: Noise class distributions of the complete URS data set. (vertical component)
 The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all working days with low wind conditions.

Data set: WorkingDays_LowWind_N

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
N	0.008-0.04	0-4	3065	0.03	1.92	1.96	15.60	21.04	0.07	61.34
N	0.008-0.04	4-8	2849	0.00	1.65	1.65	13.16	22.43	0.25	62.51
N	0.008-0.04	8-12	2493	0.04	1.28	1.32	12.56	22.86	0.16	63.10
N	0.008-0.04	12-16	1895	0.00	1.42	1.42	12.93	22.64	0.16	62.85
N	0.008-0.04	16-20	2213	0.00	0.95	0.95	11.88	23.41	0.09	63.67
N	0.008-0.04	20-24	2914	0.03	1.85	1.89	13.66	19.94	0.03	64.48
N	0.04-0.09	0-4	3520	1.31	4.66	5.97	20.17	66.88	0.00	6.99
N	0.04-0.09	4-8	3151	0.95	2.57	3.52	15.65	68.01	0.00	12.82
N	0.04-0.09	8-12	2785	0.29	1.04	1.33	15.80	69.66	0.00	13.21
N	0.04-0.09	12-16	2098	0.33	1.48	1.81	16.30	70.92	0.05	10.92
N	0.04-0.09	16-20	2537	0.35	2.21	2.56	15.06	69.41	0.04	12.93
N	0.04-0.09	20-24	3330	0.78	3.51	4.29	15.56	66.04	0.00	14.11
N	0.09-0.18	0-4	4730	43.42	26.32	69.75	15.58	14.48	0.04	0.15
N	0.09-0.18	4-8	4499	35.19	23.16	58.35	18.20	23.09	0.04	0.31
N	0.09-0.18	8-12	3935	26.58	18.55	45.13	19.24	35.43	0.05	0.15
N	0.09-0.18	12-16	2964	27.02	19.23	46.26	18.66	34.75	0.03	0.30
N	0.09-0.18	16-20	3481	32.00	22.32	54.32	20.51	24.82	0.03	0.32
N	0.09-0.18	20-24	4519	38.75	25.65	64.39	18.72	16.80	0.07	0.02
N	0.18-0.25	0-4	4745	61.73	26.39	88.11	6.01	5.86	0.02	0.00
N	0.18-0.25	4-8	4532	58.98	26.61	85.59	8.12	6.16	0.09	0.04
N	0.18-0.25	8-12	3962	53.66	25.34	79.00	10.55	10.37	0.08	0.00
N	0.18-0.25	12-16	3007	50.55	26.67	77.22	12.11	10.64	0.03	0.00
N	0.18-0.25	16-20	3521	58.68	26.53	85.20	7.67	7.10	0.03	0.00
N	0.18-0.25	20-24	4526	64.76	26.20	90.96	6.16	2.81	0.07	0.00
N	0.25-0.6	0-4	4934	73.25	11.63	84.88	6.34	8.76	0.00	0.02
N	0.25-0.6	4-8	4742	60.19	18.89	79.08	11.77	9.11	0.00	0.04
N	0.25-0.6	8-12	4161	64.31	15.77	80.08	12.02	7.81	0.00	0.10
N	0.25-0.6	12-16	3135	63.76	14.29	78.05	10.49	11.42	0.00	0.03
N	0.25-0.6	16-20	3693	68.37	16.90	85.27	8.39	6.34	0.00	0.00
N	0.25-0.6	20-24	4739	73.33	14.50	87.82	7.13	5.04	0.00	0.00
N	0.6-1	0-4	4947	17.42	22.90	40.33	43.46	16.11	0.08	0.02
N	0.6-1	4-8	4769	3.12	6.14	9.27	79.12	11.62	0.00	0.00
N	0.6-1	8-12	4167	49.60	21.86	71.47	22.94	5.59	0.00	0.00
N	0.6-1	12-16	3167	48.22	19.70	67.92	21.63	10.45	0.00	0.00
N	0.6-1	16-20	3699	25.63	43.44	69.07	26.36	4.54	0.03	0.00
N	0.6-1	20-24	4740	6.48	38.31	44.79	49.58	5.59	0.04	0.00
N	1-25	0-4	4968	0.83	2.03	2.86	25.66	67.75	1.43	2.29
N	1-25	4-8	4760	0.08	1.49	1.58	34.56	62.61	0.88	0.38
N	1-25	8-12	4161	0.58	8.60	9.18	43.84	46.31	0.58	0.10
N	1-25	12-16	3184	1.01	8.20	9.20	42.21	47.99	0.50	0.09
N	1-25	16-20	3697	0.38	5.68	6.06	43.77	49.77	0.38	0.03
N	1-25	20-24	4741	0.42	1.73	2.15	39.13	57.54	0.78	0.40
N	25-45	0-4	4941	2.02	5.18	7.21	15.04	74.88	2.57	0.30
N	25-45	4-8	4717	0.32	1.00	1.31	6.42	90.84	1.06	0.36
N	25-45	8-12	4139	0.07	0.53	0.60	13.63	83.02	2.17	0.58
N	25-45	12-16	3175	0.06	0.22	0.28	11.15	86.24	1.54	0.79
N	25-45	16-20	3683	0.16	0.98	1.14	13.41	84.39	0.62	0.43
N	25-45	20-24	4729	0.89	1.97	2.85	12.77	83.19	1.06	0.13

Table C6: Noise class distributions of the complete URS data set. (North-South component)
The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all working days with low wind conditions.

Data set: WorkingDays_LowWind_E

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
E	0.008-0.04	0-4	3073	0.03	2.31	2.34	16.37	21.90	0.07	59.32
E	0.008-0.04	4-8	2885	0.03	1.35	1.39	13.00	21.70	0.14	63.78
E	0.008-0.04	8-12	2480	0.00	1.05	1.05	12.70	22.46	0.00	63.79
E	0.008-0.04	12-16	1878	0.00	1.54	1.54	13.79	21.35	0.21	63.10
E	0.008-0.04	16-20	2212	0.00	1.40	1.40	13.97	21.43	0.27	62.93
E	0.008-0.04	20-24	2917	0.03	1.65	1.68	14.64	20.60	0.34	62.74
E	0.04-0.09	0-4	3513	1.14	5.29	6.43	21.35	65.90	0.00	6.32
E	0.04-0.09	4-8	3231	0.40	3.10	3.50	16.22	68.40	0.00	11.88
E	0.04-0.09	8-12	2772	0.22	0.94	1.15	16.92	68.90	0.00	13.02
E	0.04-0.09	12-16	2115	0.28	1.37	1.65	17.21	70.45	0.00	10.69
E	0.04-0.09	16-20	2522	0.56	2.50	3.05	17.01	68.24	0.00	11.70
E	0.04-0.09	20-24	3324	0.69	3.67	4.36	17.51	66.40	0.00	11.73
E	0.09-0.18	0-4	4751	41.95	26.67	68.62	15.49	15.81	0.06	0.02
E	0.09-0.18	4-8	4538	34.82	24.99	59.81	16.99	23.14	0.02	0.04
E	0.09-0.18	8-12	3987	26.81	19.96	46.78	18.49	34.54	0.08	0.13
E	0.09-0.18	12-16	2998	26.58	20.65	47.23	17.61	34.96	0.07	0.13
E	0.09-0.18	16-20	3501	32.56	22.62	55.18	20.11	24.45	0.11	0.14
E	0.09-0.18	20-24	4550	39.56	24.53	64.09	18.13	17.67	0.07	0.04
E	0.18-0.25	0-4	4774	61.02	26.31	87.33	6.33	6.30	0.04	0.00
E	0.18-0.25	4-8	4562	59.21	26.72	85.93	8.33	5.68	0.02	0.04
E	0.18-0.25	8-12	4007	54.68	25.93	80.61	9.33	10.03	0.02	0.00
E	0.18-0.25	12-16	3016	52.92	25.93	78.85	12.14	8.99	0.03	0.00
E	0.18-0.25	16-20	3545	57.69	27.19	84.88	8.77	6.32	0.03	0.00
E	0.18-0.25	20-24	4554	63.31	27.27	90.58	6.52	2.83	0.07	0.00
E	0.25-0.6	0-4	4959	73.36	11.57	84.94	6.86	8.21	0.00	0.00
E	0.25-0.6	4-8	4771	60.87	20.23	81.09	10.65	8.24	0.00	0.02
E	0.25-0.6	8-12	4202	64.99	16.85	81.84	11.64	6.47	0.00	0.05
E	0.25-0.6	12-16	3155	65.07	16.01	81.08	8.68	10.24	0.00	0.00
E	0.25-0.6	16-20	3714	70.76	15.51	86.27	7.78	5.95	0.00	0.00
E	0.25-0.6	20-24	4768	73.89	14.68	88.57	7.13	4.30	0.00	0.00
E	0.6-1	0-4	4975	17.25	22.79	40.04	43.46	16.50	0.00	0.00
E	0.6-1	4-8	4803	2.58	5.62	8.20	82.55	9.24	0.00	0.00
E	0.6-1	8-12	4207	48.40	27.83	76.23	19.97	3.80	0.00	0.00
E	0.6-1	12-16	3192	46.40	26.07	72.46	18.80	8.74	0.00	0.00
E	0.6-1	16-20	3721	22.60	50.17	72.78	23.62	3.60	0.00	0.00
E	0.6-1	20-24	4768	6.15	34.48	40.63	54.43	4.82	0.13	0.00
E	1-25	0-4	4994	1.30	3.14	4.45	26.47	63.42	1.80	3.86
E	1-25	4-8	4796	0.25	1.31	1.56	32.88	64.12	0.75	0.69
E	1-25	8-12	4200	0.74	7.50	8.24	42.07	48.90	0.60	0.19
E	1-25	12-16	3219	1.21	6.59	7.80	41.22	50.36	0.43	0.19
E	1-25	16-20	3721	0.46	5.59	6.05	42.27	51.25	0.24	0.19
E	1-25	20-24	4768	0.57	1.93	2.50	39.32	56.17	0.82	1.20
E	25-45	0-4	4949	3.33	2.69	6.02	15.24	71.57	4.51	2.67
E	25-45	4-8	4745	0.86	1.58	2.44	7.29	85.58	1.26	3.41
E	25-45	8-12	4185	0.10	0.48	0.57	10.70	80.38	2.32	6.02
E	25-45	12-16	3203	0.25	0.56	0.81	9.55	82.02	1.72	5.90
E	25-45	16-20	3705	1.16	1.00	2.16	13.31	78.95	1.35	4.24
E	25-45	20-24	4762	2.69	2.04	4.72	9.91	80.83	1.83	2.71

Table C7: Noise class distributions of the complete URS data set. (East-West component)
 The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all working days with low wind conditions.

Data set: WorkingDays_HighWind_Z

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
Z	0.008-0.04	0-4	38	0.00	10.53	10.53	39.47	0.00	0.00	50.00
Z	0.008-0.04	4-8	37	0.00	16.22	16.22	32.43	2.70	0.00	48.65
Z	0.008-0.04	8-12	56	0.00	17.86	17.86	25.00	3.57	0.00	53.57
Z	0.008-0.04	12-16	153	0.00	11.11	11.11	31.37	8.50	0.00	49.02
Z	0.008-0.04	16-20	94	0.00	5.32	5.32	37.23	4.26	3.19	50.00
Z	0.008-0.04	20-24	18	0.00	5.56	5.56	38.89	0.00	0.00	55.56
Z	0.04-0.09	0-4	44	0.00	0.00	0.00	43.18	54.55	0.00	2.27
Z	0.04-0.09	4-8	43	0.00	2.33	2.33	79.07	16.28	0.00	2.33
Z	0.04-0.09	8-12	65	0.00	1.54	1.54	67.69	29.23	0.00	1.54
Z	0.04-0.09	12-16	177	0.00	1.13	1.13	45.76	49.72	0.00	3.39
Z	0.04-0.09	16-20	109	0.92	2.75	3.67	41.28	50.46	0.00	4.59
Z	0.04-0.09	20-24	21	0.00	0.00	0.00	95.24	4.76	0.00	0.00
Z	0.09-0.18	0-4	58	0.00	13.79	13.79	75.86	10.34	0.00	0.00
Z	0.09-0.18	4-8	57	12.28	10.53	22.81	73.68	3.51	0.00	0.00
Z	0.09-0.18	8-12	86	9.30	11.63	20.93	61.63	17.44	0.00	0.00
Z	0.09-0.18	12-16	233	15.45	10.30	25.75	49.79	24.03	0.00	0.43
Z	0.09-0.18	16-20	144	11.11	6.94	18.06	51.39	30.56	0.00	0.00
Z	0.09-0.18	20-24	28	3.57	21.43	25.00	64.29	10.71	0.00	0.00
Z	0.18-0.25	0-4	58	56.90	29.31	86.21	13.79	0.00	0.00	0.00
Z	0.18-0.25	4-8	57	61.40	22.81	84.21	15.79	0.00	0.00	0.00
Z	0.18-0.25	8-12	86	44.19	34.88	79.07	18.60	2.33	0.00	0.00
Z	0.18-0.25	12-16	233	46.35	26.61	72.96	22.75	4.29	0.00	0.00
Z	0.18-0.25	16-20	144	39.58	31.25	70.83	24.31	4.86	0.00	0.00
Z	0.18-0.25	20-24	28	50.00	35.71	85.71	14.29	0.00	0.00	0.00
Z	0.25-0.6	0-4	59	66.10	22.03	88.14	11.86	0.00	0.00	0.00
Z	0.25-0.6	4-8	59	54.24	32.20	86.44	13.56	0.00	0.00	0.00
Z	0.25-0.6	8-12	89	64.04	17.98	82.02	16.85	1.12	0.00	0.00
Z	0.25-0.6	12-16	240	52.92	20.00	72.92	24.58	2.50	0.00	0.00
Z	0.25-0.6	16-20	148	31.08	33.78	64.86	31.08	4.05	0.00	0.00
Z	0.25-0.6	20-24	29	72.41	24.14	96.55	3.45	0.00	0.00	0.00
Z	0.6-1	0-4	59	8.47	47.46	55.93	38.98	5.08	0.00	0.00
Z	0.6-1	4-8	59	28.81	28.81	57.63	35.59	6.78	0.00	0.00
Z	0.6-1	8-12	89	37.08	32.58	69.66	26.97	3.37	0.00	0.00
Z	0.6-1	12-16	240	17.50	34.58	52.08	41.67	6.25	0.00	0.00
Z	0.6-1	16-20	148	6.76	35.81	42.57	45.95	11.49	0.00	0.00
Z	0.6-1	20-24	29	51.72	17.24	68.97	24.14	6.90	0.00	0.00
Z	1-25	0-4	59	0.00	1.69	1.69	38.98	59.32	0.00	0.00
Z	1-25	4-8	59	1.69	0.00	1.69	45.76	49.15	1.69	1.69
Z	1-25	8-12	89	0.00	1.12	1.12	48.31	49.44	0.00	1.12
Z	1-25	12-16	240	0.00	1.67	1.67	38.75	59.58	0.00	0.00
Z	1-25	16-20	147	0.68	1.36	2.04	42.18	55.10	0.00	0.68
Z	1-25	20-24	29	0.00	6.90	6.90	55.17	37.93	0.00	0.00
Z	25-45	0-4	59	0.00	3.39	3.39	30.51	62.71	3.39	0.00
Z	25-45	4-8	59	0.00	0.00	0.00	13.56	83.05	3.39	0.00
Z	25-45	8-12	89	0.00	0.00	0.00	15.73	83.15	0.00	1.12
Z	25-45	12-16	237	0.00	0.00	0.00	12.66	81.01	0.84	5.49
Z	25-45	16-20	148	0.00	0.68	0.68	12.16	85.14	0.68	1.35
Z	25-45	20-24	29	0.00	0.00	0.00	27.59	68.97	3.45	0.00

Table C8: Noise class distributions of the complete URS data set. (vertical component)
 The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all working days with high wind conditions.

Data set: WorkingDays_HighWind_N

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
N	0.008-0.04	0-4	38	0.00	2.63	2.63	34.21	2.63	0.00	60.53
N	0.008-0.04	4-8	37	0.00	2.70	2.70	32.43	18.92	2.70	43.24
N	0.008-0.04	8-12	55	0.00	3.64	3.64	12.73	21.82	0.00	61.82
N	0.008-0.04	12-16	151	0.00	1.99	1.99	22.52	13.25	0.00	62.25
N	0.008-0.04	16-20	93	0.00	5.38	5.38	16.13	18.28	0.00	60.22
N	0.008-0.04	20-24	18	0.00	22.22	22.22	38.89	5.56	0.00	33.33
N	0.04-0.09	0-4	44	0.00	0.00	0.00	75.00	15.91	0.00	9.09
N	0.04-0.09	4-8	43	0.00	0.00	0.00	76.74	20.93	0.00	2.33
N	0.04-0.09	8-12	64	0.00	0.00	0.00	53.13	45.31	0.00	1.56
N	0.04-0.09	12-16	177	0.00	0.00	0.00	45.76	45.76	0.00	8.47
N	0.04-0.09	16-20	107	0.00	0.00	0.00	39.25	54.21	0.00	6.54
N	0.04-0.09	20-24	21	0.00	0.00	0.00	90.48	4.76	0.00	4.76
N	0.09-0.18	0-4	58	0.00	15.52	15.52	62.07	22.41	0.00	0.00
N	0.09-0.18	4-8	57	8.77	12.28	21.05	59.65	19.30	0.00	0.00
N	0.09-0.18	8-12	86	1.16	9.30	10.47	58.14	31.40	0.00	0.00
N	0.09-0.18	12-16	231	9.52	5.19	14.72	42.42	42.86	0.00	0.00
N	0.09-0.18	16-20	144	4.17	6.25	10.42	39.58	50.00	0.00	0.00
N	0.09-0.18	20-24	28	3.57	17.86	21.43	57.14	21.43	0.00	0.00
N	0.18-0.25	0-4	58	44.83	24.14	68.97	29.31	1.72	0.00	0.00
N	0.18-0.25	4-8	57	50.88	26.32	77.19	22.81	0.00	0.00	0.00
N	0.18-0.25	8-12	86	45.35	25.58	70.93	26.74	2.33	0.00	0.00
N	0.18-0.25	12-16	233	45.06	18.03	63.09	27.47	9.44	0.00	0.00
N	0.18-0.25	16-20	144	38.19	20.14	58.33	29.17	11.81	0.69	0.00
N	0.18-0.25	20-24	28	46.43	28.57	75.00	21.43	3.57	0.00	0.00
N	0.25-0.6	0-4	59	61.02	15.25	76.27	20.34	3.39	0.00	0.00
N	0.25-0.6	4-8	59	49.15	27.12	76.27	23.73	0.00	0.00	0.00
N	0.25-0.6	8-12	89	56.18	12.36	68.54	29.21	2.25	0.00	0.00
N	0.25-0.6	12-16	240	50.00	12.08	62.08	30.00	7.92	0.00	0.00
N	0.25-0.6	16-20	148	28.38	33.11	61.49	27.03	11.49	0.00	0.00
N	0.25-0.6	20-24	29	58.62	31.03	89.66	10.34	0.00	0.00	0.00
N	0.6-1	0-4	59	23.73	35.59	59.32	27.12	13.56	0.00	0.00
N	0.6-1	4-8	59	35.59	32.20	67.80	20.34	11.86	0.00	0.00
N	0.6-1	8-12	89	51.69	14.61	66.29	26.97	6.74	0.00	0.00
N	0.6-1	12-16	240	39.17	24.17	63.33	32.92	3.75	0.00	0.00
N	0.6-1	16-20	148	10.81	46.62	57.43	33.11	9.46	0.00	0.00
N	0.6-1	20-24	29	48.28	24.14	72.41	17.24	10.34	0.00	0.00
N	1-25	0-4	59	1.69	8.47	10.17	44.07	44.07	0.00	1.69
N	1-25	4-8	59	3.39	3.39	6.78	50.85	42.37	0.00	0.00
N	1-25	8-12	89	3.37	12.36	15.73	46.07	38.20	0.00	0.00
N	1-25	12-16	240	0.83	8.33	9.17	47.92	42.92	0.00	0.00
N	1-25	16-20	148	0.68	6.08	6.76	47.97	44.59	0.68	0.00
N	1-25	20-24	29	6.90	13.79	20.69	48.28	31.03	0.00	0.00
N	25-45	0-4	59	0.00	1.69	1.69	20.34	76.27	1.69	0.00
N	25-45	4-8	59	0.00	0.00	0.00	10.17	88.14	1.69	0.00
N	25-45	8-12	88	0.00	1.14	1.14	15.91	80.68	2.27	0.00
N	25-45	12-16	236	0.00	0.00	0.00	14.41	83.90	1.27	0.42
N	25-45	16-20	148	0.00	0.00	0.00	12.84	86.49	0.68	0.00
N	25-45	20-24	29	0.00	0.00	0.00	17.24	75.86	6.90	0.00

Table C9: Noise class distributions of the complete URS data set. (North-South component)
The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all working days with high wind conditions.

Data set: WorkingDays_HighWind_E

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
E	0.008-0.04	0-4	38	0.00	0.00	0.00	34.21	2.63	0.00	63.16
E	0.008-0.04	4-8	37	0.00	10.81	10.81	24.32	10.81	0.00	54.05
E	0.008-0.04	8-12	53	0.00	1.89	1.89	20.75	13.21	0.00	64.15
E	0.008-0.04	12-16	151	0.00	2.65	2.65	23.18	17.88	0.00	56.29
E	0.008-0.04	16-20	93	0.00	2.15	2.15	20.43	20.43	0.00	56.99
E	0.008-0.04	20-24	18	0.00	22.22	22.22	27.78	0.00	0.00	50.00
E	0.04-0.09	0-4	44	0.00	0.00	0.00	79.55	20.45	0.00	0.00
E	0.04-0.09	4-8	43	0.00	0.00	0.00	62.79	30.23	0.00	6.98
E	0.04-0.09	8-12	65	0.00	1.54	1.54	46.15	47.69	0.00	4.62
E	0.04-0.09	12-16	175	0.00	0.00	0.00	43.43	47.43	0.00	9.14
E	0.04-0.09	16-20	108	0.00	0.00	0.00	35.19	56.48	0.00	8.33
E	0.04-0.09	20-24	21	0.00	0.00	0.00	85.71	9.52	0.00	4.76
E	0.09-0.18	0-4	57	5.26	14.04	19.30	59.65	21.05	0.00	0.00
E	0.09-0.18	4-8	57	12.28	21.05	33.33	52.63	14.04	0.00	0.00
E	0.09-0.18	8-12	85	5.88	8.24	14.12	57.65	28.24	0.00	0.00
E	0.09-0.18	12-16	229	6.99	10.48	17.47	44.98	37.12	0.00	0.44
E	0.09-0.18	16-20	143	6.29	8.39	14.69	36.36	48.25	0.00	0.70
E	0.09-0.18	20-24	28	10.71	10.71	21.43	60.71	17.86	0.00	0.00
E	0.18-0.25	0-4	57	42.11	33.33	75.44	24.56	0.00	0.00	0.00
E	0.18-0.25	4-8	57	54.39	26.32	80.70	17.54	1.75	0.00	0.00
E	0.18-0.25	8-12	85	49.41	22.35	71.76	22.35	5.88	0.00	0.00
E	0.18-0.25	12-16	229	39.74	25.76	65.50	26.64	7.86	0.00	0.00
E	0.18-0.25	16-20	143	36.36	24.48	60.84	27.97	11.19	0.00	0.00
E	0.18-0.25	20-24	28	57.14	21.43	78.57	21.43	0.00	0.00	0.00
E	0.25-0.6	0-4	58	67.24	13.79	81.03	17.24	1.72	0.00	0.00
E	0.25-0.6	4-8	59	47.46	28.81	76.27	22.03	1.69	0.00	0.00
E	0.25-0.6	8-12	88	59.09	12.50	71.59	22.73	5.68	0.00	0.00
E	0.25-0.6	12-16	236	52.54	13.14	65.68	23.73	10.59	0.00	0.00
E	0.25-0.6	16-20	147	30.61	33.33	63.95	21.77	14.29	0.00	0.00
E	0.25-0.6	20-24	29	41.38	41.38	82.76	17.24	0.00	0.00	0.00
E	0.6-1	0-4	58	29.31	37.93	67.24	24.14	8.62	0.00	0.00
E	0.6-1	4-8	59	28.81	38.98	67.80	22.03	10.17	0.00	0.00
E	0.6-1	8-12	88	55.68	19.32	75.00	18.18	6.82	0.00	0.00
E	0.6-1	12-16	236	38.98	30.08	69.07	25.85	5.08	0.00	0.00
E	0.6-1	16-20	147	17.01	42.18	59.18	33.33	7.48	0.00	0.00
E	0.6-1	20-24	29	62.07	13.79	75.86	17.24	6.90	0.00	0.00
E	1-25	0-4	58	0.00	8.62	8.62	43.10	43.10	1.72	3.45
E	1-25	4-8	59	1.69	5.08	6.78	49.15	44.07	0.00	0.00
E	1-25	8-12	88	3.41	13.64	17.05	42.05	40.91	0.00	0.00
E	1-25	12-16	236	0.85	6.36	7.20	43.64	47.88	0.42	0.85
E	1-25	16-20	147	0.68	6.12	6.80	47.62	43.54	0.68	1.36
E	1-25	20-24	29	3.45	17.24	20.69	51.72	27.59	0.00	0.00
E	25-45	0-4	58	0.00	0.00	0.00	22.41	72.41	3.45	1.72
E	25-45	4-8	58	0.00	1.72	1.72	8.62	86.21	3.45	0.00
E	25-45	8-12	87	0.00	0.00	0.00	12.64	80.46	1.15	5.75
E	25-45	12-16	232	0.00	0.00	0.00	12.50	80.60	0.43	6.47
E	25-45	16-20	147	0.00	0.00	0.00	13.61	82.31	1.36	2.72
E	25-45	20-24	29	0.00	0.00	0.00	24.14	68.97	6.90	0.00

Table C10: Noise class distributions of the complete URS data set. (East-West component)
 The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all working days with high wind conditions.

Data set: Sundays_LowWind_Z

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
Z	0.008-0.04	0-4	554	0.36	4.87	5.23	13.00	21.30	1.08	59.39
Z	0.008-0.04	4-8	639	0.16	3.44	3.60	21.75	15.02	0.31	59.31
Z	0.008-0.04	8-12	535	0.00	7.85	7.85	20.00	21.31	0.19	50.65
Z	0.008-0.04	12-16	443	0.23	9.03	9.26	21.22	17.16	0.00	52.37
Z	0.008-0.04	16-20	434	0.23	8.06	8.29	20.05	18.66	0.46	52.53
Z	0.008-0.04	20-24	617	0.00	4.38	4.38	25.61	13.61	0.81	55.59
Z	0.04-0.09	0-4	632	11.71	22.15	33.86	15.51	48.89	0.00	1.74
Z	0.04-0.09	4-8	742	3.77	16.44	20.22	26.15	48.52	0.00	5.12
Z	0.04-0.09	8-12	610	1.97	9.84	11.80	23.77	59.34	0.00	5.08
Z	0.04-0.09	12-16	514	1.56	8.17	9.73	35.80	49.22	0.00	5.25
Z	0.04-0.09	16-20	508	4.33	10.24	14.57	25.98	52.76	0.00	6.69
Z	0.04-0.09	20-24	701	6.13	18.26	24.39	25.11	46.65	0.00	3.85
Z	0.09-0.18	0-4	872	42.78	27.06	69.84	17.20	12.04	0.00	0.92
Z	0.09-0.18	4-8	984	42.78	29.37	72.15	9.76	17.58	0.00	0.51
Z	0.09-0.18	8-12	828	33.94	20.17	54.11	20.53	24.03	0.12	1.21
Z	0.09-0.18	12-16	683	39.68	18.16	57.83	18.45	23.13	0.00	0.59
Z	0.09-0.18	16-20	653	43.03	23.28	66.31	13.02	19.91	0.00	0.77
Z	0.09-0.18	20-24	958	46.97	25.78	72.76	10.44	15.76	0.00	1.04
Z	0.18-0.25	0-4	872	64.79	25.92	90.71	5.16	4.13	0.00	0.00
Z	0.18-0.25	4-8	985	62.13	23.15	85.28	7.21	7.41	0.10	0.00
Z	0.18-0.25	8-12	836	53.83	29.31	83.13	8.49	8.37	0.00	0.00
Z	0.18-0.25	12-16	684	54.24	26.02	80.26	8.77	10.96	0.00	0.00
Z	0.18-0.25	16-20	654	58.87	25.38	84.25	5.35	10.40	0.00	0.00
Z	0.18-0.25	20-24	958	62.11	26.83	88.94	6.68	4.28	0.10	0.00
Z	0.25-0.6	0-4	915	70.60	10.05	80.66	9.62	9.73	0.00	0.00
Z	0.25-0.6	4-8	1033	66.70	14.91	81.61	7.94	10.45	0.00	0.00
Z	0.25-0.6	8-12	880	64.32	11.59	75.91	10.00	14.09	0.00	0.00
Z	0.25-0.6	12-16	719	64.67	12.80	77.47	7.23	15.30	0.00	0.00
Z	0.25-0.6	16-20	667	71.66	11.39	83.06	8.25	8.55	0.00	0.15
Z	0.25-0.6	20-24	1004	70.42	13.75	84.16	9.26	6.47	0.00	0.10
Z	0.6-1	0-4	915	17.27	17.60	34.86	37.92	27.10	0.11	0.00
Z	0.6-1	4-8	1032	5.81	24.81	30.62	51.26	18.12	0.00	0.00
Z	0.6-1	8-12	880	29.20	24.09	53.30	26.48	20.23	0.00	0.00
Z	0.6-1	12-16	712	39.61	21.21	60.81	18.96	20.22	0.00	0.00
Z	0.6-1	16-20	674	41.54	19.73	61.28	22.70	16.02	0.00	0.00
Z	0.6-1	20-24	1004	11.75	50.10	61.85	28.69	9.36	0.00	0.10
Z	1-25	0-4	915	0.55	2.30	2.84	15.52	79.13	1.64	0.87
Z	1-25	4-8	1033	0.19	0.58	0.77	11.62	87.03	0.58	0.00
Z	1-25	8-12	880	0.11	0.68	0.80	22.73	75.91	0.45	0.11
Z	1-25	12-16	719	0.14	1.25	1.39	28.93	69.54	0.00	0.14
Z	1-25	16-20	712	0.28	1.12	1.40	29.07	69.52	0.00	0.00
Z	1-25	20-24	1003	0.30	0.70	1.00	23.23	75.37	0.20	0.20
Z	25-45	0-4	886	2.14	4.85	7.00	21.11	67.27	3.95	0.68
Z	25-45	4-8	1033	0.19	0.97	1.16	12.78	83.35	2.23	0.48
Z	25-45	8-12	879	0.00	0.34	0.34	8.76	89.19	0.80	0.91
Z	25-45	12-16	718	0.00	0.00	0.00	11.28	88.02	0.42	0.28
Z	25-45	16-20	713	0.14	0.42	0.56	13.60	84.29	1.40	0.14
Z	25-45	20-24	1001	0.30	1.60	1.90	14.79	81.32	2.00	0.00

Table C11: Noise class distributions of the complete URS data set. (vertical component)
The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all Sunday with low wind conditions.

Data set: Sundays_LowWind_N

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
N	0.008-0.04	0-4	549	0.00	3.10	3.10	12.93	22.40	0.00	61.57
N	0.008-0.04	4-8	620	0.00	2.42	2.42	15.32	18.55	0.16	63.55
N	0.008-0.04	8-12	522	0.00	1.15	1.15	13.22	24.14	0.00	61.49
N	0.008-0.04	12-16	432	0.00	1.39	1.39	16.67	21.53	0.00	60.42
N	0.008-0.04	16-20	421	0.00	0.71	0.71	13.54	22.57	0.00	63.18
N	0.008-0.04	20-24	602	0.00	2.49	2.49	15.61	21.93	0.17	59.80
N	0.04-0.09	0-4	621	1.61	7.89	9.50	19.32	64.73	0.00	6.44
N	0.04-0.09	4-8	707	0.99	3.54	4.53	15.28	67.75	0.00	12.45
N	0.04-0.09	8-12	596	0.00	1.34	1.34	19.46	66.11	0.00	13.09
N	0.04-0.09	12-16	491	0.20	1.83	2.04	23.01	63.54	0.00	11.41
N	0.04-0.09	16-20	485	0.62	2.27	2.89	14.02	67.01	0.00	16.08
N	0.04-0.09	20-24	688	1.45	4.51	5.96	14.68	68.02	0.00	11.34
N	0.09-0.18	0-4	860	44.65	24.65	69.30	13.84	16.63	0.12	0.12
N	0.09-0.18	4-8	972	39.81	23.77	63.58	20.68	15.53	0.00	0.21
N	0.09-0.18	8-12	808	26.73	18.81	45.54	21.91	32.30	0.00	0.25
N	0.09-0.18	12-16	667	27.59	17.09	44.68	24.59	30.73	0.00	0.00
N	0.09-0.18	16-20	641	34.32	21.37	55.69	16.22	27.46	0.16	0.47
N	0.09-0.18	20-24	940	37.34	26.28	63.62	17.23	19.15	0.00	0.00
N	0.18-0.25	0-4	861	62.25	27.18	89.43	6.74	3.83	0.00	0.00
N	0.18-0.25	4-8	972	59.67	26.44	86.11	6.48	7.41	0.00	0.00
N	0.18-0.25	8-12	820	54.63	28.54	83.17	7.56	9.02	0.24	0.00
N	0.18-0.25	12-16	672	54.76	22.47	77.23	10.12	12.65	0.00	0.00
N	0.18-0.25	16-20	635	56.69	25.35	82.05	11.02	6.93	0.00	0.00
N	0.18-0.25	20-24	943	60.55	28.31	88.87	6.57	4.56	0.00	0.00
N	0.25-0.6	0-4	904	75.33	6.86	82.19	9.07	8.74	0.00	0.00
N	0.25-0.6	4-8	1020	61.86	18.73	80.59	11.27	8.14	0.00	0.00
N	0.25-0.6	8-12	864	60.53	14.93	75.46	13.54	11.00	0.00	0.00
N	0.25-0.6	12-16	708	66.24	12.43	78.67	5.79	15.54	0.00	0.00
N	0.25-0.6	16-20	648	67.90	15.12	83.02	11.27	5.71	0.00	0.00
N	0.25-0.6	20-24	989	73.81	16.68	90.50	4.75	4.75	0.00	0.00
N	0.6-1	0-4	904	15.04	22.01	37.06	40.93	22.01	0.00	0.00
N	0.6-1	4-8	1017	5.90	16.42	22.32	62.44	15.24	0.00	0.00
N	0.6-1	8-12	864	28.13	23.73	51.85	29.28	18.87	0.00	0.00
N	0.6-1	12-16	700	36.86	25.00	61.86	20.71	17.43	0.00	0.00
N	0.6-1	16-20	665	34.44	22.86	57.29	27.37	15.34	0.00	0.00
N	0.6-1	20-24	989	9.61	42.67	52.28	40.65	7.08	0.00	0.00
N	1-25	0-4	905	2.10	4.86	6.96	27.29	61.88	1.77	2.10
N	1-25	4-8	1020	0.49	1.08	1.57	30.29	65.00	1.67	1.47
N	1-25	8-12	864	0.12	1.97	2.08	38.19	58.22	0.93	0.58
N	1-25	12-16	709	0.42	3.10	3.53	40.48	54.72	0.56	0.71
N	1-25	16-20	701	1.00	2.43	3.42	41.23	54.21	0.57	0.57
N	1-25	20-24	989	0.61	1.62	2.22	38.42	57.23	1.31	0.81
N	25-45	0-4	891	2.13	5.72	7.86	18.18	71.83	2.02	0.11
N	25-45	4-8	1020	0.20	2.16	2.35	11.18	85.59	0.78	0.10
N	25-45	8-12	863	0.00	0.23	0.23	5.91	92.82	0.70	0.35
N	25-45	12-16	707	0.00	0.42	0.42	6.22	93.07	0.28	0.00
N	25-45	16-20	701	0.29	0.71	1.00	9.99	88.45	0.57	0.00
N	25-45	20-24	988	0.81	3.04	3.85	11.23	84.01	0.91	0.00

Table C12: Noise class distributions of the complete URS data set. (North-South component)
The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all Sunday with low wind conditions.

Data set: Sundays_LowWind_E

Comp.	Freq. band /Hz	EET	#	NC1	NC2	NC1+2	NC3	NC4	NC5	NC6
				%	%	%	%	%	%	
E	0.008-0.04	0-4	552	0.18	3.26	3.44	11.23	23.19	0.36	61.78
E	0.008-0.04	4-8	622	0.00	2.25	2.25	14.79	22.19	1.29	59.49
E	0.008-0.04	8-12	518	0.00	0.97	0.97	19.31	20.85	0.77	58.11
E	0.008-0.04	12-16	418	0.00	0.96	0.96	19.14	21.53	0.48	57.89
E	0.008-0.04	16-20	413	0.00	1.21	1.21	11.62	20.34	0.97	65.86
E	0.008-0.04	20-24	606	0.00	2.15	2.15	14.69	19.14	0.66	63.37
E	0.04-0.09	0-4	628	2.07	7.96	10.03	18.15	64.81	0.00	7.01
E	0.04-0.09	4-8	706	0.57	3.26	3.82	15.86	69.69	0.00	10.62
E	0.04-0.09	8-12	587	0.00	1.19	1.19	20.44	67.12	0.00	11.24
E	0.04-0.09	12-16	480	0.21	2.50	2.71	21.88	66.88	0.00	8.54
E	0.04-0.09	16-20	478	0.63	2.93	3.56	13.39	70.08	0.00	12.97
E	0.04-0.09	20-24	684	2.34	3.95	6.29	16.37	65.50	0.00	11.84
E	0.09-0.18	0-4	865	40.81	27.28	68.09	12.02	19.65	0.00	0.23
E	0.09-0.18	4-8	976	37.50	24.39	61.89	21.21	16.80	0.10	0.00
E	0.09-0.18	8-12	806	29.40	16.50	45.91	21.34	32.63	0.00	0.12
E	0.09-0.18	12-16	676	29.29	16.86	46.15	15.53	38.31	0.00	0.00
E	0.09-0.18	16-20	625	37.60	21.60	59.20	16.32	24.00	0.00	0.48
E	0.09-0.18	20-24	948	42.19	20.78	62.97	16.24	20.68	0.00	0.11
E	0.18-0.25	0-4	867	61.94	25.37	87.31	8.65	4.04	0.00	0.00
E	0.18-0.25	4-8	976	59.73	26.74	86.48	6.15	7.38	0.00	0.00
E	0.18-0.25	8-12	828	54.95	27.05	82.00	9.30	8.57	0.12	0.00
E	0.18-0.25	12-16	677	52.29	24.37	76.66	7.98	15.07	0.30	0.00
E	0.18-0.25	16-20	623	58.75	26.48	85.23	10.91	3.69	0.16	0.00
E	0.18-0.25	20-24	948	60.13	28.16	88.29	7.70	3.90	0.11	0.00
E	0.25-0.6	0-4	909	74.59	8.14	82.73	9.79	7.48	0.00	0.00
E	0.25-0.6	4-8	1024	65.63	17.38	83.01	9.08	7.91	0.00	0.00
E	0.25-0.6	8-12	871	60.85	14.58	75.43	15.50	9.07	0.00	0.00
E	0.25-0.6	12-16	713	67.32	11.92	79.24	5.33	15.43	0.00	0.00
E	0.25-0.6	16-20	653	67.84	15.77	83.61	11.49	4.90	0.00	0.00
E	0.25-0.6	20-24	994	74.45	17.81	92.25	3.42	4.33	0.00	0.00
E	0.6-1	0-4	909	15.62	20.24	35.86	42.35	21.78	0.00	0.00
E	0.6-1	4-8	1021	5.97	13.61	19.59	68.07	12.34	0.00	0.00
E	0.6-1	8-12	872	20.87	30.50	51.38	32.91	15.71	0.00	0.00
E	0.6-1	12-16	707	29.70	31.97	61.67	23.20	15.13	0.00	0.00
E	0.6-1	16-20	665	29.02	26.47	55.49	31.28	13.23	0.00	0.00
E	0.6-1	20-24	994	9.36	36.02	45.37	47.99	6.64	0.00	0.00
E	1-25	0-4	909	3.19	6.93	10.12	25.19	59.63	1.87	3.19
E	1-25	4-8	1024	0.59	2.25	2.83	28.42	64.84	1.17	2.73
E	1-25	8-12	871	0.34	3.79	4.13	36.39	57.75	0.57	1.15
E	1-25	12-16	715	0.84	4.20	5.03	38.60	54.41	0.56	1.40
E	1-25	16-20	705	1.28	3.97	5.25	41.13	51.35	0.57	1.70
E	1-25	20-24	994	0.80	2.62	3.42	37.42	54.63	1.01	3.52
E	25-45	0-4	889	3.60	2.25	5.85	19.46	67.49	4.84	2.36
E	25-45	4-8	1024	2.54	1.07	3.61	10.06	80.37	3.42	2.54
E	25-45	8-12	869	1.84	2.19	4.03	5.75	85.73	0.81	3.68
E	25-45	12-16	715	2.52	1.26	3.78	6.99	85.17	0.70	3.36
E	25-45	16-20	705	2.41	1.99	4.40	8.51	83.12	1.42	2.55
E	25-45	20-24	994	2.31	1.81	4.12	10.87	79.88	2.52	2.62

Table C13: Noise class distributions of the complete URS data set. (East-West component)
The table displays the NC distributions in percent for the frequency bands between 0.008-45 Hz obtained from all Sunday with low wind conditions.

List of Figures

Figure 3.1: Station network of the URS project.	18
Figure 3.2: Map with used GSN stations in the USA.	19
Figure 3.3: Block diagram of the BHZ channel (vertical component, 20 Hz) of GSN station HRV.	22
Figure 3.4: Frequency responses of the complete BHZ channel and the seismometer of GSN station HRV.	23
Figure 3.5: Filter coefficients b_i and phase response spectra of two DAA stages of HRV.BHZ.	26
Figure 3.6: Filter coefficients b_i with amplitude and phase response spectra of stages 3 and 5 of HRV.BHZ.	29
Figure 4.1: Standard Gaussian distributed stochastic process with histogram.	34
Figure 4.2: Time series of vertical component urban seismic noise with their histograms.	36
Figure 5.1: Spectrograms of the vertical-component USN in Bucharest during March 2004.	46
Figure 5.2: Frequency-wave number (FK) analysis of vertical-component USN in Bucharest.	47
Figure 5.3: Maps with seismic noise amplitudes (colour) and classes (symbols) at the URS station sites.	52
Figure 5.4: Maps with seismic noise amplitudes and classes at the URS station sites (0.6-1 Hz).	53
Figure 5.5: Seismic noise amplitudes and classes in the frequency range 0.6-1 Hz.	54
Figure 5.6: Seismic noise amplitudes and classes for the vertical-component USN (0.04-0.6 Hz).	55
Figure 5.7: South-North profiles of the seismic noise amplitudes and classes (0.25-0.6 Hz).	56
Figure 5.8: South-North profiles of the seismic noise amplitudes and classes (0.008-0.25 Hz).	57
Figure 5.9: Map of seismic noise amplitudes and classes in Bucharest (0.008-0.04 Hz).	59
Figure 5.10: Time-domain and spectral H/V ratios observed in the Bucharest area.	68
Figure 5.11: The U-matrix with a possible clustering of the SOM for the vert.-comp. USN (0.6-1Hz).	73
Figure 5.12: Amplitude components of the USN (0.6-1 Hz) data vectors after the SOM classification.	73
Figure 5.13: Relation between the average wind velocity and the SOM classification.	74
Figure 5.14: Relation between weekday and daytime of the time windows with the SOM classification.	76
Figure 6.1: Processing scheme for the cross-correlation processing after Bensen et al. (2007).	83
Figure 6.2: Processing scheme for the cross-correlation processing realised for this thesis.	84
Figure 6.3: A comparison of CCFs obtained from the same data with different normalisation schemes.	88
Figure 6.4: Comparison of overlapping and non-overlapping time windows and their computational costs.	90
Figure 6.5: Influence of the time window length on CCFs obtained with different spectral whitening strategies.	92

Figure 6.6: Influence of the time window length on CCFs obtained with spectral whitening.	94
Figure 6.7: Histogram-based spectral analysis of single time window CCFs.	95
Figure 6.8: Significance of the calculated waveform symmetry coefficient (WSC) values.	97
Figure 6.9: Comparison of 12-months CCFs obtained with different time domain normalisations (ram+one bit).	99
Figure 6.10: Comparison of 12-months CCFs obtained with different time domain normalisations (ram+wpcf).	101
Figure 6.11: Influence of the time window length on the CCFs obtained with different normalisation schemes.	102
Figure 6.12: Influence of normalisation and time window length on the obtained CCFs (without data selection).	103
Figure 6.13: Number of time windows in percent selected by the WSA and TSA data selection approaches.	105
Figure 6.14: Influence of the TSA data selection approach on the obtained CCFs.	106
Figure 6.15: Influence of the TSA data selection approach on the CCF waveform symmetry.	107
Figure 6.16: Influence of the WSA data selection approach on the obtained CCFs.	108
Figure 6.17: Influence of the WSA data selection approach on the CCF waveform symmetry.	109
Figure 6.18: Distance-dependent illustration of the CCFs (7-150 s) of all ten station pairs.	110

List of Tables

Table 3.1: Summary of the delay times of channel IU.HRV.BHZ.....	29
Table 3.2: Summary of the instrument response metadata analysis.	31
Table 4.1: The time series parameters used for the classification (with abbreviations).	35
Table 4.2: The classification scheme used to classify time series of seismic noise.	38
Table 4.3: Results of the classification test with synthetic Gaussian noise.	40
Table 5.1: Selection criteria for earthquakes which may be observed at the URS network.	44
Table 5.2: Frequency ranges for the statistical time-series analysis of urban seismic noise.	48
Table 5.3: All working days selected for the first noise classification analysis.	49
Table 5.4: The noise class distributions for the vertical-comp. USN in Bucharest (selected working days)	50
Table 5.5: Comparison of the noise class distributions for selected / all working days (low wind conditions).....	61
Table 5.6: Comparison of the noise class distributions for working days / Sundays (low wind conditions).....	62
Table 5.7: Median noise amplitudes on working days and Sundays.....	63
Table 5.8: Comparison of the noise class distributions for working days with low / high wind velocities	64
Table 5.9: Comparison of the noise class distributions for the components Z, N and E.	66
Table 5.10: The 15 largest earthquakes which occurred during the URS experiment.....	77
Table 5.11: The 15 closest regional earthquakes which occurred during the URS experiment.	78
Table 5.12: The five largest local earthquakes which occurred in the Vrancea subduction zone.....	78
Table 6.1: Overview about the applied normalisation schemes and the relevant figures.	85
Table 6.2: Interstation distances with corresponding signal and noise time windows of several station pairs.	87

List of Abbreviations

1B	1 Bit, one-bit, normalisation (chapter 6)
ADC	Analogue-to-Digital Converter
CCF	Cross-Correlation Function
CLT	Central Limit Theorem
DAA	Digital Anti-Alias
DHI	Direct Hydrocarbon Indicator
DMC	Data Management Centre
FDSN	International Federation of Digital Seismograph Networks
FIR	Finite Impulse Response
FK	Frequency-wavenumber (array technique)
GSN	Global Seismographic Network
IIR	Infinite Impulse Response
int68	68% amplitude interval (chapter 4)
IRIS	Incorporated Research Institutions for Seismology
KABBA	KARlsruhe BroadBand Array
KaSP	Karlsruhe Seismology Processing toolbox for MATLAB
ram	running absolute mean (normalisation, chapter 6)
pf	peakfactor (chapter 4)
psd	power spectral density
SEED	Standard for the Exchange of Earthquake Data
SOM	Self-Organizing Map
SPAC	SPatial AutoCorrelation (array technique)
SW	Spectral Whitening (of cross correlation functions)
TSA	Time Series Approach (data selection, chapter 6)
TSSW	Time Series Spectral Whitening
URS	URban Seismology
USN	Urban Seismic Noise
wpcf	waveform preserving normalisation of CCFs (chapter 6)
wpts	waveform preserving normalisation of time series (chapter 6)
WSA	Waveform Symmetry Approach (data selection, chapter 6)

References

- Aki, K., 1957. Space and time spectra of stationary stochastic waves with special reference to microtremors. *Bulletin of the Earthquake Research Institute*, **35**, 415-456.
- Aki, K. & Richards, P. G., 2002. *Quantitative Seismology, Second Edition*. University Science Books, Sausalito.
- Aster, R. C., McNamara, D. E. & Bromirski, P. D., 2008. Multidecadal climate-induced variability in microseism. *Seismological Research Letters*, **79**, 194-202.
- Beauduin, R., Lognonne, P., Montagner, J. P., Cacho, S., Karczewski, J. F. & Morand, M., 1996. The effects of the atmospheric pressure changes on seismic signals. *Bull. Seism. Soc. Am.*, **86**, 1760–1769.
- Bendat, J. S. & Piersol, A. G., 1994. *Random data: Analysis and measurement procedures*, 2. ed., rev. and expanded, 8. print, A Wiley-Interscience publication, New York.
- 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.
- Bertelli, B., 1872. Osservazioni sui piccolo movimenti dei pendoli in relazione ad alcuni fenomeni meteorologiche. *Bullettino Meteorologico dell' Osservatorio dell' Collegio Romano*, Roma, Italy, **101**. (In Italian).
- Bokelmann, G. H. & Baisch, S., 1999. Nature of narrow-band signals at 2.083 Hz. *Bull. Seism. Soc. Am.*, **89**, 156–164.
- Bonnefoy-Claudet, S., Cotton, F. & Bard, P.-Y., 2006a. The nature of noise wavefield and its applications for site effects studies: A literature review. *Earth-Sci. Rev.*, **79**, 205–227.
- Bonnefoy-Claudet, S., Cornou, C., Bard, Y.-B., Cotton, F., Moczo, P., Kristek, J. & Fäh, D., 2006b. H/V ratio: a tool for site effects evaluation. Results from 1-D noise simulations. *Geophysical Journal International*, **167**, 827-837.
- Bonnefoy-Claudet, S., Köhler, A., Cornou, C., Wathelet, M. & Bard, P.-Y., 2008. Effects of Love Waves on Microtremor H/V Ratio. *Bulletin of the Seismological Society of America*, **98**, 288-300.
- Brenguier, F., Shapiro, N. M., Campillo, M., Nercessian, A. & Ferrazzini, V., 2007. 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations. *Geophysical Research Letters*, **34**, L02305.
- Brenguier, F., Shapiro, N. M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O. & Nercessian, A., 2008a. Towards forecasting volcanic eruptions using seismic noise. *Nature Geoscience*, **1**, 1-5.
- Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M. & Larose, E., 2008b. Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations. *Science*, **321**, 1478-1481.
- Bromirski, P. D. & Duennebieer, F. K., 2002. The near-coastal microseism spectrum: Spatial and temporal wave climate relationships. *Journal of Geophysical Research*, **107**, 5.1-5.20.

- Bromirski, P. D., Flick, R. E. & Graham, N., 1999. Ocean wave height determined from inland seismometer data: Implications for investigating wave climate changes in the NE Pacific. *Journal of Geophysical Research*, **104**, 20.753-20.766.
- Brooks, L. A., Townend, J., Gerstoft, P., Bannister, S. & Lionel, C., 2009. Fundamental and higher-mode Rayleigh wave characteristics of ambient seismic noise in New Zealand. *Geophysical Research Letters*, **36**, L23303.
- Bull, C. R., Bull, R. M. & Rastin, B. C., 1992. On the sensitivity of the chi-square test and its consequences. *Measurement Science and Technology*, **3**, 789-795.
- Burtin, A., Bollinger, L., Vergne, J., Cattin, R. & Nábělek, J. L., 2008. Spectral analysis of seismic noise induced by rivers: A new tool to monitor spatiotemporal changes in stream hydrodynamics. *Journal of Geophysical Research*, **113**, B05301.
- Bussat, S. & Kugler, S., 2009. Recording noise – Estimating shear-wave velocities: Feasibility of offshore ambient-noise surface-wave tomography (answt) on a reservoir scale. *SEG Expanded Abstracts*, **28**, 1627–1631.
- Campillo, M. & Paul, A., 2003. Long-Range Correlations in the Diffuse Seismic Coda. *Science*, **299**, 547–549.
- Campillo, M., 2006. Phase and Correlation in Random Seismic Fields and the Reconstruction of the Green Function. *Pure and Applied Geophysics*, **163**, 475–502.
- Capon, J., Greenfield, R. J. & Kolker, R. J., 1967. Multidimensional Maximum-Likelihood Processing of a Large Aperture Seismic Array. *Proceedings of the IEEE*, **55**, 192-211.
- Capon, J., 1969. High-Resolution Frequency-Wavenumber Spectrum Analysis. *Proceedings of the IEEE*, **57**, 1408–1418.
- Chen, Q.-F., Li, L., Li, G., Chen, L., Peng, W.-T., Tang, Y., Chen, Y. & Wang, F.-Y., 2004. Seismic features of vibration induced by train. *Acta Seismologica Sinica*, **17**, 715–724.
- Chen, Y.-J., Ju, S.-H., Ni, S.-H. & Shen, Y.-J., 2007. Prediction methodology for ground vibration induced by passing trains on bridge structures. *Journal of Sound and Vibration*, **302**, 806–820.
- Coward, D., Blair, D., Burman, R. & Zhao, C., 2003. Vehicle-induced seismic effects at a gravitational wave observatory. *Review of Scientific Instruments*, **74**, 4846–4854.
- Coward, D., Turner, J., Blair, D. & Galybin, K., 2005. Characterizing seismic noise in the 2-20 Hz band at a gravitational wave observatory. *Review of Scientific Instruments*, **76**, 44501-1–44501-5.
- Curtis, A., Gerstoft, P., Sato, H., Snieder, R. & Wapenaar, K., 2006. Seismic interferometry - turning noise into signal. *The Leading Edge*, **25**, 1082–1092.
- Duvall, T. L., Jefferies, S. M., Harvey, J. W. & Pomerantz, M. A., 1993. Time-distance helioseismology. *Nature*, **362**, 430-432.
- Endrun, B., Ohrnberger, M. & Savvaidis, A., 2010. On the repeatability and consistency of three-component ambient vibration array measurements. *Bulletin of Earthquake Engineering*, **8**, 535-570.
- Fäh, D., 1997. Microzonation of the city of Basel. *Journal of Seismology*, **1**, 87-102.
- Fäh, D., Stamm, G. & Havenith, H. B., 2008. Analysis of three-component ambient vibration measurements. *Geophys. J. Int.*, **172**, 199–213.

- Fiala, P., Degrande, G. & Augusztinovicz, F., 2007. Numerical modelling of ground-borne noise and vibration in buildings due to surface rail traffic. *Journal of Sound and Vibration*, **301**, 718–738.
- Forbriger, T., 2007. Low-frequency limit for H/V studies due to tilt. *Abstracts of the 67th Annual Meeting of the Deutsche Geophysikalische Gesellschaft, March 26-27, 2007, Aachen, Germany*.
- Gerstoft, P., Sabra, K. G., Roux, P., Kuperman, W. A. & Fehler, M. C., 2006. Green's function extraction and surface-wave tomography from microseisms in southern California. *Geophysics*, **71**, 23-31.
- Gerstoft, P., Shearer, P. M., Harmon, N. & Zhang, J., 2008. Global P, PP, and PKP wave microseisms observed from distant storms. *Geophysical Research Letters*, **35**, L23306.
- Giardini, D., 2009. Geothermal quake risks must be faced. *Nature*, **462**, 848-849.
- Gilles, P. M., Duvall, T. L., Scherrer, P. H. & Bogart, R. S., 1997. A subsurface flow of material from the Sun's equator to its poles. *Nature*, **390**, 52-54.
- Gouédard, P., Stehly, L., Brenguier, F., Campillo, M., Colin de Verdière, Y., Larose, E., Margerin, L., Roux, P., Sánchez-Sesma, F. J., Shapiro, N. M. & Weaver R. L., 2008a. Cross-correlation of random fields: mathematical approach and applications. *Geophysical Prospecting*, **56**, 375-393.
- Gouédard, P., Cornou, C. & Roux, P., 2008b. Phase-velocity dispersion curves and small-scale geophysics using noise correlation slantstack technique. *Geophys. J. Int.*, **172**, 971–981.
- Gouédard, P., Roux, P., Campillo, M. & Verdel, A., 2008c. Convergence of the two-point correlation function toward the Green's function in the context of a seismic-prospecting data set. *Geophysics*, **73**, 47-53.
- Gray, R. M., 1984. Vector Quantization. *IEEE ASSP Magazine*, 4-29.
- Grevemeyer, I., Herber, R. & Essen, H.-H., 2000. Microseismological evidence for a changing wave climate in the northeast Atlantic Ocean. *Nature*, **408**, 349-352.
- Groos, J. C. & Ritter, J. R. R., 2009. Time domain classification and quantification of seismic noise in an urban environment. *Geophysical Journal International*, **179**, 1213–1231.
- Groos, J. & Ritter, J. R. R., 2010. Seismic noise: A challenge and opportunity for seismological monitoring in densely populated areas. In: Ritter, J. & Oth, A. (Eds.), *Induced Seismicity, Cahiers du Centre Européen de Géodynamique et de Séismologie*, **30**, 87-97.
- Groos, J. C., Bussat, S. & Ritter, J. R. R., 2010. Effects of different processing schemes and automated data selection on seismic noise cross-correlation functions. Submitted to *Geophysical Journal International*.
- Gutenberg, B., 1911. Die seismische Bodenunruhe. *Doctoral thesis*, Universität Göttingen, Germany.
- Gutenberg, B., 1924. Die seismische Bodenunruhe (Schwingungen des Erdbodens durch Industrie, Verkehr, meteorologische u. a. Ursachen) und ihr Zusammenhang mit den Nachbargebieten insbesondere Geologie und Meteorologie. *Sammlung geophysikalischer Schriften*, **3**.
- Hanssen, P. & Bussat, S., 2008. Pitfalls in the analysis of low frequency passive seismic data. *First Break*, **26**, June, 111-119.

- Hao, H. & Ang, T. C., 1998. Analytical modeling of traffic-induced ground vibrations. *Journal of Engineering Mechanics*, **124**, 921–928.
- Hyvärinen, A., Karhunen, J. & Oja, E., 2001. *Independent Component Analysis*, John Wiley & Sons, New York.
- Kar, C. & Mohanty, A. R., 2006. Monitoring gear vibrations through motor current signature analysis and wavelet transform. *Mechanical Systems and Signal Processing*, **20**, 158-187.
- Karlström, A., 2006. An analytical model for ground vibrations from accelerating trains. *Journal of Sound and Vibration*, **293**, 587–598.
- Kennett, B. L. N. & Engdahl, E. R., 1991. Traveltimes for global earthquake location and phase identification. *Geophys. J. Int.*, **105**, 429–465.
- Kohonen, T., 2001. *Self-Organizing maps*, Springer Series in Information Sciences, **30**, Third Extended Edition, Springer Berlin, Heidelberg, New York, 1995, 1996, 2001.
- Koper, K. D., Seats, K. & Benz, H., 2010. On the composition of Earth's short-period seismic noise field. *Bulletin of the Seismological Society of America*, **100**, 606-617.
- Köhler, A., Ohrnberger, M., Scherbaum, F., Wathelet, M. & Cornou, C., 2007. Assessing the reliability of the modified three-component spatial autocorrelation technique. *Geophysical Journal International*, **168**, 779-796.
- Köhler, A., 2009. *Recognition and investigation of temporal patterns in seismic wavefields using unsupervised learning techniques*. Dissertation, University of Potsdam, Germany.
- Köhler, A., Ohrnberger, M. & Scherbaum, F., 2010. Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organising Maps. *Geophysical Journal International*, **182**, 1619-1630.
- Kurrle, D. & Widmer-Schmidrig, R., 2008. The horizontal hum of the Earth: A global background of spheroidal and toroidal modes. *Geophysical Research Letters*, **35**, L06304.
- Lachet, C. & Bard, P.-Y., 1994. Numerical and theoretical investigations on the possibilities and limitations of Nakamura's technique. *Journal of Physics of the Earth*, **42**, 377-397.
- Lacoss, R. T., Kelly, E. J. & Toksöz, M. N., 1969. Estimation of seismic noise structure using arrays. *Geophysics*, **34**, 21-38.
- Li, H., Bernardi, F. & Michelini, A., 2010. Surface wave dispersion measurements from ambient seismic noise analysis in Italy. *Geophys. J. Int.*, **180**, 1242–1252.
- Lin, F.-C., Moschetti, M.P. & Ritzwoller, M. H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps. *Geophysical Journal International*, **173**, 281-298.
- Longuet-Higgins, M. S., 1950. A theory of the origin of microseism. *Phil. Trans. Roy. Soc. London, Ser. A*, **243**, 1-35.
- Mandrescu N., Radulian M. & Marmureanu G., 2004. Site conditions and predominant period on seismic motion in the Bucharest urban area. *Revue Roumaine de Geophysique*, **48**, 37-48.
- Marsaglia, G. & Tsang, W. W., 2000. The Ziggurat Method for Generating Random Variables. *Journal of Statistical Software*, **5**, 1–7.

- Meier, U., Shapiro, N. M. & Brenguier, F., 2010. Detecting seasonal variations in seismic velocities within Los Angeles basin from correlations of ambient seismic noise. *Geophys. J. Int.*, **181**, 985-996.
- Milana, G., Barba, S., Del Pezzo, E. & Zambonelli, E., 1996. Site Response from Ambient Noise Measurements: New Perspectives from an Array Study in Central Italy. *Bull. Seism. Soc. Am.*, **86**, 320–328.
- Mohammed, A. Y., Berteussen, K. A., Small, J. & Barkat, B., 2007. A low frequency, passive seismic experiment over a carbonate reservoir in Abu Dhabi. *First Break*, **25**, November, 71-73.
- Montgomery, M. R., 2008. The Urban Transformation of the Developing World. *Science*, **319**, 761–764.
- Nakamura, Y., 1989. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quarterly Report Railway Technology Research Institute*, **30**, 25-30.
- Nakamura, Y., 1996. Real-time information systems for hazards mitigation. *Proceedings of the 11th World Conference on Earthquake Engineering*, Acapulco, Mexico.
- Nakamura, Y., 2000. Clear identification of fundamental idea of Nakamura's technique and its applications. *Proceedings of the 12th World Conference on Earthquake Engineering*, Auckland, New Zealand.
- Nogoshi, M. & Igarashi, T., 1971. On the amplitude characteristics of microtremor (part 2). *Journal of Seismological Society of Japan*, **24**, 26-40. (In Japanese).
- Oncescu, M. C., Marza, V. I., Rizescu, M. & Popa, M., 1999. The Romanian Earthquake Catalogue between 984-1997 in Wenzel, F., Lungu, D. & Novak, O., 1999. *Vrancea earthquakes tectonics, hazard and risk mitigation*, Kluwer Academic Publishers, Dordrecht, Boston.
- 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.
- Percival, D. B. & Walden, A. T., 1998. *Spectral analysis for physical applications: Multitaper and conventional univariate techniques*, Reprinted (with corr.), Cambridge University Press, Cambridge.
- Plesinger, A. & Wielandt, E., 1974. Seismic noise at 2 Hz in Europe. *Journal of Geophysics*, **40**, 131-136.
- Rebeur-Paschwitz, E. v., 1889. The earthquake of Tokio, April 18, 1889. *Nature*, **40**, 294-295.
- Ritter, J. R. R., Balan, S. F., Bonjer, K.-P., Diehl, T., Forbriger, T., Marmureanu, G., Wenzel, F. & Wirth, W., 2005. Broadband Urban Seismology in the Bucharest Metropolitan Area. *Seismological Research Letters*, **76**, 574–580.
- Ritter, J. R. R. & Groos, J., 2007. Kyrills seismischer Fingerabdruck – Der Orkan Kyrill ließ auch den Boden erzittern. *Spektrum der Wissenschaft*, **March**, 19.
- Ritter, J. R. R., Thun, J., Groos, J., Wenzel, F. & Hauck, C., 2009. METSEIS: An experiment to quantify meteorological and seismological interactions. In: Sens-Schönfelder, C., Ritter, J., Wegler, U. & Große, C. (eds.), 2009. Noise and diffuse

wavefields, *Mitteilungen der Deutschen Geophysikalischen Gesellschaft*, Sonderband II/2009, 101-106.

Roux, P., Sabra, K. G., Gerstoft, P., Kuperman, W. A. & Fehler, M. C., 2005. P-waves from cross-correlation of seismic noise. *Geophysical Research Letters*, **32**, L19303.

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.

Saito, T., 2010. Love-wave excitation due to the interaction between a propagating ocean wave and the sea-bottom topography. *Geophysical Journal International*, doi: 10.1111/j.1365-246X.2010.04695.x.

Scales, J. A. & Snieder, R., 1998. What is noise? *Geophysics*, **63**, 1122-1124.

Scherbaum, F. & Bouin, M.-P., 1997. FIR Filter Effects and Nucleation Phases. *Geophysical Journal International*, **130**, 661-668.

Scherbaum, F., 2001. *Of Poles and Zeros, Fundamentals of Digital Seismology 2nd Edition*. Kluwer Academic Publishers, Dordrecht.

Scherbaum, F., Hinzen, K. G. & Ohrnberger, M., 2003. Determination of shallow shear wave velocity profiles in the Cologne, Germany area using ambient vibrations. *Geophys. J. Int.*, **152**, 597–612.

Sèbe, O., Forbriger, T. & Ritter J. R. R., 2009. The shear wave velocity underneath Bucharest city, Romania, from the analysis of Love waves. *Geophys. J. Int.*, **176**, 965–979.

SEED Reference Manual, 2010. *SEED Reference Manual, Standard for the Exchange of Earthquake Data, SEED Format Version 2.4, May 2010*. Incorporated Research Institutions for Seismology, Washington, DC.

Sens-Schönfelder, C. & Wegler, U., 2006. Passive image interferometry and seasonal variations of seismic velocities at Merapi volcano, Indonesia. *Geophysical Research Letters*, **33**, L21302.

Sens-Schönfelder, C., 2008. Synchronizing seismic networks with ambient noise. *Geophysical Journal International*, **174**, 966-970.

Shapiro, N. M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, **31**, L07614.

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.

Sleman, R., van Wettum, A. & Trampert, J., 2006. Three-Channel correlation analysis: A new technique to measure instrumental noise of digitizers and seismic sensors. *Bulletin of the Seismological Society of America*, **96**, 258-271.

Smith, W. H. F. & Wessel, P., 1990. Gridding with continuous curvature splines in tension. *Geophysics*, **55**, 293-305.

- Stehly, L., Campillo, M. & Shapiro, N. M., 2006. A study of the seismic noise from its long-range correlation properties. *Journal of Geophysical Research*, **111**, B10306.
- Stehly, L., Campillo, M. & Shapiro, N. M., 2007. Traveltime measurements from noise correlation: stability and detection of instrumental time-shifts. *Geophysical Journal International*, **171**, 223-230.
- Sudhaus, H. & Ritter, J., 2009. Broadband frequency-dependent amplification of seismic waves across Bucharest, Romania. *Journal of Seismology*, DOI 10.1007/s10950-008-9140-0.
- Tarvainen, M., 1999. Recognizing explosion sites with a self-organizing network for unsupervised learning. *Phys. Earth planet. Int.*, **113**, 143-154.
- Tsai, V. C. & Moschetti, M. P., 2010. An explicit relationship between time-domain noise correlation and spatial autocorrelation (SPAC) results. *Geophysical Journal International*, **182**, 454-460.
- United Nations. *World Urbanization Prospects*. New York, 2006.
- van der Baan, M., 2009. The origin of SH-wave resonance frequencies in sedimentary layers. *Geophysical Journal International*, **178**, 1587-1596.
- Vesanto, J. & Alhoniemi, E., 2000. Clustering of the self-organizing map. *IEEE Trans. Neural Network*, **11**, 586-600.
- Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., Team, S. & Oy, L., 2000. Som toolbox for matlab. *Techn. Ber. Helsinki University of Technology*.
- Walker, D., 2008. Recent developments in low frequency spectral analysis of passive seismic data. *First Break*, **26**, February, 69-77.
- Ward, H. & Crawford, R., 1966. Wind-induced vibrations and building modes. *Bull. Seism. Soc. Am.*, **56**, 793-813.
- Wapenaar, K. & Fokkema, J., 2006. Green's function representations for seismic interferometry. *Geophysics*, **71**, 33-46.
- Wapenaar, K., Slob, E. & Snieder, R., 2006. Unified Green's function retrieval by cross-correlation. *Physical Review Letters*, **97**, 234301.
- Weaver, R. L. & Lobkis, O. I., 2001. Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies. *Physical Review Letters*, **87**, 134301.
- Weaver, R. L., 2005. Information from Seismic Noise. *Science*, **307**, 1568-1569.
- Wegler, U. & Sens-Schönfelder, C., 2007. Fault zone monitoring with passive image interferometry. *Geophysical Journal International*, **168**, 1029-1033.
- Wenzel, F., Lungu, D. & Novak, O., 1999. *Vrancea earthquakes tectonics, hazard and risk mitigation*, Kluwer Academic Publishers, Dordrecht, Boston.
- Wessel, P. & Smith, W. H. F., 1998. New, improved version of Generic Mapping Tools released, *EOS*, **79**, 579.
- Withers, M. M., Aster, R. C., Young, C. J. & Chael, E. P., 1996. High-frequency analysis of seismic background noise as a function of wind speed and shallow depth. *Bull. Seism. Soc. Am.*, **86**, 1507-1515.

Yang, Y. & Ritzwoller, M. H., 2008. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet. *Geochem. Geophys. Geosyst.*, **9**, Q02008.

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.

Zhang, J., Gerstoft, P. & Shearer, P. M., 2009. High-frequency P-wave seismic noise driven by ocean winds. *Geophysical Research Letters*, **36**, L09302.

Zhang, J., Gerstoft, P. & Shearer, P. M., 2010. Resolving P-wave travel-time anomalies using seismic array observations of oceanic storms. *Earth and Planetary Science Letters*, **292**, 419-427.

Ziehm, J., 2006. Seismisches Monitoring einer Großstadt, Analyse von H/V-Spektralverhältnissen, *Diploma thesis*, Geophysical Institute, Universität Karlsruhe (TH), Germany, 89pp.

Zürn, W., Exß, J., Steffen, H., Kroner, C., Jahr, T. & Westerhaus, M., 2007. On reduction of long-period horizontal seismic noise using local barometric pressure. *Geophysical Journal International*, **171**, 780-796.

Used hard- and software

This thesis was written on a personal computer with the operating system *Microsoft Windows Vista* using the office processing software *Microsoft Office 2007*. Some schematic figures were constructed with *Microsoft Power Point 2007*. The format conversion of the figures was done with *The GNU Image Manipulation Program – GIMP* on a personal computer with the Linux operating system openSUSE 11.

The data processing was done on personal computers with the Linux operating systems openSUSE 10 and openSUSE 11. The data processing was done with *MATLAB for Linux* (The Mathworks) using the new written *Karlsruhe Processing (KaSP) Toolbox for MATLAB* which is described in Appendix A. The *KaSP-Toolbox for MATLAB* is available as free software under the GNU General Public License.

The analysis of the noise classification results with Self-Organizing Maps (SOM) was done with the *SOM Toolbox 2.0 for MATLAB* written by Esa Alhoniemi, Johan Himberg, Jukka Parviainen and Juha Vesanto. The *SOM Toolbox 2.0* is available as free software under the GNU General Public License.

Most figures were created with *MATLAB*. The maps of the Bucharest area were created with the *Generic Mapping Tools - GMT* by Wessel & Smith (1998).

The FK-analyses were done with *SeismicHandler* by Klaus Stammler.

The facilities of the IRIS Data Management System, and specifically the IRIS Data Management Center, were used for access to waveforms and metadata required in this study (see section 3.2). The software *rdseed* of the IRIS DMC was used to obtain the waveforms and metadata from the provided SEED volumes. The software *evalresp* of the IRIS DMC was used to cross-check the transfer functions calculated with my MATLAB-functions in the KaSP-Toolbox as discussed in section 3.3.

The software package *xcut* written and maintained by Rainer Plokarz at the Geophysical Institute of the Karlsruhe Institute of Technology was used to access the waveforms of the Karlsruhe BroadBand Array KABBA (see section 3.1).

Danksagung / Acknowledgment

PD Dr. Joachim Ritter danke ich für die Betreuung meiner Promotion und die Übernahme des Hauptreferats. Ohne seine große Unterstützung beim Schreiben zahlreicher Anträge und der Durchführung eines Drittmittel-Projektes mit StatoilHydro wäre diese Arbeit schlicht nicht möglich gewesen. Weiterhin bedanke ich mich sehr für das große Vertrauen, die Unterstützung und die vielen Freiheiten, die ich bei der Durchführung meiner Promotion hatte. Besonders wertvoll für mich waren die wöchentlichen Arbeitsgruppentreffen, die mich immer zu neuen Erkenntnissen und Ideen geführt haben. Danken möchte ich auch für die zahlreichen Konferenzteilnahmen und für deren Vorbereitung in Form von umfassend diskutierten Probevorträgen.

Prof. Dr. Friedemann Wenzel danke ich für die Übernahme des Korreferats und für die zahlreichen wertvollen Anregungen zu meiner Arbeit. Besonders bedanken möchte ich mich auch für das Erstellen mehrerer Gutachten und für die Möglichkeit, eine *Feasibility Study of Young Scientists* am Geophysikalischen Institut durchführen zu können.

Thomas Forbriger danke ich für die ungezählten Diskussionen in den letzten Jahren zu allen Aspekten der Seismologie. Sie haben immer mein Verständnis vertieft und viele Ideen angeregt. Auch Britta Wawerzinek und Lisa Rehor danke ich für die Diskussionen zur Lösung allerhand fachlicher Probleme.

Sascha Bussat und Peter Hanssen von StatoilHydro (heute Statoil) danke ich für die Durchführung eines gemeinsamen Forschungsprojektes in 2008 und den bleibenden Kontakt darüber hinaus. Dieses Projekt war der Anstoß für meine Arbeiten zur Berechnung von Kreuzkorrelationsfunktionen des seismischen Rauschens, die heute einen großen Teil dieser Dissertation ausmachen. Besonders bedanken möchte ich mich für die hervorragende Betreuung und die vielen Diskussionen während meines produktiven und lehrreichen Aufenthalts am StatoilHydro Research Centre in Bergen.

Bedanken möchte ich mich bei allen Kolleginnen und Kollegen am Institut, besonders bei Petra Knopf, Rainer Plokarz, Werner Scherer, Gaby Bartman und Claudia Payne, an die ich mich immer Hilfe suchend wenden konnte.

Meiner Zimmerkollegin Britta möchte ich für die vielen Jahre der tollen und entspannten Zusammenarbeit danken.

Mein letztes großes Dankeschön geht an die vielen Menschen, die auch außerhalb des Instituts diese Arbeit auf vielfältigste Weise unterstützt haben. Meinen Eltern danke ich für die immer währende Hilfsbereitschaft in allen Lebenslagen. Ein großer Dank geht auch an alle Korrekturleser und Korrekturleserinnen dieser Arbeit!

Acknowledgment

The author was supported by a graduate scholarship (Landesgraduiertenförderung) of the state of Baden-Württemberg at the Karlsruhe Institute of Technology.

This work was supported by grants of the Deutsche Forschungsgemeinschaft (CRC-461), Universität Karlsruhe (TH), Germany, and the National Institute for Earth Physics (NIEP), Bucharest-Magurele, Romania. The URS project was a collaboration between the Universität Karlsruhe (TH) and the NIEP. Data were acquired with the **K**arlsruhe **B**road**B**and **A**rray, KABBA.

The facilities of the IRIS Data Management System, and specifically the IRIS Data Management Center, were used for access to waveforms and metadata required in this

study. The IRIS DMS is funded through the National Science Foundation and specifically the GEO Directorate through the Instrumentation and Facilities Program of the National Science Foundation under Cooperative Agreement EAR-0552316. The Global Seismographic Network (GSN) is a cooperative scientific facility operated jointly by the Incorporated Research Institutions for Seismology (IRIS), the United States Geological Survey (USGS), and the National Science Foundation (NSF).

Curriculum Vitae

Name: Jörn Christoffer Groos

Born: 11.05.1981 in Mannheim, Germany

School

08/1987-07/1991 Grundschule Malsch (elementary school)

08/1991-06/2000 Leibniz Gymnasium Östringen (secondary school)
Final degree: Abitur

Civil Service

09/2000-08/2001 Kreisaltenpflegeheim Nussloch

Academic studies

10/2001-04/2007 Studies of geophysics, Universität Karlsruhe (TH)
Final degree: Diplom-Geophysiker

05/2007-12/2010 PhD studies, Karlsruhe Institute of Technology