Applications of ambient seismic noise: clock error detection and group velocity estimation in land and ocean bottom seismograms

Sarah Katharina Hable

SIGILL·UNIVERSITAT·LUDOVIC·MAXIM

München 2019

APPLICATIONS OF AMBIENT SEISMIC NOISE: CLOCK ERROR DETECTION AND GROUP VELOCITY ESTIMATION IN LAND AND OCEAN BOTTOM SEISMOGRAMS

Sarah Katharina Hable

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Geowissenschaften
der Ludwig-Maximilians-Universität München

vorgelegt von
Sarah Katharina Hable

MÜNCHEN, DEN 06.03.2019

Erstgutachter: Prof. Dr. Heiner Igel
Zweitgutachter: Prof. Dr. Karin Sigloch

Tag der mündlichen Prüfung: 06.06.2019

“Our definition of signal and noise is subjective in the sense that a given part of the data is “signal” for those who know how to analyze and interpret the data, but it is “noise” for those who do not.”

Aki & Richards (1980)

SUMMARY

Ambient seismic noise generated by ocean waves is continuously present in seismograms and has previously been considered as undesired, disturbing signal. However, it has been shown that cross-correlations of noise recorded at two seismic stations converge towards the Green's function. This function describes the underground properties between the two stations. Ever since this finding, a wide range of noise applications has been described, which are still under development. Temporal changes of noise cross-correlations can be used for detection of clock errors in seismic data, while group velocities derived from cross-correlations are the basis for tomographic studies.

In the first part of this thesis, an extensive clock error study of land stations and ocean bottom seismometers (OBSs) is presented. A new multiple-component approach is applied, which enhances the accuracy ($\sim$20 ms) of the detected clock errors significantly. Moreover, this approach allows the retrieval of clock errors with high temporal resolution of 1-2 days, even for large interstation distances ($\sim$300 km). The application of the described approach to data sets with low timing quality could highly increase their usability for structural studies.

The second part of this thesis deals with group velocity curves that are mainly retrieved from OBS cross-correlations with interstation distances of up to $\sim$2000 km. A joint inversion of the noise group velocities together with earthquake data from a prior study yields a high-resolution crustal S-wave velocity model of the western Indian Ocean. This model highly reflects tectonic structures in this region, like ocean ridges and plateaus. These results demonstrate the feasibility of large-scale OBS noise tomography of ocean basins, while prior studies were limited to smaller scales.

In the last part of this thesis, first steps towards a seismological image of the island La Réunion are presented. Group velocity curves are derived from noise cross-correlations between island stations. In agreement with gravity studies, the group velocities indicate a spacious high-velocity body beneath the ancient volcano of the island. The inversion of the group velocities will yield a tomographic model of the island, which may be used as starting velocity model for future seismic surveys.

The majority of the seismic stations used in this thesis were installed in the western Indian Ocean on and around La Réunion during the RHUM-RUM project. This reflects the high applicability of noise cross-correlations (from data quality inspection to crustal imaging) within the same data set. The detailed method descriptions of this work provide a valuable guideline for future studies, that deal with land or OBS seismograms.

ZUSAMMENFASSUNG

Seismisches Hintergrundrauschen, das von Ozeanwellen erzeugt wird, ist kontinuierlich präsent in Seismogrammen und wurde früher als unerwünschtes Störsignal betrachtet. Allerdings wurde erwiesen, dass sich Kreuzkorrelationen von Rauschaufzeichnungen zweier seismischer Stationen der Greenschen Funktion annähern. Diese Funktion beschreibt die Eigenschaften des Untergrunds zwischen den beiden Stationen. Seit dieser Erkenntnis wurde eine große Bandbreite an Anwendungen mit Rauschen beschrieben, die nach wie vor weiterentwickelt werden. Zeitliche Veränderungen von Rausch-Kreuzkorrelationen können dazu genutzt werden, um Zeitfehler in den seismischen Daten festzustellen, während Gruppengeschwindigkeiten, die von Kreuzkorrelationen abgeleitet werden, als Grundlage für tomographische Studien dienen.

Im ersten Teil dieser Doktorarbeit wird eine umfangreiche Studie bezüglich Zeitfehler in Landstationen und Ozeanbodenseismometern (OBS) vorgestellt. Ein neuer Ansatz, der auf mehreren Komponenten beruht und der die Genauigkeit (∼20 ms) der aufgespürten Zeitfehler deutlich verbessert, wird angewandt. Darüber hinaus erlaubt dieser Ansatz Zeitfehler mit hoher zeitlicher Auflösung von 1-2 Tagen zu gewinnen, sogar für lange Distanzen (∼300 km) zwischen den Stationen. Der Einsatz des beschriebenen Ansatzes bei Datensätzen mit geringer zeitlicher Qualität könnte deren Verwendbarkeit für strukturelle Studien erheblich erhöhen.

Der zweite Teil dieser Arbeit behandelt Gruppengeschwindigkeitskurven, die hauptsächlich von OBS-Kreuzkorrelationen mit Distanzen bis zu ∼2000 km stammen. Eine gemeinsame Inversion der Rausch-Gruppengeschwindigkeiten zusammen mit Erdbebendaten einer früheren Studie ergeben ein hochauflösendes S-Wellengeschwindigkeitsmodell der Kruste im westlichen Indischen Ozean. Dieses Modell spiegelt die tektonischen Strukturen in dieser Region, wie Ozeanrücken und Plateaus, deutlich wider. Diese Ergebnisse beweisen, dass Rausch-Tomographien von Ozeanbecken mit OBS im großen Maßstab möglich sind, während frühere Studien auf kleinere Maßstäbe begrenzt waren.

Im letzten Teil dieser Arbeit werden erste Schritte für ein seismologisches Bild der Insel La Réunion vorgestellt. Gruppengeschwindigkeitskurven werden von Rausch-Kreuzkorrelationen zwischen Inselstationen gewonnen. Die Gruppengeschwindikeiten weisen in Übereinstimmung mit Gravitationsstudien auf einen ausgedehnten Körper mit hoher Geschwindigkeit unter dem ehemaligen Vulkan der Insel hin. Die Inversion der Gruppengeschwindigkeiten wird ein tomographisches Modell der Insel ergeben, das als Geschwindigkeits-Startmodell für zukünftige seismische Studien dienen könnte.

Der Großteil der seismischen Stationen, die in dieser Doktorarbeit verwendet werden, wurden im westlichen Indischen Ozean auf und rund um La Réunion während des RHUM-RUM Projektes installiert. Dies spiegelt die vielen Verwendungsmöglichkeiten von Rausch-Kreuzkorrelationen (von der Überprüfung der Datenqualität bis zu Krustenabbildungen) innerhalb desselben Datensatzes. Die detaillierten Methodenerläuterungen dieser Arbeit liefern eine wertvolle Orientierungshilfe für zukünftige Studien, die sich mit Land- oder OBS-Seismogrammen beschäftigen.

Contents

List of Figures

List of Tables

Abbreviations

CC	**C**orrelation **C**oefficient
CCF	**C**ross-**C**orrelation **F**unction
CIR	**C**entral **I**ndian **R**idge
GPS	**G**lobal **P**ositioning **S**ystem
NHNM	**N**ew **H**igh **N**oise **M**odel
NLNM	**N**ew **L**ow **N**oise **M**odel
OBS	**O**cean **B**ottom **S**eismometer
OVPF	**O**bservatoire **V**olcanologique du **P**iton de la **F**ournaise
PCC	**P**hase **C**ross-**C**orrelation
PPSD	**P**robabilistic **P**ower **S**pectral **D**ensity
PSD	**P**ower **S**pectral **D**ensity
RCF	**R**eference **C**orrelation **F**unction
RHUM-RUM	**R**éunion **H**otspot and **U**pper **M**antle - **R**éunions **U**nterer **M**antel
SNR	**S**ignal-to-**N**oise **R**atio
SWIR	**S**outh**W**est **I**ndian **R**idge
tf-PWS	**t**ime-**f**requency domain **P**hase-**W**eighted **S**tacking

Chapter 1

Introduction

1.1 Ambient seismic noise

The first seismographs that could monitor ground motions continuously as a function of time have been invented end of the 19th century (Dewey & Byerly, 1969). While earthquakes are identified by temporary, large amplitude deflections from a baseline, smaller deflections are permanently present in seismograms. These background deflections are considered as noise. The main sources of the noise are essentially located near the earth's surface, as seismic noise is mainly composed of surface waves (Rayleigh and Love waves) (e.g. Friedrich et al., 1998; Ekström, 2001). The manifold noise sources comprise tides, strong wind, ocean waves, human activity, industry, cars and trains, self-noise of the instrument, etc. (Bonnefoy-Claudet et al., 2006).

Bonnefoy-Claudet et al. (2006) gave a rough classification of noise sources dependent on the frequency: low frequency noise (<1 Hz) originates mainly from natural sources like ocean swell, high frequency noise (>5 Hz) is typically produced by cultural sources like traffic and intermediate frequency noise (∼1-5 Hz) is generated by both natural (mainly meteorological activity) and cultural sources.

This study deals with ambient seismic noise of periods between 1 s and 50 s. In this period range, two prominent peaks can be observed in the seismic noise spectrum: the primary microseism that peaks around 14 s and the stronger secondary microseism that peaks roughly at 7 s (e.g. Gutenberg, 1951). Fig. 1.1 shows the power spectral density of an hour-long seismic record. The typical period bands of the secondary microseism (3-10 s) and primary microseism (10-20 s) are marked by blue and orange colour, respectively. In both period bands, high amplitude peaks are visible that can be associated with the secondary and primary microseismic peak.

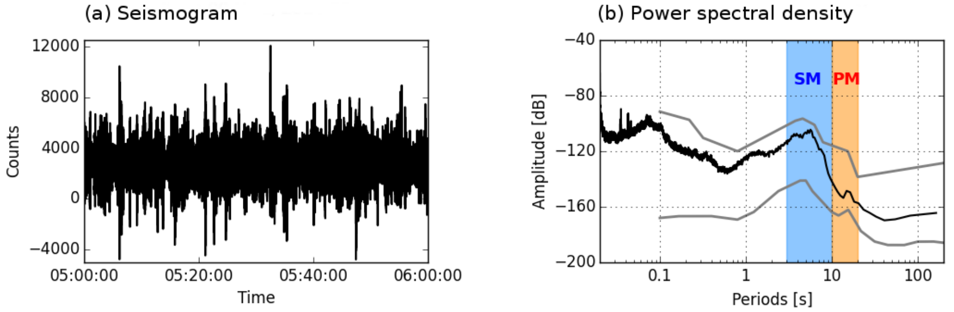


Figure 1.1: (a) Seismogram of one hour (2014-02-09, 05:00am to 06:00am) of station SALA (vertical component) and (b) corresponding power spectral density. The blue- and orange-shaded areas mark the period bands that can be associated with the secondary microseism (SM, 3-10 s) and the primary microseism (PM, 10-20 s). The grey lines in (b) indicate the New High Noise Model and the New Low Noise Model introduced by Peterson et al. (1993) as typical upper and lower noise level bounds of terrestrial stations.

The origin of these microseismic peaks has been investigated since decades (e.g. Gutenberg, 1911; Longuet-Higgins, 1950; Gutenberg, 1951; Hasselmann, 1963). The primary microseism is characterized by the same period as the ocean swell (typically 10-20 s). This argues for a direct interaction of ocean waves with the coast as source mechanism. The interaction induces pressure variations on the sloping seafloor in shallow water, which excites seismic waves (Hasselmann, 1963). This linear coupling of ocean wave energy into seismic energy is supported by the strong correlation of noise levels with ocean wave heights at coastal regions (e.g. Barruol et al., 2006; Juretzek & Hadziioannou, 2017).

The secondary microseismic peak is characterized by half of the period of the ocean swell. Thus, the secondary microseism is proposed to be generated by non-linear interaction of two ocean waves that propagate with the same period, but in opposite directions. The interference of the waves generates then a depth-independent pressure fluctuation that is transmitted through the water column and induces seismic waves at the ocean bottom with a frequency that is twice to the frequency of the generating ocean waves (e.g. Longuet-Higgins, 1950; Hasselmann, 1963).

Based on numerical modelling, Ardhuin et al. (2011) described three classes of how counter-propagating ocean waves can potentially be produced in nature. In class I, ocean waves with opposite direction are generated by a single storm with a wide directional spectrum, like a cyclone. In class II, the secondary microseism is generated near the shore by interference of ocean waves with their own coastal reflections. Class III relates the generation of the secondary microseism to the interaction of waves from two counter-propagating storms with similar wavelengths. According to these classes, noise sources of the secondary microseism have been located both in shallow water related to coastal reflections of class II (e.g. Bromirski & Duennebier, 2002; Bromirski et al., 2013) and in deep water related to storms of class I and III (e.g. Obrebski et al., 2012; Davy et al., 2015).

The described theories of microseismic noise sources are commonly accepted as main excitation mechanisms of the primary and secondary microseisms. These mechanisms are primarily responsible for the generation of Rayleigh waves, as they are equivalent to vertical forces that act on the ocean bottom (Tanimoto et al., 2015). However, recent studies showed that Love waves are also present in the microseisms (e.g. Friedrich et al., 1998; Nishida et al., 2008) and that their contribution to the microseismic noise field can even exceed the contribution of Rayleigh waves (e.g. Tanimoto et al., 2015; Juretzek & Hadziioannou, 2016). Thus, the exact excitation mechanism of Love waves in the microseisms is still under investigation (e.g. Juretzek & Hadziioannou, 2017; Ziane & Hadziioannou, 2019).

1.2 State of the art: noise as source of information

For long time, ambient seismic noise has mainly been considered as signal-disturbing part of a seismogram, especially when the signal is so tiny that it gets overprinted by background noise. With increasing understanding of the nature of the noise, applications that use noise as source of information emerged. Since the 1950s, array techniques based on noise like spatial auto-correlation analysis (e.g. Aki, 1957, 1965) and frequency-wavenumber ($f-k$) analysis (Capon et al., 1967; Lacoss et al., 1969) have been developed to derive surface wave dispersion curves for S-wave velocity profiles. In the 1970s, the usage of noise for seismic city microzonation was detected (Nogoshi & Igarashi, 1971; Nakamura, 1989). These techniques have been further developed and are still commonly used in seismic studies (e.g. Nakahara, 2012; Gal et al., 2014; Wassermann et al., 2016).

1.2.1 Cross-correlation functions

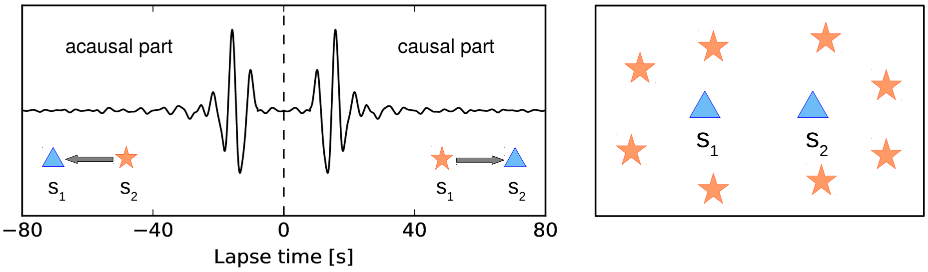
An amazing success story of ambient seismic noise began in the early 2000s, when the cross-correlation technique was adopted to seismology and opened a wide range of applications. Lobkis & Weaver (2001) developed the basic principle of noise cross-correlations in ultrasound experiments, while Shapiro & Campillo (2004) first showed that cross-correlations of ambient seismic noise converge towards the Green's function between two seismometers. The Green's function corresponds to the seismogram that can be recorded at one station as response to a delta pulse acting as input at another station. Thus, the Green's function reflects the properties of the medium along the propagation path between the two stations.

In practice, seismograms of two seismic stations are cross-correlated yielding a two-sided cross-correlation function (CCF). To reduce the effect of high-amplitude signals like earthquakes, either a special pre-processing of the seismograms before the cross-correlation is required (Bensen et al., 2007) or a phase cross-correlation is applied which is based on amplitude-insensitive phase alignments in contrast to conventional cross-correlation (Schimmel, 1999; Schimmel et al., 2011). Fig. 1.2(a) illustrates a schematic sketch of a CCF between two stations s_1 and s_2 (blue triangles), where both sides are symmetric: the causal side with positive lapse times represents the seismogram measured at station s_2 as response to a delta pulse that has acted as virtual source at the location of station s_1; the acausal side with negative lapse times shows the impulse response recorded at station s_1 as output to a delta pulse at station s_2. The symmetry of both sides can theoretically be expected as the waves sample the same medium, but it can only be achieved when the noise sources (orange stars in Fig. 1.2) are evenly distributed and the same wave energy is travelling from station s_1 to station s_2 as vice versa. In nature, the spatial distribution of the noise sources is generally uneven, which causes asymmetric CCFs as depicted in Fig. 1.2(b).

The convergence of a CCF towards the Green's function is limited by the distance between both involved stations: the larger the interstation distance, the lower the signal-to-noise ratio of the CCF and the poorer the convergence towards the Green's function. The convergence can be enhanced by stacking CCFs of shorter time increments (e.g. 1 day) over a longer time duration (e.g. several months or years) (Bensen et al., 2007). This procedure is used for structural studies that image the earth's subsurface and requires CCFs that remain stable over time. However, temporal changes in the CCFs are measurable in certain situations. Stehly et al. (2007) described three mechanisms that can cause temporal fluctuations in the CCFs: velocity variations in the propagation medium, location changes of the noise sources, and clock errors in the seismic data.

Velocity changes in the propagation medium appear as dilated or compressed CCFs. Such changes can be measured in the vicinity of faults after large earthquakes, e.g. due to strong ground shaking (e.g. Brenguier et al., 2008b; Wegler et al., 2009; Hobiger et al., 2012). Likewise, magma movements at volcanoes can be associated with velocity changes (e.g. Brenguier et al., 2008a; Sens-Schönfelder et al., 2014). Furthermore, Salvermoser et al. (2015) showed that velocity changes measured in a concrete bridge can be related to temperature variations.

(a) Symmetric CCF



(b) Asymmetric CCF

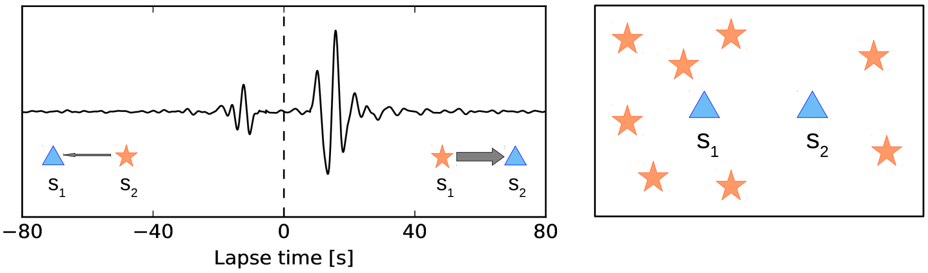


Figure 1.2: Schematic sketch of a (a) symmetric and (b) asymmetric cross-correlation function (CCF) between stations s_1 and s_2 (blue triangles). In (a), the causal and the acausal part of the CCF are symmetric, since the noise sources (orange stars) are evenly spaced around the seismic stations. In (b), the causal side is characterized by a much higher amplitude than the acausal side, as much more wave energy is propagating from station s_1 to station s_2 due to the concentration of noise sources around station s_1.

Spatial changes of the noise sources would affect the waveforms of the causal and acausal CCF's side independently of each other (Stehly et al., 2007). This effect may be too small to influence structural studies significantly.

Clock errors in seismic records cause a shift in the CCFs, i.e. waves appear earlier on the causal side and later on the acausal side or vice versa. A summary of clock error studies is given in Section 1.2.2.

1.2.2 Clock error detection

Only a small number of studies investigated the correction of clock errors in seismic data (e.g. Sabra et al., 2005a; Stehly et al., 2007; Sens-Schönfelder, 2008; Gouédard et al., 2014; Hannemann et al., 2014; Le et al., 2018), although an accurate timing is a crucial requirement for many seismological applications like tomographic studies. In seismic land stations, a frequent synchronization of the internal station clock with a GPS signal usually ensures a proper timing, but can fail due to GPS signal loss. Ocean bottom seismometers (OBSs), that are deployed on the ocean bottom, can only be synchronized with a GPS signal on ship before deployment and after recovery. For the period during the OBS deployment, an unverified linear clock drift of the internal OBS clock is assumed (e.g. Gouédard et al., 2014; Stähler et al., 2016).

With the inspection of the temporal evolution of CCFs, clock errors can be continuously monitored in land stations (Stehly et al., 2007; Sens-Schönfelder, 2008) as well as in OBSs (Sabra et al., 2005a; Gouédard et al., 2014; Hannemann et al., 2014; Le et al., 2018). The main limitation of this method is the distance between the stations. High-accurate clock corrections can only be achieved for small interstation distances of a few kilometres (Sens-Schönfelder, 2008). For larger interstation distances, clock errors can only be obtained with low accuracy and/or low temporal resolution (e.g. Stehly et al., 2007), which may obscure the true clock error magnitude (Xie et al., 2018a).

Prior studies used only one component, either the vertical (Sens-Schönfelder, 2008; Hannemann et al., 2014) or the hydrophone component (Sabra et al., 2005a; Gouédard et al., 2014; Hannemann et al., 2014; Le et al., 2018), while a multi-component approach that could improve the clock error determination has not been performed so far (see Chapter 2).

1.2.3 Crustal imaging

The CCF method is a powerful seismological tool to investigate underground structures, when temporal CCF changes (e.g. due to clock errors) can be excluded. As ambient seismic noise is recorded worldwide, CCFs enable seismic tomography studies in seismically quiet areas (e.g. Nicolson et al., 2012). But also in seismically active areas, this technique provides independent structural information in addition to classical earthquake-based tomography studies (e.g. Ma & Dalton, 2017).

Typically, group velocity curves are derived from stacked CCFs and subsequently inverted for a tomographic image of the earth's subsurface (Bensen et al., 2007). This method illuminates especially the crust and the upper mantle, as CCFs are dominated by surface waves that are sensitive to shallower earth layers.

Since Shapiro & Campillo (2004) have shown the great potential of this method for crustal imaging, the number of noise tomography studies has exploded. The interstation distances used in these studies comprise a few tens of metres and kilometres to investigate local structures (e.g. Brenguier et al., 2007; Mordret et al., 2014, 2015), a few hundreds of kilometres to survey regional and continental underground properties (e.g. Shapiro et al., 2005; Guo et al., 2013; Zigone et al., 2015; Goutorbe et al., 2015; Corela et al., 2017; Lu et al., 2018; Xie et al., 2018b), and even thousands of kilometres to derive global images of the earth (Nishida et al., 2009; Haned et al., 2016).

The vast majority of these noise tomography studies is based on seismic records of land stations, while only a limited number of studies used OBS seismograms (Yao et al., 2011; Mordret et al., 2014; Zha et al., 2014; Ball et al., 2016; Corela et al., 2017; Ryberg et al., 2017). The OBS studies are based on stations that are spaced between a few metres (e.g. Mordret et al., 2014) and a few hundred kilometres (e.g. Yao et al., 2011; Zha et al., 2014; Ball et al., 2016; Corela et al., 2017; Ryberg et al., 2017), while large-scale OBS noise tomography across distances of thousands of kilometres has not been performed yet (see Chapter 3).

The CCF method is a powerful method for crustal imaging and is applicable in all areas where seismic stations are installed. This explains the numerous noise tomography studies that have been conducted so far. However, there are still areas on earth that have never been investigated by this method or even by other seismological imaging techniques, e.g. the Piton des Neiges on the island La Réunion (see Chapter 4).

1.3 Objectives and outline of this thesis

This thesis intends to improve existing applications of noise cross-correlations and to expand previous limitations of these applications. The work presented in this thesis is divided into three chapters, where clock corrections (Chapter 2) as well as structural studies (Chapter 3 and Chapter 4) are addressed. The main part of the used data set of each chapter stems from OBSs and/or land stations that were installed on and around the island La Réunion during the Réunion Hotspot and Upper Mantle - Réunions Unterer Mantel (RHUM-RUM) experiment (Barruol & Sigloch, 2013; Stähler et al., 2016). An important part of Chapter 2 and Chapter 3 is the evaluation of the usability of OBSs compared to commonly used land stations.

1. Chapter 2 presents an extensive clock error study. Noise cross-correlations are calculated for an unprecedented large data set for clock error detection. The survey comprises seismic stations in different environments (land, ocean bottom), various interstation distances (10-370 km), different recording systems (seismometers, hydrophones), and various clock error types (non-linear, linear, jumps, gaps). Each processing step is documented and elucidated to provide a guideline for future studies. The great benefit of using multiple components instead of a conventional single-component approach (e.g. Sens-Schönfelder, 2008; Hannemann et al., 2014) is emphasized. The multi-component procedure allows the retrieval of clock errors with high temporal resolution and high accuracy (∼20 ms), even for large

interstation distances of >300 km. The demonstrated routine feasibility of high-accuracy clock corrections for such large interstation distances could previously not be envisaged.

2. In Chapter 3, a noise tomography of the crust and uppermost mantle of the western Indian Ocean around La Réunion is presented. Mazzullo et al. (2017) derived a 3-D S-wave velocity model for this region based on earthquake data. However, their data set was not suited to reliably invert for crustal structure. In order to improve the crustal model of Mazzullo et al. (2017), group velocity curves are retrieved from noise cross-correlations that are sensitive to shallower layers (<40 km). High-quality group velocities can be derived from OBS noise cross-correlations of interstation distances of up to >2000 km, while prior studies reported successful OBS group velocities for at most a few hundreds of kilometres. A joint inversion with the data set of Mazzullo et al. (2017) yields a high-resolution crustal S-wave velocity model. The velocity anomalies of this model can clearly be associated with tectonic structures indicating that the noise group velocities provide a valuable contribution to the model of Mazzullo et al. (2017).

3. In Chapter 4, group velocities of the island La Réunion are presented. Gravity studies detected a large hypovolcanic body beneath the Piton des Neiges, an ancient volcano of the island (e.g. Malengreau et al., 1999; Gailler & Lénat, 2012). No seismological evidence for this body existed up to now, as prior seismic studies have concentrated mainly on the active volcano of the island, the Piton de la Fournaise. Analogous to the method used in Chapter 3, group velocity curves are measured from noise cross-correlations between land stations installed on La Réunion. Inspection of the group velocities support the findings of the gravity studies and provide the first seismological indication of a high-velocity body beneath the Piton des Neiges.

Chapter 2

Corrections of clock errors in land and ocean bottom seismograms

This chapter was published in *Geophysical Journal International* in September 2018 under the title *Clock errors in land and ocean bottom seismograms: high-accuracy estimates from multiple-component noise cross-correlations* (Hable et al., 2018).

ABSTRACT

Many applications in seismology rely on the accurate absolute timing of seismograms. However, both seismological land stations and ocean bottom seismometers (OBSs) can be affected by clock errors, which cause the absolute timing of seismograms to deviate from a highly accurate reference time signal, usually provided by GPS satellites. Timing problems can occur in land stations when synchronization with a GPS signal is temporarily or permanently lost. This can give rise to complicated, time-dependent clock drifts relative to GPS time, due to varying environmental conditions. Seismometers at the ocean bottom cannot receive GPS satellite signals, but operate in more stable ambient conditions than land stations. The standard protocol is to synchronize an OBS with a GPS signal immediately before deployment and after recovery. The measured timing deviation, called “skew”, is assumed to have accumulated linearly over the deployment interval, an assumption that is plausible but usually not verifiable.

In recent years, cross-correlations of ambient microseismic noise have been put to use for correcting timing errors, but have been limited to interstation distances of at most a few tens of kilometres without reducing the temporal resolution. We apply noise cross-correlations to the evaluation of clock errors in four broad-band land stations and 53 wideband and broad-band OBSs, which were installed on and around the island of La Réunion in the western Indian Ocean during the RHUM-RUM (Réunion Hotspot and Upper Mantle - Réunions Unterer Mantel) experiment. We correlate all three seismic components, plus a hydrophone channel in OBS stations. Daily cross-correlation functions are derived for intermediate distances ($\sim$20 km) for land-to-land station pairs; stable, 10-day stacks are obtained for very large interstation distances up to >300 km for land-to-OBS, and OBS-to-OBS configurations. Averaging over multiple station pairs, and up to 16 component pairs per station, improves the accuracy of the method by a factor of four compared to the single-channel approaches of prior studies. The timing accuracy of our method is estimated to be $\sim$20 ms standard deviation, or one sample at a sampling rate of 50 Hz. In land stations, non-linear clock drifts and clock jumps of up to six minutes are detected and successfully corrected. For 52 out of 53 OBSs, we successfully obtain drift functions over time, which validate the common assumption of linear clock drift. Skew values that were available for 29 of these OBSs are consistent with our independent estimates within their observational error bars. For 23 OBSs that lacked skew measurements, linear OBS clock drifts range between 0.2 ms/day and 8.8 ms/day. In addition to linear drift, three OBSs are affected by clock jumps of $\sim$1 s, probably indicating a missing sample problem that would otherwise have gone undetected. Thus, we demonstrate the routine feasibility of high-accuracy clock corrections in land and ocean bottom seismometers over a wide range of interstation distances.

2.1 Introduction

A crucial requirement of many seismological processing methods is the correct absolute timing of seismograms. For the internal clocks of seismological land stations, accurate timing can usually be ensured by frequent synchronization with GPS satellites acting as a highly accurate, external reference clock. This synchronization can fail when reception of the GPS signal is lost, in which case the internal clock usually starts to drift noticeably relative to reference time (e.g. Sens-Schönfelder, 2008). The estimation and correction of such clock errors is a long-standing problem in seismology.

An analogous timing problem occurs in ocean bottom seismometers (OBSs), because GPS satellite signals do not reach the seafloor. Only two GPS synchronizations can be attempted: immediately before deployment and after recovery. If both GPS connections are successful, a timing deviation between the internal clock of the data logger that records the seismometer and hydrophone components and GPS clock is obtained, the so-called skew value. A first-order correction of internal clock timing can then be attempted by assuming that the skew accumulated gradually and linearly over the deployment interval (e.g. Gouédard et al., 2014; Hannemann et al., 2014). This is plausible because the rate of quartz oscillators in station clocks is expected to depend mainly on ambient temperature, which tends to remain very stable in deep water, although ultimately these assumptions remain unverified. If one or both GPS synchronizations failed, then not even this simple linear clock correction is possible.

Recent years have seen the rise of an independent observational method for estimating and correcting station clock errors, using cross-correlation functions (CCFs) of ambient seismic noise. It is based on the principle that noise CCFs feature a causal and an acausal part, which should occur at time lags of equal magnitude but opposite sign. Violation of this expectation can indicate the presence of clock errors, an "unphysical" biasing process in the sense that it is unrelated to seismic wave propagation (Stehly et al., 2007). Since Lobkis & Weaver (2001) developed the basic principle of noise cross-correlations in ultrasound experiments, CCFs have found wide application in seismology, including structural studies of the crust (e.g. Shapiro et al., 2005; Sabra et al., 2005b), global tomography (e.g. Haned et al., 2016), and the monitoring of seismic velocity changes along faults (e.g. Brenguier et al., 2008b; Wegler et al., 2009) and on volcanoes (e.g. Sens-Schönfelder & Wegler, 2006; Brenguier et al., 2008a; Sens-Schönfelder et al., 2014). A smaller number of studies investigated the use of noise cross-correlations for the correction of seismometer clocks. Stehly et al. (2007) and Sens-Schönfelder (2008) used CCFs to detect clock errors in land stations, while Sabra et al. (2005a), Gouédard et al. (2014), Hannemann et al. (2014) and Le et al. (2018) measured clock drifts of OBSs. In most cases, CCFs were calculated between stations with relatively small interstation distances, ranging from several metres (Sabra et al., 2005a), to several kilometres (Sens-Schönfelder, 2008; Gouédard et al., 2014), and up to tens of kilometres (Hannemann et al., 2014; Le et al., 2018). Obtaining CCFs for larger interstation distances usually came at expense of temporal resolution (Stehly et al., 2007, distance 200 km, monthly CCFs). Such long time averaging intervals may obscure the true time-dependence and magnitude of clock drifts (Xie et al., 2018a). Prior work used only a single-instrument channel: either the vertical seismic component of land stations (Sens-Schönfelder, 2008) and OBSs (Hannemann et al., 2014), or the hydrophone component of OBSs (Sabra et al., 2005a; Gouédard et al., 2014; Hannemann et al., 2014; Le et al., 2018). We are aware of only one study (Stehly et al., 2007) that uses several component pairs in order to discriminate between Rayleigh and Love waves.

In this study, we demonstrate the great potential of the noise cross-correlation technique for even larger interstation distances (>300 km) and finer time resolution (CCFs from daily or 10-day intervals). We also demonstrate a severalfold increase in accuracy of clock error estimates when correlating all three seismogram components in the case of land stations, plus a hydrophone channel (H) in the case of OBSs. We successfully retrieve CCFs between OBSs and land stations; prior work on this is very limited (e.g. Corela et al., 2017; Tian & Ritzwoller, 2017).

Section 2.2 describes our data, recorded by seismometers that were operating for several months to three years on and around the island of La Réunion in the western Indian Ocean, as part of the RHUM-RUM experiment (Réunion Hotspot and Upper Mantle - Réunions Unterer Mantel, 2011 to 2015). Section 2.3 describes the various processing steps for obtaining multi-channel CCFs and estimating the clock errors they imply. Section 2.4 presents results, subdivided by the three types of station pairs encountered: land-to-land clock errors; land-to-OBS; and OBS-to-OBS. The island stations on La Réunion were spaced by tens of kilometres, versus 200 km or more between two OBSs. Section 2.5 is a Discussion, followed by Conclusions in Section 2.6.

We used the open-source toolboxes Python (Rossum, 1995), ObsPy (Beyreuther et al., 2010; Megies et al., 2011; Krischer et al., 2015), and obspyDMT (Hosseini & Sigloch, 2017) for data downloading, processing, and plotting all figures except Figs 2.1 and 2.11, which were generated with the open-source mapping toolbox GMT (Wessel et al., 2013).

2.2 Data description

The RHUM-RUM experiment was deployed between 2011 and 2015 with the primary objective of imaging crust and mantle beneath the volcanic hotspot of La Réunion. The island, 70 km long and 50 km wide, is located in the western Indian Ocean, 800 km east of Madagascar and 200 km south-west of Mauritius (Fig. 2.1a). Its hotspot volcanism, time-progressive hotspot track, and associated large igneous province (Deccan Traps) has long marked it as a strong candidate for hosting a deep mantle plume underneath (Barruol & Sigloch, 2013). In order to investigate this hypothesis, 48 wideband OBSs from the German DEPAS pool and 9 broad-band OBSs from the French INSU pool were deployed around the island for roughly 13 months (October 2012 to November 2013), at water depths of 2200-5400 m. An extensive performance report of the OBSs is given by Stähler et al. (2016). In addition, 10 broad-band land stations were installed on La

Réunion, operating for roughly three years (mid-2012 to mid-2015). Additional RHUM-RUM island stations on Mauritius, Rodrigues, the Seychelles, Madagascar and the Îles Éparses were not part of our investigation. In total, the RHUM-RUM array covered 2000×2000 km^2 of the Indian Ocean and ocean islands.

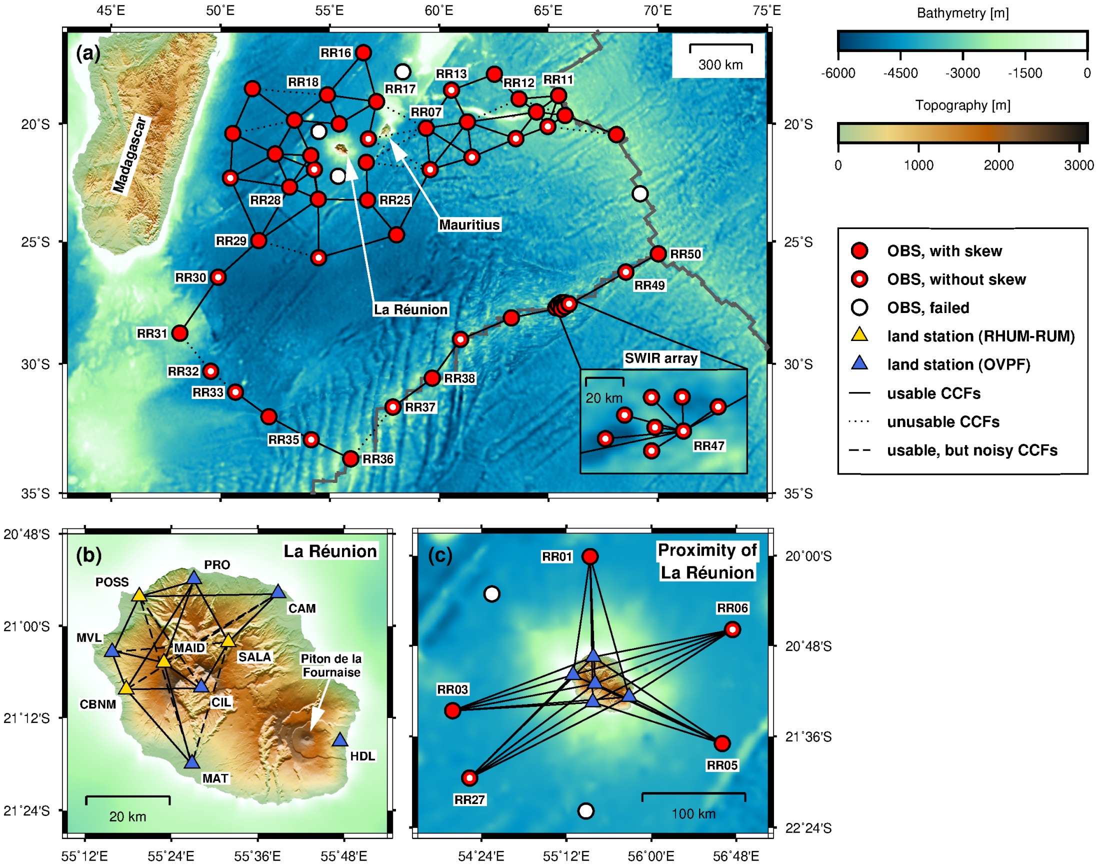


Figure 2.1: Maps of seismological broad-band stations used in this study, installed on and around La Réunion island. Circles represent ocean bottom seismometers deployed by the RHUM-RUM experiment: coloured red if a skew measurement was available; red with a white dot if no skew had been obtained; white if the OBS failed completely. Yellow triangles denote RHUM-RUM land stations; blue triangles denote permanent land stations operated by the OVPF volcano observatory. Stations explicitly mentioned in the text are labelled by their station names. Black lines connect pairs of neighbouring stations for which cross-correlations were attempted: OBS-to-OBS correlations in panel (a); land-to-land correlations in panel (b); and land-to-OBS correlations in panel (c). Solid black lines denote usable correlations; dotted lines denote unusable correlations, and dashed lines usable but noisy correlations. Grey lines in (a) indicate plate boundaries. Figure generated with the GMT toolbox (Wessel et al., 2013).

This study arose from the initial discovery of timing errors of up to several minutes in four out of ten stations on La Réunion (yellow triangles in Fig. 2.1b). All four stations featured the same type of data logger (RefTek RT 130), in which a software failure was causing loss of GPS synchronization. Upon detection of these timing problems, the GPS receiver units of the four malfunctioning stations were repaired while the data loggers and seismic sensor were kept running in the field, with their internal clocks adrift relative to GPS reference time. After reinstallation, the four repaired GPS units provided proper clock synchronizations until the entire array was dismantled approximately three months later.

The OBS deployment also encountered an unexpectedly high rate of GPS synchronization failures. Successful synchronization both before and after deployment, and thus a skew value, could be obtained for only 29 of the 57 OBSs (red circles in Fig. 2.1a). For these stations, the timing of the seismograms was corrected based on the assumption of linear clock drift (Stähler et al., 2016). For another 24 OBSs, GPS synchronization upon recovery failed because the OBS clocks had shut down prematurely (red circles with white dot in Fig. 2.1a). The likely reason were sudden, sharp voltage drops of the (poor-quality?) lithium batteries of the OBS clocks (Stähler et al., 2016). High-quality seismograms but unknown clock errors for these 24 OBSs prompted the extension of our terrestrial clock study to the oceanic realm. The remaining four OBSs had failed completely (white circles in Fig. 2.1a).

Our clock error study is divided into three parts: First, we evaluate the clock errors of the four land stations (CBNM, MAID, POSS, SALA) whose GPS units temporarily failed, setting their internal clocks adrift. We calculate daily CCFs between these erroneous stations and five reference land stations (CAM, CIL, MAT, MVL, PRO) whose clocks were properly synchronized to GPS throughout the RHUM-RUM deployment. These five permanent stations are operated by the volcano observatory of La Réunion, the Observatoire Volcanologique du Piton de la Fournaise (OVPF). Interstation distances on the island were 19 km on average (Fig. 2.1b). Second, we calculate CCFs between the five OBSs located closest to La Réunion (RR01, RR03, RR05, RR06, RR27) and five GPS-synchronized land stations operated by OVPF (CIL, HDL, MAT, MVL, PRO), in order to estimate the OBS clock drifts and to compare these estimates to measured skew values, which are available for RR01, RR03 and RR05 (Fig. 2.1c). Despite significant larger interstation distances of ∼140 km, we retrieve reliable results with a temporal resolution of 10 days for the CCFs. This prompted us to take the third step, of expanding the clock error study to the entire OBS network. We compute CCFs between neighbouring OBSs with interstation distances of up to 370 km (Fig. 2.1a). At this scale, verified skew-corrected OBSs serve as reference stations for OBSs that lack skew measurements.

2.3 Methodology

2.3.1 Pre-processing

Before the CCF calculation can be performed, continuous broad-band waveform data are pre-processed in order to remove contaminating signals that could overprint ambient noise, such as earthquakes. Our procedure largely follows Bensen et al. (2007). The pre-processing steps are applied to day-long time series for each station and component (land stations: E, N, Z; OBSs: 1, 2, Z, H). Time series with missing samples must either be rejected or filled with some *ad hoc* number. Filling of large gaps could however distort the resulting CCF. To overcome this problem and to reduce the effect of the numerous small gaps in our data, we divide the 24-hour long time series into 1-hour windows with an overlap of 50%, which yields 47 hour-long windows per day (records of 00:00am to 00:30am and 11:30pm to 00:00am are used only once). Gaps of less than 500 samples (corresponding to 5 s and 10 s for sampling rates of 100 Hz and 50 Hz, respectively) are filled with interpolated values, whereas windows featuring longer gaps are rejected. The amount of rejected windows is less than 1% for the land stations. The majority of the OBSs contains no gaps, except for 3 stations, where we exclude 13% (RR39), 24% (RR53) and 33% (RR33) of the hour-long windows. Even then a sufficient number of windows is obtained each day to calculate daily CCFs (see Section 2.3.2). The division into hour-long windows is redundant if the number of gaps is negligible. Next, the instrument response is removed from each hour-long window, a standard processing step (e.g. Bensen et al., 2007), even though strictly speaking it may not be necessary since we only rely on the CCF stability over time, not on its physical correctness. The instrument correction is followed by the removal of each window's mean value and trend. Then a zero-phase bandpass filter from 0.01 Hz to 10 Hz is applied. This wide frequency band gives us more flexibility in testing various narrower frequency ranges without repeating the entire, computationally expensive pre-processing before settling on a specific frequency band (see below: 0.05-0.5 Hz) for further processing.

A subsequent optional downsampling is performed as follows: the RHUM-RUM OBSs sampled at 50 Hz, 62.5 Hz, or 100 Hz, depending on instrument type (Stähler et al., 2016). For consistency and usability, all OBS data are downsampled to 50 Hz. Land stations that are correlated with OBSs are also downsampled to 50 Hz; for all other land stations, their original sampling rate of 100 Hz is retained. Although more severe downsampling would render the CCF computations less time consuming, it is avoided in order to estimate clock errors to the highest possible temporal resolution.

To reduce the effect of highly energetic signals such as earthquakes, each hour-long seismogram is amplitude-clipped at twice its standard deviation of that hour-long time window. These clipped time series are then spectrally whitened between 0.05 Hz and 0.5 Hz (2-20 s period). This period range is chosen because it contains the primary (∼14 s) and secondary microseisms (∼7 s), which are present in the La Réunion region throughout the year (Davy et al., 2015). Finally, 1-bit normalization is applied to the seismograms with the same purpose as the amplitude clipping. Larose et al. (2004) demonstrated in acoustic laboratory experiments that this technique improves the signal-to-noise ratio (SNR). It is therefore a widely used method in ambient noise pre-processing (e.g. Shapiro & Campillo, 2004; Shapiro et al., 2005; Hobiger et al., 2012).

2.3.2 CCF calculation

CCFs are generally calculated using

$$CCF_{k_1k_2}(s_1, s_2, t) = \int_{\tau_1}^{\tau_2} d_{k_1}(s_1, \tau) \cdot d_{k_2}(s_2, \tau - t)\, d\tau, \tag{2.1}$$

where $d_{k_1}(s_1, \tau)$ represents the seismogram of station s_1 and component k_1, while $d_{k_2}(s_2, \tau - t)$ denotes the time-reversed seismogram of station s_2 and component k_2. The times τ_1 and τ_2 indicate the start and end times of the CCF, respectively, and its lapse time is given by t. CCFs consist of a causal and an acausal part, which should be symmetric for homogeneously distributed noise sources. The causal part represents the response (the so called Green's function) of station s_2 to a delta pulse at location of station s_1, while the acausal part corresponds to the impulse response of s_1 with a source at position s_2.

In nature, noise sources are often distributed unevenly, which leads to asymmetric CCFs (Stehly et al., 2007). Temporal CCF changes can be caused by several mechanisms, which are summarized by Stehly et al. (2007). Velocity changes in the subsurface would affect a CCF's causal and acausal parts in analogous ways, by dilating or compressing the CCF's waveforms on either side of t=0, corresponding to slower or faster wave propagation, respectively. Such changes in wave propagation are expected in the vicinity of faults after large earthquakes (e.g. Brenguier et al., 2008b) or close to volcanoes, reflecting magma movements (e.g. Brenguier et al., 2008a; Sens-Schönfelder et al., 2014). However, there are no large earthquake faults on La Réunion and the island's active volcano, the Piton de la Fournaise, is too far away to have an effect on our CCFs (see Fig. 2.1b).

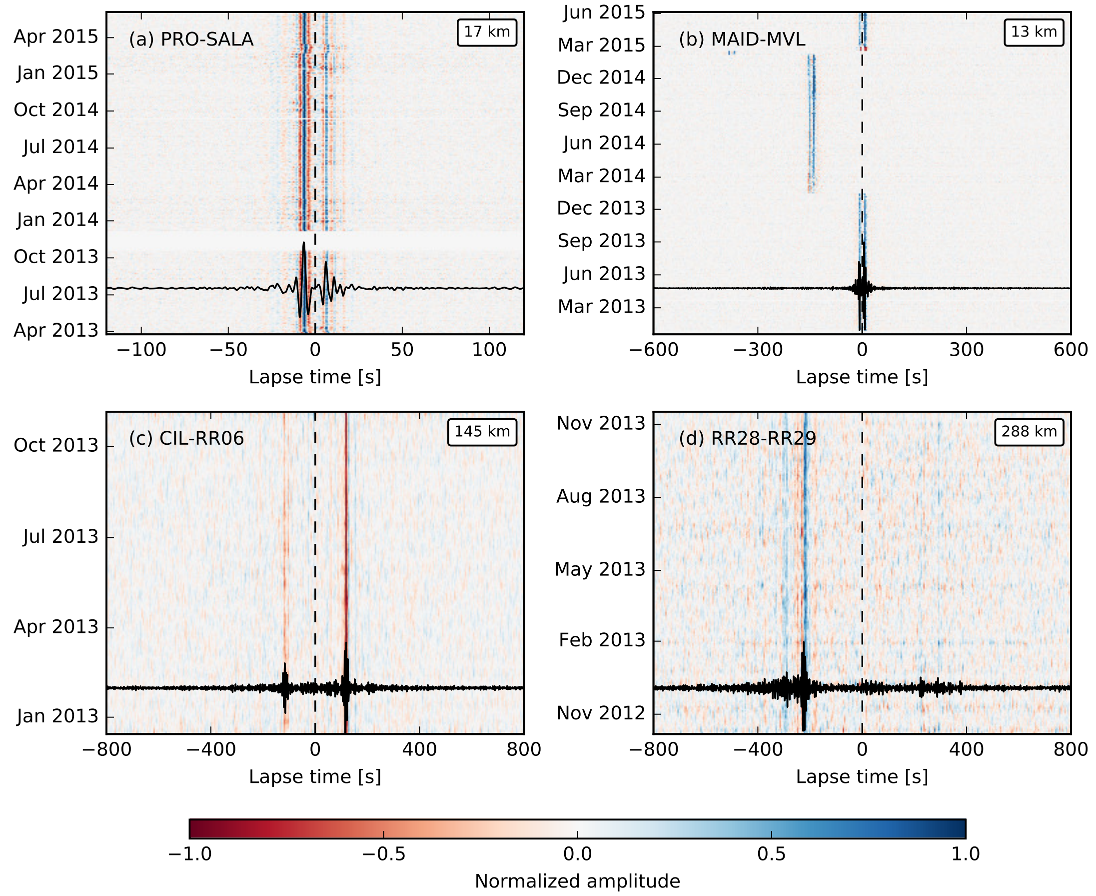


Figure 2.2: Examples of cross-correlation functions (CCFs) for the different types of station configurations and interstation distances. Colour shading represents the normalized amplitudes of the time series. (a) Station pair PRO-SALA (land-to-land correlation, distance 17 km, ZN-component, daily stacks). (b) Station pair MAID-MVL (land-to-land, distance 13 km, ZE-component, daily stacks). (c) Station pair CIL-RR06 (land-to-OBS, 145 km, EZ-component, 10-day stacks). (d) Station pair RR28-RR29 (OBS-to-OBS correlation, 288 km, HH-component, 10-day stacks). In each panel, the black time series represents the RCF, that is, the stack over all individual CCFs. The dashed vertical line indicates zero lapse time; interstation distance is given in the top right of each panel.

Temporal changes in the locations of noise source should be observed as shape changes of a CCF's causal and acausal parts independently of each other. This effect can be neglected because our tropical and subtropical stations record microseismic noise sources (2-20 s) that show little spatial seasonality (Davy et al., 2015). This is confirmed by the marked stability over time in our CCFs (Fig. 2.2).

In contrast to the above, instrument clock drifts manifest themselves by time-shifting the entire CCF, so that its causal part moves closer to t=0 and its acausal part moves further away, or vice versa. We evaluate such shifts as described in Section 2.3.3. CCFs would be affected in the same way by the occurrence of a phase change in the instrument response.

We compute CCFs of our pre-processed, hour-long traces according to eq. (2.1). Daily CCFs are obtained by averaging all hourly CCFs in a day (up to 47). For the 3 OBSs (RR33, RR39, RR53) affected by numerous gaps, we can use the sufficient number of ∼30-40 windows per day for the calculation. The start time τ_1 and end time τ_2 of the hourly CCFs depend on interstation distance and on the expected clock error: we use lapse times from -120 s to +120 s for correlations of two land stations (Fig. 2.2a). An exception was station MAID, which showed severe clock jumps of several minutes, necessitating start and end times of ±600 s (Fig. 2.2b). For the larger interstation distances of land-to-OBS and OBS-to-OBS correlations, lapse times run from -800 s to +800 s (Figs 2.2c and d).

We characterize the quality of the daily CCFs by defining their SNR as the ratio of the maximum absolute value of a signal window to the standard deviation of a noise window:

$$SNR = \frac{max(|CCF_{signal}(t)|)}{std(CCF_{noise}(t))}. \tag{2.2}$$

The time windows are based on interstation distance. The signal window is chosen in the early lapse times (land-to-land correlations: -25 s to +25 s, land-to-OBS: -150 s to +150 s, OBS-to-OBS: -400 s to +400 s) expected to contain high energetic wavetrains; the noise window is defined in the late lapse times (land-to-land: ±80 s to ±120 s, land-to-OBS and OBS-to-OBS: ±650 s to ±800 s), where noise is dominant. For land-to-land station pairs, we reject daily CCFs if their SNR is below 7. For land-to-OBS and OBS-to-OBS correlations, this SNR threshold is chosen as 1, since we expect a generally lower SNR for these large-distance correlations. The mean SNR (SNRs averaged over the data period) ranges from 3.5 to 20, but we chose this much lower rejection threshold of 1 in order to reject only obviously useless windows, not those that are noisy but may still contain usable information. To increase the SNR of land-to-OBS and OBS-to-OBS correlations, we stack up to 10 daily CCFs (of SNR $\geq$1) from a 10-day time window into 10-day stacks, where the 10-day stacking window is moved in increments of one day over the entire data period.

Next we calculate a reference correlation function (RCF) by stacking all daily CCFs that passed the SNR thresholding. For the RCF calculation for station MAID we exceptionally consider only the time period prior to the first extreme clock jump in January 2014 (about one year of data). RCFs are displayed as black curves in Fig. 2.2.

Comparing the different station pairs of Fig. 2.2 reveals that clear wave arrivals in the CCFs can be retrieved continuously over the entire operation period. The SNR of daily CCFs decreases with increasing interstation distance (compare PRO-SALA and RR28-RR29). The examples of CIL-RR06 (land-to-OBS) and RR28-RR29 (OBS-to-OBS) demonstrate that stable, 10-day-stacked CCFs can be obtained for interstation distances of several hundred kilometres. The high-amplitude wave trains of PRO-SALA and MAID-MVL most likely correspond to Rayleigh and Love waves, whereas the wave packages of CIL-RR06 and RR28-RR29 arrive significantly later than expected for surface wave velocities of roughly 3 km/s. They may be associated with Scholte waves (Scholte., 1947) because their velocity is in the range of 0.8-1.5 km/s (Flores-Mendez et al., 2012; Le et al., 2018). Scholte waves of similar velocities were used by Le et al. (2018) to examine OBS clock errors. For estimating clock errors, it is not necessary to know the exact nature of the slow wave packages as long as we are able to retrieve clear and stable wave trains throughout the study period, which is clearly the case in Fig. 2.2.

Clock drifts are often so tiny that they cannot be detected by eye in the CCF plots, for example, Fig. 2.2(c). Minor CCF drifts are visible for PRO-SALA (e.g. in March 2015). By contrast, MAID-MVL is dominated by several large clock jumps: +2.5 minutes in January 2014, an additional 3.5 min jump in early February 2015, and a negative jump back to the initial state in late February 2015, after reinstallation of the repaired GPS unit.

2.3.3 Clock error measurement

For every station pair and component pair, the clock error of each day is measured by performing a cross-correlation between the RCF and the daily CCFs (for land-to-land station pairs) or the 10-day stacks (for land-to-OBS and OBS-to-OBS pairs). The clock error is taken to be the time shift that maximizes the Pearson correlation coefficient (CC), which is the maximum amplitude of the cross-correlation between RCF and CCF. Each daily or 10-day CCF is associated with a CC value; the average over all such values is CC_{av}, the average CC achieved over the duration of the deployment, for a given station pair and component pair. If the CC of an individual CCF drops below a certain threshold, chosen as $0.85 \cdot CC_{av}$, then this CCF is rejected in order to ensure a consistently high quality of the daily (or 10-daily) clock error estimates.

Per station pair, this yields up to 9 daily clock error measurements ϵ_i from 3×3 component pairs for land-to-land correlations (EE, EN, EZ, NE, NN, NZ, ZE, ZN, ZZ), and also for land-to-OBS correlations (E1, E2, EZ, N1, N2, NZ, Z1, Z2, ZZ). OBS-to-OBS station pairs yield up to 16 clock error estimates ϵ_i from 4×4 component pairs (11, 12, 1H, 1Z, 21, 22, 2H, 2Z, H1, H2, HH, HZ, Z1, Z2, ZH, ZZ).

The relatively large scatter that affects clock error estimates derived from a single component can be decreased substantially by forming a weighted average over estimates from all 9 or 16 components, following Hobiger et al. (2012):

$$\epsilon(t) = \frac{\sum\limits_{i=1}^{N} CC_i^2(t) \cdot \epsilon_i(t)}{\sum\limits_{i=1}^{N} CC_i^2(t)}, \tag{2.3}$$

where ϵ denotes our best estimate of the clock error of one station pair, CC_i is the CC of component pair i, and N is the number of usable component pairs. The correlation coefficients CC_i^2 serve as weights for the individual clock errors ϵ_i. Similarly, a weighted average $CC(t)$ is calculated for each station pair by averaging over its component pairs (Hobiger et al., 2012):

$$CC(t) = \frac{\sum\limits_{i=1}^{N} CC_i^3(t)}{\sum\limits_{i=1}^{N} CC_i^2(t)}. \tag{2.4}$$

Fig. 2.3 visualizes the substantial benefit of averaging over component pairs, on the example of station pair CIL-SALA (time period December 2013 to February 2015). Fig. 2.3(a) shows $\epsilon(t)$, our best estimate of the CIL-SALA clock error obtained as the weighted average (eq. 2.3) over the error estimated from nine components, which are individually shown in Figs 2.3(b)-(j). The orange curve, repeated in all 10 panels, is the best spline fitting function of polynomial degree 3 to $\epsilon(t)$ in Fig. 2.3(a). Standard deviations σ to this curve are calculated for the averaged clock error $\epsilon(t)$ as well as for individual component pairs, and are shown in the bottom right of each panel (Figs 2.3a-j). It is evident from these values and also from visual comparison that Fig. 2.3(a) shows a much reduced scatter compared to the other nine panels. In general, the σ of a weight-averaged $\epsilon(t)$ is typically two to three times smaller than the σ of its individual constituent components. This observation can be expected from statistics, where it is known as standard error: the standard deviation is reduced by the square root of the number of measurements (Hughes & Hase, 2010):

$$\sigma = \frac{\sigma_i}{\sqrt{n}}. \tag{2.5}$$

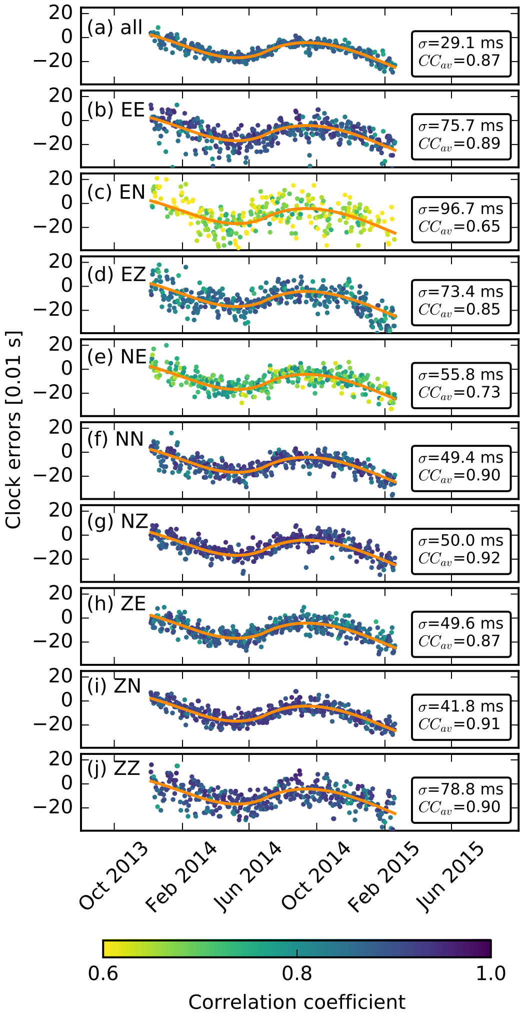


Figure 2.3: Clock drift of island station SALA relative to island station CIL from December 2013 to February 2015, as estimated from multi-component cross-correlations. (a) shows $\epsilon(t)$, the clock drift estimates over time obtained from averaging over all nine individual component pairs. Each dot represents the result from one daily CCF. (b)-(j) show the same clock drift as estimated by the nine individual component pairs (East-with-East, East-with-North, etc.). Clock drift estimates are given in units of 0.01 s, corresponding to one sample. The clock errors of (a) are fitted by a polynomial function given by the orange curve, which is also plotted in all subsequent panels. Boxes on the right of each panel state measurement uncertainty σ in milliseconds, defined as the standard deviation of the ensemble of daily measurements from the orange fitting curve. σ thus quantifies the accuracy of our clock error estimates $\epsilon(t)$: the σ and point scatter in panel (a) are strongly reduced compared to those of any individual component in (b)-(j). CC_{av} is the average cross-correlation coefficient of the daily CCFs (average taken over the entire deployment period).

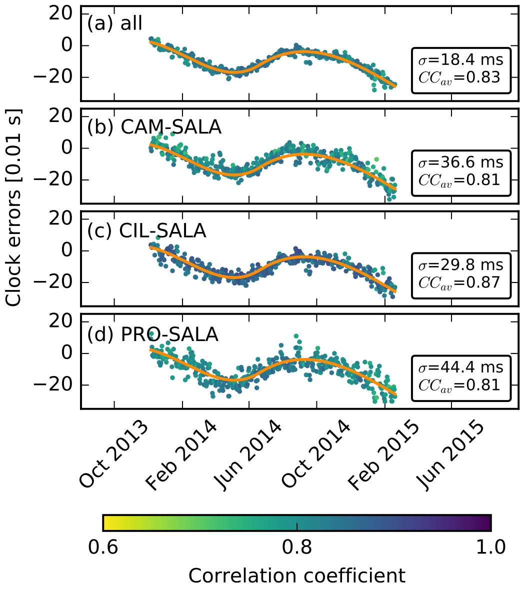


Figure 2.4: Clock drift of island station SALA relative to the three island stations CAM, CIL, and PRO used for correlation, for the same time period as in Fig. 2.3. (a) shows daily $\epsilon(t)$ estimates after averaging over all three station pairs, together with the best-fitting polynomial curve (in orange). Panels (b)-(d) show daily $\epsilon(t)$ estimates for the three individual station pairs (CAM-SALA, CIL-SALA, PRO-SALA), each already averaged over up to nine component pairs, and superimposed by the same polynomial curve as in (a). Boxes on the right contain uncertainty estimates σ and averaged CCF correlation coefficients, as defined in caption of Fig. 2.3. Note that panel (c), CIL-SALA, is identical to panel (a) in Fig. 2.3.

In our case, σ_i represents the standard deviation of an individual channel pair and n indicates the number of channel pairs used for averaging. Thus, we would expect that the standard deviation is reduced by a factor of three when 9 component pairs are used instead of one.

For the average and each component in Figs 2.3(a)-(j), we also show the time-averaged correlation coefficient CC_{av}, that is, CC values averaged over the period of December 2013 to February 2015. High CC values are seen to generally coincide with tightly clustered curves, and are thus an indication for a high-confidence clock error measurement. The exception is component pair NE with a relatively low CC_{av} of 0.73 (e.g. compared to ZZ with CC_{av}=0.90), and yet its standard deviation is significantly lower (NE: σ=55.8 ms, ZZ: σ=78.8 ms).

The clock error estimate of a station can be further improved by averaging over all station pairs including this station. This step is analogous to the average over component pairs in eqs (2.3) and (2.4), but with N now indicating the number of station pairs. Averaging over multiple station pairs can only be performed when it is known that only the station common to all station pairs can be affected by timing problems, whereas the timing of all other stations is assured (e.g. due to GPS synchronization over the entire recording period). Else more than one clock error would be present, with no clear path separating them out. Fig. 2.4 demonstrates the benefit on the example of station SALA, where the averaged error estimate in Fig. 2.4(a) is much less scattered, and has a lower σ of 18.4 ms, than individual estimates from station pairs CAM-SALA (σ=36.6 ms), CIL-SALA (σ=29.8 ms), or PRO-SALA (σ=44.4 ms) in Figs 2.4(b)-(d). Jointly considering Fig. 2.3(j) (σ=78.8 ms for component pair ZZ) and Fig. 2.4(a) (σ=18.4 ms), our example shows that compared to the conventional method of using only the ZZ-component of one station pair, clock error accuracy can be improved by a factor of 4 by averaging over all available components and station pairs.

2.4 Results

2.4.1 Correlations of land stations to land stations

The four problematic RHUM-RUM island stations SALA, CBNM, POSS and MAID, which had suffered from temporary outages of their GPS units, were correlated with five stations (CAM, CIL, MAT, MVL, PRO), which are permanently operated by the OVPF volcano observatory and served us as reference stations. This yielded 20 station pairs, with interstation distances ranging between 10 km and 43 km. Measurements over the nine component pairs were averaged as described in Section 2.3.3. For averaging over station pairs, only the three or four best station pairs were taken into account, as indicated by the solid black lines connecting station locations in Fig. 2.1(b). The OVPF stations are designed to record data only when GPS signal is available (Valérie Ferrazzini, personal communication, 15/09/2015), and should therefore never be afflicted by clock errors. We confirmed this by correlating the reference stations with each other, and not observing any clock errors. (Note that this also serves as an independent check on the validity of our method.) We conclude that the clock error measurements presented below originated solely from the RHUM-RUM island stations.

Fig. 2.5 shows our best estimates of the relative clock errors of SALA, CBNM, POSS and MAID over time. Each dot represents the clock error of one day, so that a rising trend indicates an increasing cumulative error. Clock errors are given in units of samples (1 sample $\equiv$ 0.01 s), except for station MAID, whose large errors are plotted in seconds. Positive clock errors are caused by clocks that run fast, causing the waveforms to appear delayed. Gaps in the data reflect periods when the recording stopped until it was restarted during station servicing visits, which happened every few months. Since the clock errors of these four stations were caused by loss of GPS synchronization, the

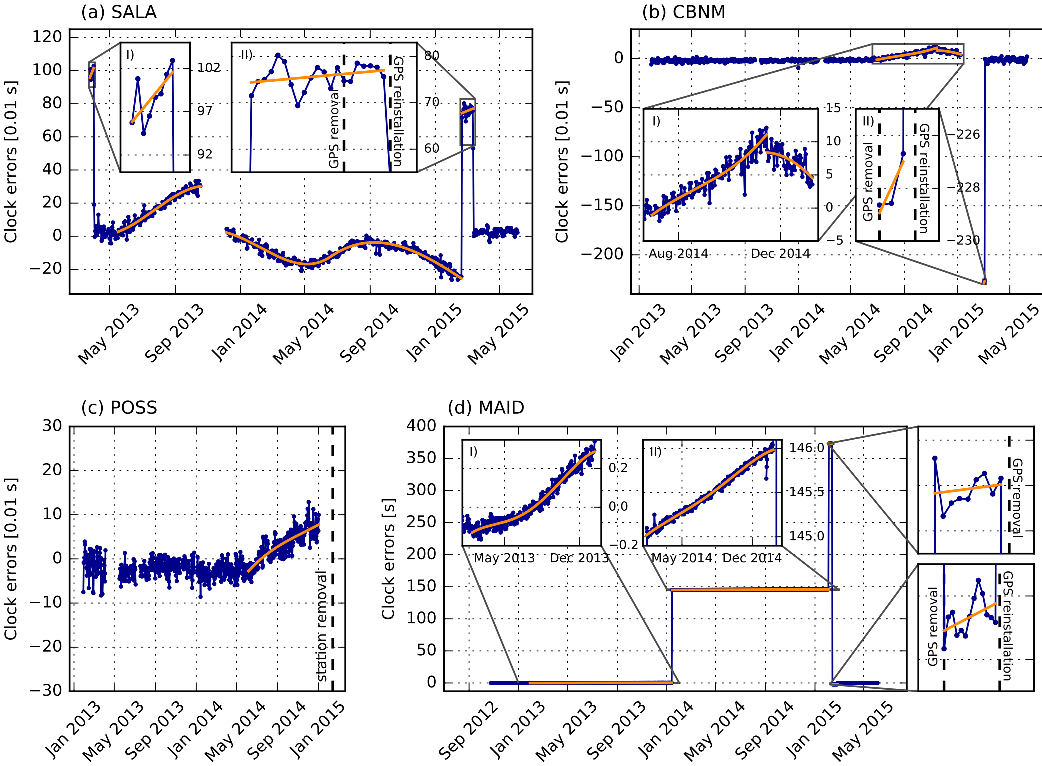


Figure 2.5: Best estimates of clock drift for all four investigated island stations, obtained by averaging over all available component and station pairs. Blue dots represent daily clock error estimates for (a) SALA, (b) CBNM, and (c) POSS, given in units of 0.01 s, corresponding to one sample. Panel (d) shows results for station MAID, in units of seconds. The dates of GPS unit removals and reinstallations are represented by vertical dashed lines. Time intervals during which the clocks were unsynchronized and thus adrift are fitted by polynomial functions (orange curves). Note that part of panel (a), drift of station SALA from Dec. 2013 to Feb. 2015, is identical to panel (a) in Fig. 2.4.

time period immediately after reinstallation of a repaired GPS unit can be regarded as reference period during which the station clock is known to have worked properly. The clock error estimates for such a reference period need not be centred around time zero, because the clock error measurement is a relative measurement: one-day or 10-day CCFs are correlated with an RCF obtained by averaging CCFs over much or all of the operating period. This means that in general RCFs can be contaminated by clock errors as well.

For station SALA in Fig. 2.5(a), comparison of the reference period (after GPS reinstallation in mid-March 2015, see zoom inset II on the right) with earlier times indicates that the station had no GPS signal when it was installed in late March 2013. After eight days (zoom inset I on the left), the station probably latched onto a satellite signal and synchronized, which caused a clock jump of around 1 s, followed by a period of no significant drift during April 2013. The GPS connection appears to have been lost again in

May 2013, as evidenced by the onset of a gradually accumulating, positive clock error. Subsequent clock errors manifest as gradual, non-linear clock drifts (fitted by orange polynomial splines), as well as clock jumps of 0.3-1 s.

In stations CBNM (Fig. 2.5b) and POSS (Fig. 2.5c), the clock worked normally during most of the operation period, until it started to drift in late June 2014 in CBNM (left inset I and orange fitted curve in Fig. 2.5b). After a recording stop of CBNM in January 2015, the station's GPS sensor was removed and reinstalled three days later, which manifests as a jump of 2.3 s (right inset II), back to relative clock error values around zero. Similarly, the clock of station POSS started to drift in early June 2014 (orange line in Fig. 2.5c). Station POSS was removed from the field in February 2015, and its repaired GPS sensor was reinstalled at station MAID, whose GPS sensor could not be repaired. Station MAID (Fig. 2.5d) exhibits striking clock jumps of up to 6 minutes, which had already been identified in the CCF plot of Fig. 2.2(b) (see Section 2.3.2). Zooming into time periods between jumps reveals non-linear drift behaviour (orange curves), as for the other stations.

The purpose of computing the fitting polynomials is to correct for clock errors according to these best-fitting curves, instead of using the scattered daily estimates. This follows the rationale that true clock drifts are probably rather smooth because the physical oscillators are unlikely to run fast on one day and slow on the next. Polynomial degrees are chosen according to the length of the drift interval: linear if the interval is of only a few days (e.g. SALA, insets I and II of Fig. 2.5a), whereas for longer periods a polynomial degree up to 4 is used (e.g. MAID, insets I and II of Fig. 2.5d). The non-linear clock drifts for longer periods are in the range of -2.1 ms/day to +2.4 ms/day for SALA and -1.3 ms/day to +1.5 ms/day for CBNM. POSS shows a moderate clock drift with 0.5 ms/day on average, while the drifts for MAID range between 0.4 ms/day and 2.9 ms/day.

2.4.1.1 Implementation of clock corrections on the archived RHUM-RUM data and meta-data

The ultimate purpose of estimating clock errors is to correct seismograms archived at the data centre such that end users can rely on downloading correctly timed data and meta-data. For implementing this timing correction, we use the polynomials fitted to the estimated clock error curves, rather than individual daily values of the clock error curves themselves. The RHUM-RUM time series are archived as miniSEED at the RESIF data centre (http://dx.doi.org/10.15778/RESIF.YV2011) in data records of 10 s to 15 s length. Hence the correction procedure (Wayne Crawford, personal communication,

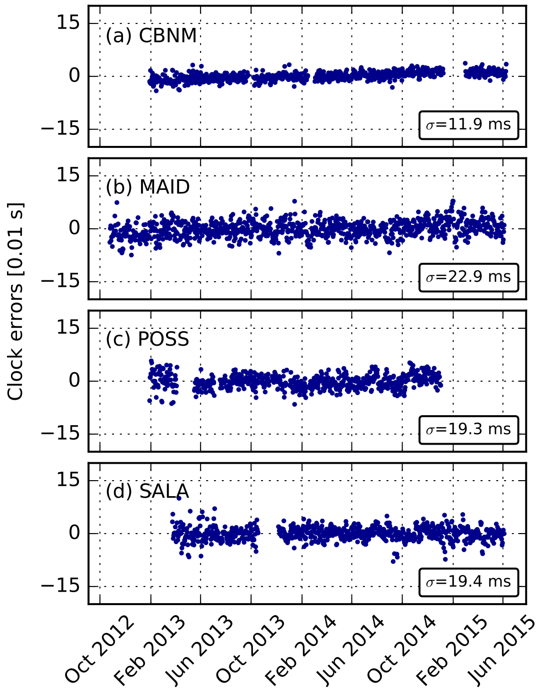


Figure 2.6: Residual clock errors of land stations (a) CBNM, (b) MAID, (c) POSS and (d) SALA after timing correction, that is, subtraction of the orange curve segments from the daily estimates in Fig. 2.5. Estimation uncertainty σ, here defined as standard deviation from zero, is given in the bottom right of each panel. They quantify the accuracy of our method for clock error estimation.

17/12/2017) looks up the value of a fitting curve in absolute time increments of 10-15 seconds and modifies the header of each data record accordingly, more specifically header fields 8 (start time of record); 16 (time correction); and 12 (bit 1: time correction applied). The time-corrected RHUM-RUM seismograms will be made available through the RESIF data centre in 2018. Time-corrected seismograms at the data centre have a data quality flag of “Q” whereas the uncorrected seismograms have a data quality flag of “D”.

The user experience of this clock correction might differ depending on a user’s data handling software, and by the length of the time series requested. For example, ObsPy (Beyreuther et al., 2010; Megies et al., 2011; Krischer et al., 2015) works with a single sampling rate per time series. Hence if user requests two days of data, divided into two consecutive chunks, then the last sample of the first chunk will be wrong by the equivalent of one day’s drift, whereas the first sample of the second time series will be essentially correct, and spacing of these two samples will differ from the (constant) sampling interval within each chunk.

In order to verify our clock error corrections, the entire cross-correlation procedure was repeated on the time-corrected seismograms, divided into 24-hour chunks. Fig. 2.6 shows the results: flat clock error values around zero and over each station's entire operation period indicate that all clock errors have been corrected successfully, even the large clock jumps of station MAID. The scatter of the residual clock errors around zero can be used to assess the accuracy of our method, as discussed in Section 2.5.1.

This same approach to rewriting meta-data at the data centre and verification of the corrected seismograms has been applied to clock error estimates obtained for OBS stations, as described in Sections 2.4.2 and 2.4.3.

2.4.2 Correlations of land stations to the nearest OBSs

Clock drifts of the five OBSs deployed closest to the island of La Réunion (RR01, RR03, RR05, RR06, RR27) are evaluated by computing their CCFs with five reference land stations (CIL, HDL, MAT, MVL, PRO). These are essentially the same OVPF reference stations as used for land-to-land correlations, except that HDL is used instead of CAM, because CAM was not installed until July 2013. Correlation of five OBSs with five reference land stations yielded 25 station pairs, with interstation distances ranging between 99 km and 174 km. Clock error averaging over the nine component pairs and five station pairs is performed according to eq. (2.3). We chose not to include the hydrophone component in land-to-OBS correlations because those already benefitted from the large number of station pairs to average over (see Section 2.5.1).

Immediately before deployment, each OBS clock was synchronized with the GPS clock, so that the clock error is known to have been zero at the very beginning of the operation period. Immediately after OBS recovery, a second GPS synchronization attempt was made; if successful, the difference of OBS clock and GPS clock yielded the "skew", or cumulative OBS clock error. Skew is defined as OBS time minus GPS time, so that a positive skew value indicates an OBS clock that runs fast compared to the GPS clock. The time difference (i.e. the OBS time error) is thought to be due to the temperature-dependent oscillator of the internal OBS clock. Due to near-constant temperatures at the bottom of the deep ocean (4 °C), a constant clock drift rate during the operation period is usually anticipated (e.g. Gouédard et al., 2014; Hannemann et al., 2014; Stähler et al., 2016), which suggests that seismograms should be corrected by linear interpolation of the skew for the total period of OBS operation. (Lingering doubts remained whether relatively large clock drifts might occur during an OBS' free fall to the seafloor upon deployment, and/or rapid rise during recovery, discussed in Section 2.5.5.)

Prior to this study, the timing of RHUM-RUM stations RR01, RR03 and RR05, for which a skew had been successfully obtained, had been corrected by a linear skew interpolation of daily time series (Stähler et al., 2016). No correction was possible for RR06 and RR27, because the OBS clock had stopped working before recovery and thus the second GPS synchronization had failed.

Fig. 2.7 shows daily clock error estimates for the five OBSs. To the extent that OBS clocks actually drift linearly, we would expect to observe no clock errors for RR01, RR03 and RR05 in Figs 2.7(a)-(c). This is indeed the case to good approximation, with flat curves scattering around zero, and in particular no or only weak differing trends near the very beginning and end of the recording periods. This provides independent confirmation that linear skew interpolation is an appropriate clock error correction for OBSs. The tiny jumps at the beginning and end are very likely processing artefacts, rather than a true clock drift, due to the fact that the first and last stacks contain only one CCF, whereas the second and second-to-last stacks contain two CCFs.

Processing the non-corrected seismograms of RR06 and RR27 with our method reveals clear clock drifts that are well approximated by a linear fit, with slopes of 0.91 ms/day for RR06 and 0.40 ms/day for RR27 (Figs 2.7d and e). After applying these linear timing corrections and re-running the processing (calculation of daily CCFs, computation of 10-day stacks and RCF, clock error estimation), a small clock drift remained measurable. This may be due to the usage of 10-day CCF stacks (instead of daily), which flattens the clock drift slightly. In addition, the correlation with a RCF affected by clock errors may play a role (Sens-Schönfelder, 2008; Gouédard et al., 2014). A second iteration of linear clock drift correction is necessary and sufficient to remove the clock error completely. Thus, the full clock drift is the sum of drifts from the first and second iterations (RR06: 0.91 ms/day + 0.15 ms/day = 1.06 ms/day, RR27: 0.40 ms/day + 0.11 ms/day = 0.51 ms/day). The necessity of several iterations will be explained in more detail in Section 2.4.3.

These are very encouraging results, indicating that our 10-day CCF stacking period still offers sufficiently high timing resolution to successfully estimate and remove OBS clock drifts. The 10-day stacking interval is necessary to obtain sufficiently high SNRs for stable CCFs when OBS interstation distances are up to several hundred kilometres. Our study is the first to present successful OBS cross-correlations and clock estimates over such large distances with the comparatively short stacking length of 10 days.

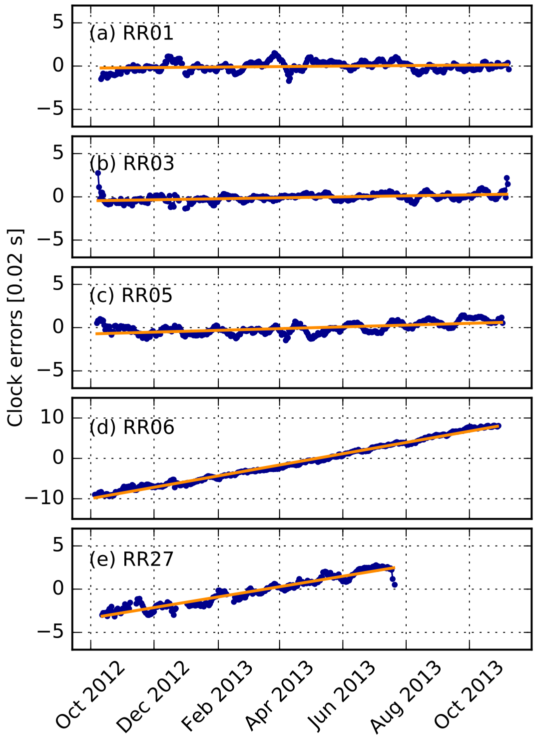


Figure 2.7: Daily clock drift estimates from the first iteration for ocean bottom stations, obtained through correlation with nearby island stations (land-to-OBS). (a) Station RR01, (b) RR03, (c) RR05, (d) RR06, and (e) RR27. Drift values are given in units of 0.02 s, corresponding to one sample. Seismograms of RR01, RR03, and RR05 had been linearly skew-corrected prior to correlation. The observation that their daily estimates scatter around zero with no discernible trend over deployment time implies that their clock drift rate was indeed almost constant and accurately quantified by the GPS skew measurement. No such independent skew measurements were available for RR06 and RR27. Their daily estimates show clear, almost linear trends over deployment time, again confirming that clock drift rates were almost constant. Drift rates are estimated as the slopes of the best-fitting orange lines, and listed in Table 2.1 for all OBSs.

2.4.3 Correlations of OBSs to OBSs

The encouraging outcome from land-to-OBS correlations in Section 2.4.2 led us to expand our study to the entire OBS network. This permits to verify measured skews, the general applicability of linear skew interpolation, and the estimation of clock drifts for OBSs that lack skew measurements.

We calculate 10-day stacked CCFs of 89 OBS station pairs, using neighbouring stations with distances ranging from 16 km to 374 km, averaging ∼209 km (Fig. 2.1a). We include the hydrophone component, which is particularly useful because the seismic channels of several OBSs failed, whereas the hydrophones worked very reliably (Stähler et al., 2016). The up to 16 component pairs are averaged according to eq. (2.3), as in previous sections.

By comparing the clock error curves over time of individual component pairs, we detected that the H-component of several OBSs shows a behaviour that differs from the seismometer components. In most cases, the deviations between the curves were limited to the first few months of the operation period. We find that this behaviour can be attributed to a change in hydrophone noise levels, probably related to a protracted settling period on the seafloor, which amounts to an unexpected malfunctioning of the hydrophone model used. Details are still under investigation, see Appendix A. To avoid a degradation of our clock error estimates, we do not admit CCFs that included a hydrophone channel prior to its settling into proper operation, which is characterized by a higher dynamic range of the instrument response (Fig. 2.15c). Eq. (2.5) predicts a reduction in standard deviation by a factor of $\sqrt{12} \approx 3.5$ from averaging over 12 component pairs, compared to the hypothetically possible $\sqrt{16} = 4$ in the case of 16 component pairs. This relatively small loss in estimation accuracy led us to exclude the questionable hydrophone components.

We proceed in two stages: (1) For the 29 OBSs where skew values had been obtained, we verify these skews and the validity of their linear interpolation over the recording period for the purpose of clock correction. (2) For the 24 OBSs without skew, we estimate and correct clock drifts, using as reference stations the previously verified OBSs with skews.

For a given OBS with a skew value, Stage 1 estimates the full time dependence of its clock errors by attempting to average over several correlating OBS pairs, as in previous sections (eq. 2.3 again, where N is now the number of neighbouring stations). However, the requirement that all correlating OBSs need to be in the “with skew” group, means that only a few correlating pairs may be available (e.g. only RR16-RR17 and RR16-RR18 for station RR16). In some cases, only an indirect skew verification can be carried out, for example, for RR31, which can only be correlated with OBSs that lack a skew measurement. RR29-RR30 and RR30-RR31 show the same clock drift, indicating that the source of the clock drift is the station lacking a skew measurement (RR30). The timing of the station RR31 can therefore be inferred as accurate, especially since the timing of RR29 can be verified directly with several other station pairs. We find that 26 out of 29 skew-corrected stations show no additional clock drift as expected, which verifies both the measured skew values and the validity of their linear interpolation.

The three exceptions are station RR12, which shows a small linear clock drift of -0.65 ms/day, and stations RR07 and RR11, each of which experienced one clock jump of roughly one second during their respective operation periods (Fig. 2.8a). By comparing

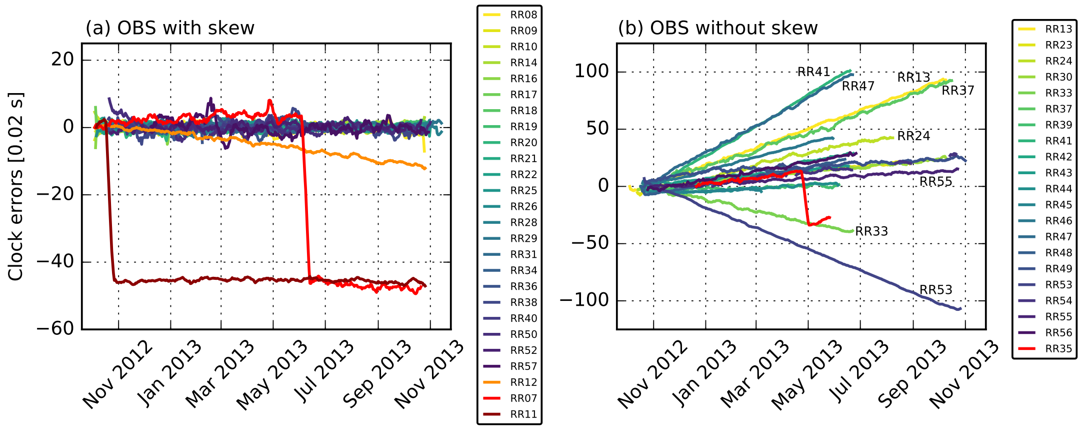


Figure 2.8: Daily clock drift estimates for ocean bottom stations, obtained through correlation with neighbouring OBSs (OBS-to-OBS). (a) Clock errors of all OBSs that could be linearly time-corrected prior to correlation, thanks to existing skew measurements. Drift values are given in units of 0.02 s, corresponding to one sample. The vast majority of stations show no trend away from zero, providing direct observational evidence for the validity of linear skew correction. Three aberrant stations displaying clock drift or jumps are discussed in the text. (b) Clock error estimates for OBSs that could not be time-corrected *a priori*, for lack of skew measurements. Almost all stations that clearly deviate from zero do so in a near-linear fashion, an observational confirmation of near-constant rates of clock drift. RR35 shows a clock jump, discussed in the text.

raw data with skew-corrected data, we find that the negative clock drift of RR12 was caused by a sign error: seismograms had mistakenly been corrected and archived using a skew value of +0.11 s instead of -0.11 s (the latter reported correctly by Stähler et al. (2016)). This mistake produced an induced skew of -0.22 s, which corresponds to a clock drift of -0.57 ms/day. This induced clock drift is in good agreement with our measured drift of -0.65 ms/day. The presumable origin of the clock jumps in RR07 and RR11 is discussed in Section 2.5.4. Due to the usage of 10-day CCF stacks, the date of the jumps can only be estimated within a two-day time window. Clock error corrections for RR07, RR11 and RR12 were conducted as for OBSs without skew, see discussion of Stage 2 below.

For the estimation and correction of clock errors of 24 OBSs without skews in Stage 2, the 26 OBSs with successfully verified skews from Stage 1 are regarded as reference stations. As before, we attempt to average eq. (2.3) over several station pairs in order to estimate an accurate clock drift time series for a “skewless” OBS (e.g. the pairs RR29-RR30 and RR30-RR31 for “skewless” station RR30). However, the requirement that one station in each station pair needs to be a reference station tends to cut down on the number of successful station pairs due to larger interstation distances. We exclude from the averaging process station pairs with low correlation coefficients (CC_{av} <0.35-0.4, dotted black lines in Fig. 2.1), because this tends to flatten the clock drift curve. This flattening behaviour can particularly be observed in the first iteration (Fig. 2.9). The station pairs

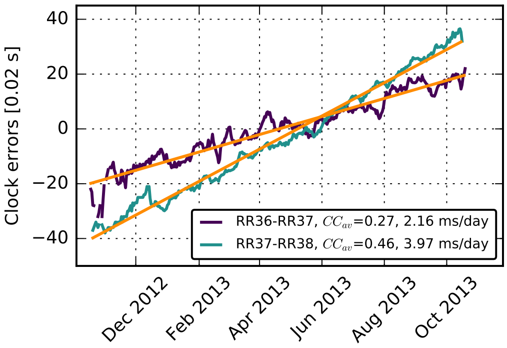


Figure 2.9: Practical complications in estimating drift rates of OBS clocks, on the example of station RR37. Plot shows estimates from the first iteration for two correlation pairs RR36-RR37 and RR37-RR38 (OBS-to-OBS correlations), and their best linear fits (orange lines) with slopes of 2.16 ms/day and 3.97 ms/day. These slopes are clearly different and yet should be identical because both RR36 and RR38 are skew-corrected (thus not contributing any drift). The explanation is that clock drift tends to be underestimated for low CC_{av} <0.35-0.4, such as here. The solution is to execute several iterations of CCF/RCF computation, daily drift estimates, and linear timing correction, as discussed in Section 4.3. Table 2.1 gives the required number of iterations for each OBS (3 for RR37).

can be included in the averaging process in later iterations if the flattening behaviour has disappeared. We suspect that this flattening effect stems from the alteration of the RCF by clock errors, with the effect that some parts of the RCF may be destructively stacked rather than constructively, such that the RCF does not resemble individual CCFs. This explanation is supported by the observation that the flattening effect is mainly visible in the first iteration. Moreover, the effect is more severe the lower the CC and the higher the drift rate.

Special consideration was required for the SWIR array, a subarray of 8 OBSs on the Southwest Indian Ridge that were spaced at much closer distances (16-42 km) than the other OBSs (see inset in Fig. 2.1a). For all 8 OBSs, the skew measurement had failed. Hence we initially estimated and corrected the clock drift of only one SWIR array station (RR47) by correlating with reference stations outside the subarray. RR47 was then used as a reference station to correlate with the remaining seven SWIR array OBSs. Note that errors made on the clock drift of RR47 (e.g. drift underestimation) would propagate to the clock drift estimation of the other SWIR OBSs.

All investigated OBSs show clock errors that clearly follow a linear trend over time (Fig. 2.8b). For most stations, the correction process required to completely remove these trends consists of several iterations of the kind described in Section 2.4.2. The required number of iterations typically increases if the clock drift is large, and/or if the CCFs have low SNRs (linked to large interstation distances and associated low CC_{av})

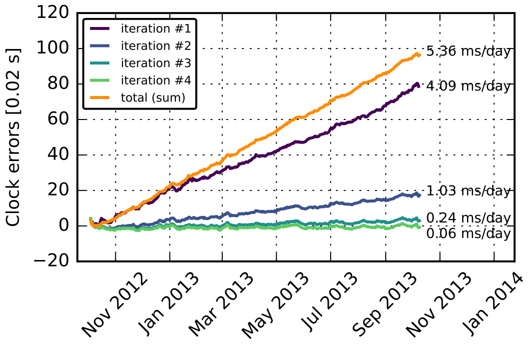


Figure 2.10: Iterative procedure for estimating OBS clock drift rate, on the example of RR13. Coloured lines are estimates of clock error over deployment time. Three iterations of linear estimation and correction are necessary. Drift rate estimates decrease in every iteration (#1: 4.09 ms/day; #2: 1.03 ms/day; #3: 0.24 ms/day) until the rate of the fourth iteration (0.06 ms/day) is indistinguishable from residual scatter, signalling convergence. Our best estimate of the actual (total) clock drift rate is 5.36 ms/day, the sum of partial rates from iterations 1-3. For better visual reference, the clock drifts of individual iterations are aligned to approximately share the same starting point as the summed drift curve.

(Gouédard et al., 2014). For instance, four iterations are required for RR47 (CC_{av}=0.44, total drift: 8.03 ms/day), whereas one iteration is sufficient for RR12 (CC_{av}=0.60, total drift: -0.65 ms/day). Fig. 2.10 visualizes this iterative process of clock drift reduction on the example of station RR13. In each iteration, we correct the absolute timing of the seismograms until the slope fitted to the clock drift measurements has dropped to less than 0.1 ms/day. This value is considered to be indistinguishable from zero clock drift, from comparisons to clock drift estimates performed on skew-corrected OBSs. Total clock drift is obtained as the sum of drift estimates from the individual iterations, as in Section 2.4.2. On the example of RR13 in Fig. 2.10, total clock drift is the sum of the drifts from the first, second, and third iterations (4.09 ms/day + 1.03 ms/day + 0.24 ms/day). Since the drift estimate of the fourth iteration (0.06 ms/day) has dropped below the threshold value, this iteration is regarded as a control iteration, which signals that the correction procedure has converged. To generate the graphical representation of drifts in Fig. 2.8, total clock drifts are determined by comparing the CCFs of the first iteration with the RCF of the last included iteration (for RR13: third iteration). In contrast to individual iterations, clock drift starts at approximately zero units because the RCF of the last iteration is almost free from contaminating effects, given that clock errors have nearly been removed prior to stacking individual CCFs into the RCF. The clock drifts in Fig. 2.8 can be fitted by linear drift rates that essentially correspond to the sum of drifts from individual iterations.

Table 2.1: Summary of clock drift estimates for all 57 RHUM-RUM OBS stations. Clock drifts are stated in units of ms/day, that is, as best-fitting slope to the temporal succession of individual clock error estimates, and also as cumulative drift in ms over a hypothetical, 365-day long deployment. σ is the estimated uncertainty on individual clock error estimates, which is not time-dependent and quantifies our method's accuracy. For example, the clock of RR06 ran fast, accumulating an error of 388.0 ms ± 8.4 ms over the course of one year. Light blue shading indicates stations affected by statistically significant clock drift. Non-shaded rows denote OBSs where drift rates are not distinguishable from zero within their σ bounds, they coincide with stations that were linearly skew-corrected prior to processing. Three exceptions, discussed in the text, are skew-corrected stations that were additionally affected by apparent clock jumps (RR07, RR11) or by a clock drift due to a sign error of the applied skew correction (RR12). In the case of clock jumps (RR07, RR11, RR35), the affected stations are indicated by dark blue shading, and two clock drift rate estimates are given: for before and after the jump. Jump occurrences can be constrained to within the two-day intervals given in brackets below the jump magnitudes. CC_{av} is the average cross-correlation coefficient of a station's CCFs from the first iteration compared to the RCF of the last included iteration. D_{av} is the average distance to neighbouring stations; N_{it} is the number of CCF iterations required to converge on the stated clock drift estimates. N_{sc} denotes the number of station and component pairs used for averaging, for example, RR25 was correlated with three stations, but for one of these stations only the H-component was usable resulting in a N_{sc} of 36 (16+16+4 component pairs) for RR25. "Skew correction" states the magnitude of an *a priori* linear timing correction, if applied. In notes, OBSs correlated with land stations are marked "land-to-OBS", whereas the default is OBS-to-OBS. "SWIR" denotes stations in a densely spaced subarray.

OBS	Clock drift rate [ms/day]	Drift per year [ms]	Uncertainty σ [ms]	Clock jump [s] (Date of jump)	CC_{av}	D_{av} [km]	N_{it}	N_{sc}	Skew correction [ms/day]	Notes
RR01	0.019	6.9	10.7	-	0.60	125	1	45	1.688	land-to-OBS
RR02	-	-	-	-	-	-	-	-	-	OBS failed
RR03	0.038	13.9	9.0	-	0.50	144	1	45	2.032	land-to-OBS
RR04	-	-	-	-	-	-	-	-	-	OBS failed
RR05	0.067	24.5	12.8	-	0.50	137	1	45	2.341	land-to-OBS
RR06	1.063	388.0	8.4	-	0.55	144	2	45	-	land-to-OBS
RR07	0.261/-0.440	95.3/-160.6	23.0	-0.999 (2013-06-07/08)	0.48	199	2	16	1.357	-
RR08	-0.071	-25.9	25.2	-	0.31	335	1	16	3.585	-
RR09	0.019	6.9	18.3	-	0.49	236	1	32	5.588	-
RR10	-0.022	-8.0	20.1	-	0.50	200	1	32	0.999	-
RR11	1.896/-0.015	692.0/-5.5	10.2	-0.990 (2012-10-21/22)	0.59	117	2	32	1.554	-
RR12	-0.652	-238.0	11.7	-	0.60	137	1	24	correct: -0.284 applied: +0.284	-
RR13	5.365	1958.2	16.3	-	0.42	197	3	32	-	-
RR14	0.098	35.8	19.5	-	0.53	168	1	12	6.041	-
RR15	-	-	-	-	-	-	-	-	-	OBS failed
RR16	-0.020	-7.3	18.1	-	0.62	257	1	32	4.111	-
RR17	0.035	12.8	12.3	-	0.64	219	1	64	4.714	-
RR18	-0.020	-7.3	11.1	-	0.47	244	1	80	0.914	-
RR19	0.085	31.0	18.4	-	0.49	208	1	80	4.324	-
RR20	0.063	23.0	20.0	-	0.46	244	1	32	1.046	-
RR21	0.021	7.7	18.0	-	0.56	230	1	32	0.689	-
RR22	-0.075	-27.4	14.9	-	0.57	190	1	76	2.287	-
RR23	1.522	555.5	24.8	-	0.50	264	2	16	-	-
RR24	3.494	1275.3	36.5	-	0.47	272	2	5	-	-
RR25	0.025	9.1	20.2	-	0.45	206	1	36	1.088	-
RR26	-0.074	-27.0	13.0	-	0.55	213	1	76	1.605	-
RR27	0.516	188.3	9.6	-	0.49	156	2	45	-	land-to-OBS
RR28	0.012	4.4	14.2	-	0.56	185	1	76	7.789	-
RR29	0.075	27.4	16.8	-	0.48	312	1	32	8.467	-
RR30	1.464	534.4	16.8	-	0.45	279	2	28	-	-
RR31	-0.192	-70.1	18.7	-	0.42	306	1	12	-2.085	-
RR32	-	-	-	-	-	-	-	-	-	no reliable result

RR33	-3.298	-1203.8	20.4	-	0.37	180	4	16	-	-
RR34	-0.098	-35.8	20.4	-	0.37	180	1	16	-3.233	-
RR35	1.878/6.367	685.5/2324.0	22.5	-0.942 (2013-04-26/27)	0.45	197	3	8	-	-
RR36	0.200	73.0	33.6	-	0.44	187	1	4	7.669	-
RR37	5.108	1864.4	36.3	-	0.48	213	3	4	-	-
RR38	-0.050	-18.3	36.3	-	0.48	213	1	4	-0.150	-
RR39	0.299	109.1	22.4	-	0.44	230	1	8	-	-
RR40	-0.056	-20.4	17.4	-	0.44	247	1	16	0.475	-
RR41	8.773	3202.1	16.3	-	0.78	42	2	16	-	SWIR
RR42	2.928	1068.7	12.9	-	0.71	32	2	4	-	SWIR
RR43	1.404	512.5	12.7	-	0.81	25	1	16	-	SWIR
RR44	0.723	263.9	21.1	-	0.84	18	1	16	-	SWIR
RR45	0.209	76.3	18.8	-	0.78	16	1	16	-	SWIR
RR46	3.902	1424.2	14.7	-	0.85	20	2	16	-	SWIR
RR47	8.026	2929.5	17.4	-	0.44	247	4	16	-	SWIR
RR48	2.572	938.8	19.2	-	0.81	23	1	16	-	SWIR
RR49	0.913	333.2	36.8	-	0.61	171	1	4	-	-
RR50	-0.083	-30.3	36.8	-	0.60	171	1	4	4.339	-
RR51	-	-	-	-	-	-	-	-	-	OBS failed
RR52	0.051	18.6	19.7	-	0.36	263	1	16	2.419	-
RR53	-6.135	-2239.3	11.0	-	0.60	93	2	32	-	-
RR54	1.746	637.3	25.0	-	0.49	203	2	8	-	-
RR55	0.827	301.8	17.8	-	0.53	169	2	12	-	-
RR56	2.594	946.8	22.3	-	0.44	287	2	20	-	-
RR57	-0.127	-46.4	43.9	-	0.46	211	1	4	3.419	-

We successfully estimate total absolute clock drifts for 23 out of 24 OBSs that lacked a skew measurement. These absolute drift rates range from 0.21 ms/day to 8.77 ms/day. For one OBS (RR32), no reliable clock error estimates could be obtained because this station's seismometer failed, its hydrophone was affected by high noise, and correlating adjacent stations RR31 and RR33 also happened to be noisy (Stähler et al., 2016).

For RR35, we detect a clock jump of 0.94 s in addition to a linear clock drift (Fig. 2.8b). This resembles the jumps at "with skew" stations RR07 and RR11, see discussion in Section 2.5.4.

Table 2.1 summarizes our estimates of total clock errors (both drifts and jumps) for all RHUM-RUM OBSs (including results from land-to-OBS correlations), together with parameters such as the number of required iterations and the average CC. The high CC for the SWIR OBSs can be attributed to their small distances to reference station RR47 (∼25 km). The timing of the OBS records affected by clock errors was corrected according to calculated linear drifts (see OBSs marked blue in Table 2.1) analogous to the procedure of the land stations described in Section 2.4.1.

2.5 Discussion

2.5.1 Method accuracy

Using cross-correlations of ambient noise, we successfully correct timing errors of broadband land seismometers as well as ocean bottom seismometers, including hydrophone channels. Per station pair, we average up to 16 component pairs and multiple station pairs, with interstation distances ranging from 10 km to 374 km. The accuracy of our method can be evaluated by the standard deviation σ of the clock errors calculated from the time corrected seismograms. Thus, we assess the method accuracy after clock correction by calculating the standard deviation of the residual clock errors. The evaluation of this residual scattering around the baseline of zero is the same accuracy criterion as calculating the standard deviation of the clock errors to the fitting function (land stations: polynomial function, OBSs: linear function), as performed in Section 2.3.3 (see Figs 2.3 and 2.4). These residual clock error variations are caused by the imperfect convergence of the CCFs to a perfectly stable function, where large interstation distances, short time windows for CCF stacking and the number of usable station and component pairs (N_{sc}) determine the limits of achievable accuracy in clock error estimates (Stehly et al., 2007; Sens-Schönfelder, 2008).

The σ estimates for clock errors of the land stations range from 11.9 ms to 22.9 ms (see Fig. 2.6), while individual component pairs show σ values of 30-114 ms. This enhancement stems from our averaging process, where we took 27 or 36 station and component pairs into account which should reduce the standard deviation statistically by $\sqrt{27}$ or $\sqrt{36}$ (see eq. 2.5).

The σ estimates for the OBSs immediately adjacent to La Réunion (which were correlated with land stations) are even smaller, ranging between 8.4 ms (RR06) and 12.8 ms (RR05)(Table 2.1). This can most likely be attributed to the use of 10-day CCF stacks in the land-to-OBS correlations, as opposed to daily CCFs in the land-to-land correlations. Additionally, we were able to average over a higher number of station and component pairs (N_{sc}=45), which clearly improves the accuracy compared to individual component pairs (σ values of 22-62 ms).

The σ estimates for clock errors of more distant OBSs (from OBS-to-OBS correlations and 10-day stacks) range between 10.2 ms and 43.9 ms, with a mean of 20.7 ms (Table 2.1). These σ values are found to depend strongly on interstation distance and on the number of station and component pairs (N_{sc} of 4-80) available for averaging, see Fig. 2.11. As for land-to-land and land-to-OBS correlations, the σ estimates are significantly reduced compared to those of the individual component pairs (11-91 ms) due

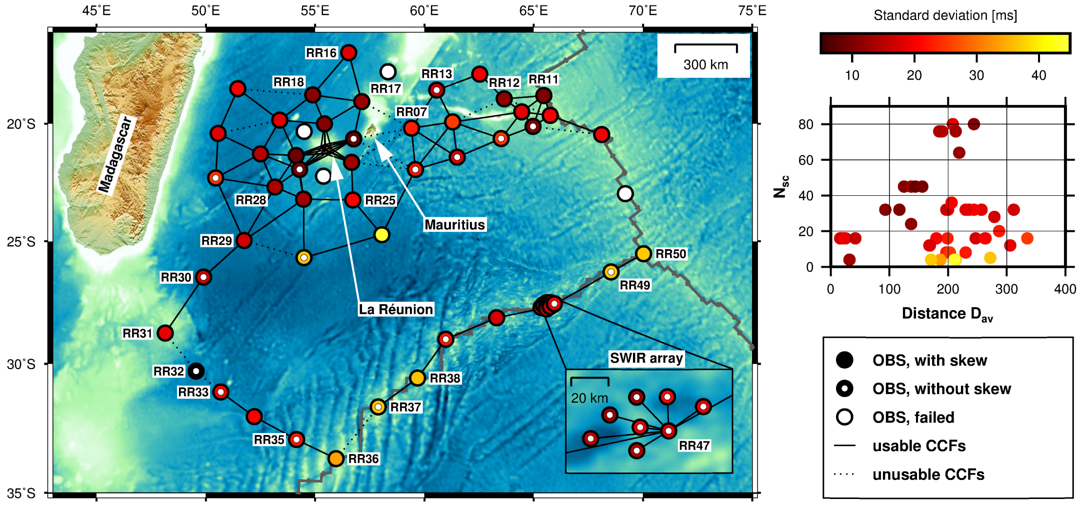


Figure 2.11: Graphical summary of the uncertainties (standard deviation σ) affecting our OBS clock error estimates. The σ values tend to be low (and estimation accuracy high) where interstation distances D_{av} are small and/or in the centre of the network (corresponding to a high N_{sc}), where many neighbouring stations are available for correlation. Right panel shows σ values (same colouring as in map) as a function of N_{sc} and D_{av}, confirming this relationship. (Solid and dotted black lines denote successful and unsuccessful correlation pairs, respectively.) RR32 is shaded black because no reliable estimates could be obtained. Figure generated with the GMT toolbox (Wessel et al., 2013).

to the performed averaging. For instance, the σ value of RR45 (σ=18.8 ms) is quite similar to the value of RR19 (σ=18.4 ms), even though the interstation distance is much shorter for RR45 (16 km) than for RR19 (208 km). However, RR19 benefitted from the higher number of station and component pairs (N_{sc}=80) available for averaging, compared to N_{sc}=16 for RR45. Fig. 2.12 visualizes the considerable benefit of averaging over multiple component and station pairs for the OBSs: σ estimates are plotted as a function of CC for the 1076 component pairs, for the 87 station pairs (averaged over the component pairs) and for the 52 stations (averaged over component and station pairs). The σ values decrease with increasing CC, and higher CCs are usually correlated with smaller interstation distances. Averaging over component pairs tremendously reduces the standard deviation (orange dots in Fig. 2.12, cf. Fig. 2.3); this effect is further enhanced by averaging over multiple station pairs (red dots in Fig. 2.12, cf. Fig. 2.4). Additional relationships between distance, SNR, CC and σ value of the 1076 component pairs are plotted in Appendix B. Note that σ values and interstation distances are quite similar for the land network and the SWIR OBS array due to a trade-off between N_{sc} and stacking duration. At the SWIR OBSs, the lower N_{sc} (that should lead to a higher σ value compared to the land stations) cancels the effect of using 10-day stacks (that should lower the σ value compared to 1-day stacks).

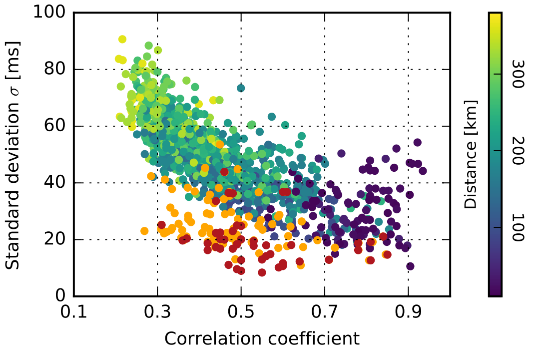


Figure 2.12: Beneficial effect of averaging on the accuracy of clock drift estimates, on the example of OBS-to-OBS correlations. Standard deviations σ are shown as a function of the correlation coefficient for 1076 component pairs, where colouring indicates distance. Orange dots denote σ values for the 87 station pairs (after averaging over component pairs). Red dots represent σ for the 52 OBSs (after averaging over component and station pairs).

Compared to the stations' sampling rates of 50 Hz, 62.5 Hz and 100 Hz, that is, sampling intervals of 20 ms, 16 ms and 10 ms, the σ uncertainties on clock drift estimates are on the order of just one sample. The clock drift estimates themselves are typically of a few milliseconds per day (Table 2.1), which means that within recording intervals of well over a month, clock drift rates can be reliably estimated. Even for shorter recording durations a clock drift estimate is possible (e.g. Fig. 2.5a, inset II: clock errors range between 70 and 80 samples). Such approximate clock drift corrections are much preferable to none at all.

2.5.2 Comparison with prior work

A comparison with the parameters and achieved accuracies of prior work on clock error estimation is given in Table 2.2. Stehly et al. (2007) reported a standard deviation σ of 47-110 ms for an interstation distance between land stations of roughly 200 km, using CCFs stacked over one month. Le et al. (2018) give a σ value of 200 ms for OBS interstation distances of 60-270 km and using 11-day stacks of the HH-component, as compared to our σ uncertainties of 11.1-43.9 ms on drift estimates derived from OBSs spaced by 200-335 km (Table 2.1). For this we stacked CCFs over only 10 days, but with averaging over as many station and component pairs as available. Sens-Schönfelder (2008) found standard deviations of 8-63 ms for land seismometer spaced by 0.2 km to 5 km and using day-long CCFs, as compared to 11.9-22.9 ms for our land stations, which were spaced by 10-43 km and also used day-long CCFs. Hence we achieved the same or better accuracy as prior work on clock drift estimates, but for larger interstation distances and/or at higher temporal resolution. This demonstrates the significant benefit of averaging clock drift estimates over multiple component and station pairs.

In Fig. 2.3, we have defined the error σ of our clock drift measurements as the standard deviation of individual daily clock drift estimates from a fitted polynomial spline of degree 3. Our study shows that for any single-component CCF, most of this error (i.e. point scatter in any of panels Figs 2.3b-j) is random noise, because it averages out to a several times smaller scatter and σ in Fig. 2.3(a), which shows the average over all nine component pairs. Although the predominantly random nature of measurement noise on single-component estimates could be suspected by prior workers (because they observed increasing σ for decreasing SNR and/or shorter stacking windows), the suppression of this noise by our averaging technique provides direct confirmation that it is indeed mainly random, as opposed to systematic noise or true clock drifting on the time scale of the stacking window. The small clock error wiggles that can be observed in some OBSs (especially in Fig. 2.7) could possibly be linked to tides, as the duration of one wiggle is roughly 1 month. This observation could merit a future study.

For OBSs, the *a priori* expectation of linear clock drift due to constant temperature at the ocean bottom is clearly confirmed by our results. This is in agreement with the findings of Gouédard et al. (2014), Hannemann et al. (2014) and Le et al. (2018), who obtained OBS clock drifts that are to first order linear. A temperature dependence is not evident for the investigated land stations. For example, station MAID shows positive clock drift rates over more than two years (February 2013 to February 2015, divided by a jump in January 2014) without any seasonal variations (Fig. 2.5d).

Table 2.2: Accuracy of clock error estimates: comparison of our method to prior studies. The most pertinent parameters are station type (land versus OBS), typical interstation spacing, and length of the time windows for CCF stacking. Uncertainty σ is the estimated, typical accuracy of clock error estimates in each study. The exact definition of σ differed somewhat across studies: for Stehly et al. (2007) and Sens-Schönfelder (2008) it is very similar to ours. Gouédard et al. (2014) and Le et al. (2018) used a bootstrapping approach that defines σ as the standard deviation of their 200 bootstrap realizations. Hannemann et al. (2014) defines σ as deviation of CCF results from measured OBS skews, that is, they take the applicability of strictly linear OBS drift for granted.

Study	Type	Distance [km]	CCF stacking window	Accuracy σ [ms]
This Study	land stations	10-43	1 day	11.9-22.9
	OBSs	16-374	10 days	8.4-43.9
Stehly et al. (2007)	land stations	~200	1 month	47-110
Le et al. (2018)	OBSs	60-270	11 days	~200
Sens-Schönfelder (2008)	land stations	0.2-5	1 day	8-63
Gouédard et al. (2014)	OBSs	~4.5	1 day	~5
	OBSs	10-15	1 month	n/a
Hannemann et al. (2014)	OBSs	<75	20 days	87-129

The use of several iterations in estimating clock error measurements was first proposed by Sens-Schönfelder (2008), in order to iteratively suppress the alteration of the RCF by clock errors. We find that this procedure is especially necessary when dealing with CCF stacking windows longer than one day, that is, for all OBSs. By contrast, the timing of our land stations can be fully corrected in one iteration (Fig. 2.6). Besides CCF stacking length, the number of required iterations is primarily determined by the magnitude of the timing errors and by the SNR of the CCFs (Gouédard et al., 2014).

We find that poor SNRs (low CC values) can lead to an underestimation of clock drift. This can be avoided by excluding station pairs with average CC below 0.35-0.4, in agreement with Sens-Schönfelder (2008), who adopted an acceptance threshold of $CC > 0.4$. The OBS (RR32) for which we could not obtain a reliable result exhibits a CC_{av} of 0.23, which is clearly below our CC threshold.

2.5.3 Consistency with skew measurements and lab experiments

For OBSs, both the magnitude and sign of estimated clock drifts are in good agreement with measured skew values, where available (Stähler et al., 2016). The majority of the RHUM-RUM OBSs were equipped with clocks that ran too fast (Table 2.1), yielding delayed waveforms.

Stähler et al. (2016) made an attempt to measure OBS clock drifts in the laboratory, on the clocks of 7 "skewless" RHUM-RUM recorders (the others were unavailable, having been re-deployed elsewhere). We find that their results are inconsistent with our drift estimates, which is very likely due to the experimental shortcoming of running the recorders at ambient lab temperatures rather than at temperatures characteristic of the ocean bottom (see Appendix C for details). Stähler et al. (2016) had similarly noted that the drift rates found in their experiments were much lower than those implied by their skew measurements, and that for the one clock (RR11) that both had a skew and ran in the lab, the two values obtained were inconsistent.

2.5.4 Apparent clock jumps - true clock failures or missing data samples?

In addition to linear clock drifts, our CCFs detect apparent clock jumps at three OBSs (RR07 and RR11 with skew; RR35 without skew, Fig. 2.8). If these jumps had been produced by actual jumps of the physical station clock, we would expect the clock error to return to zero upon resynchronization with GPS after OBS recovery (GPS

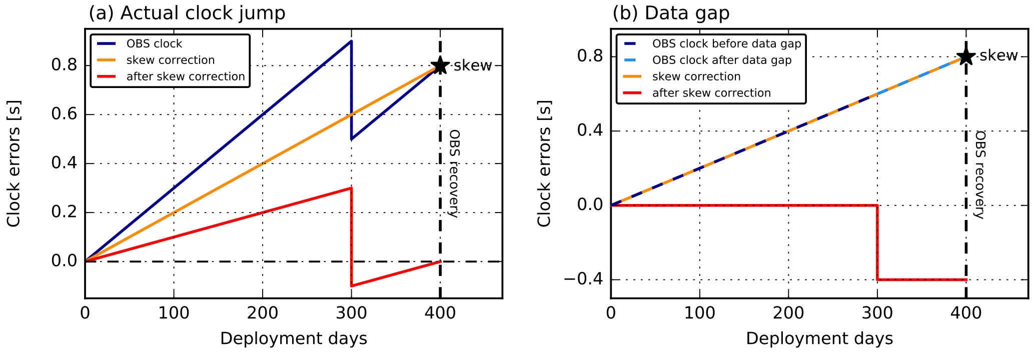


Figure 2.13: A jump in a clock error curve of an OBS can be due either to a true, physical jump of the data logger clock, or to a batch of missing samples on disk. Our method can distinguish between these two cases if a skew measurement is available (as for stations RR07 and RR11, cf. Fig. 2.8a). (a) Schematic visualization of the effect of an actual clock jump. Blue curve depicts a scenario where the recorder clock is affected by a constant drift rate, plus a clock jump 300 days into the deployment. Due to the jump, the measured skew value (star) underestimates the true clock drift rate (slopes of orange versus blue curves). Red curve shows the residual clock error after linear skew correction, which is the curve that our method would (approximately) estimate. The timing of the last sample is corrected properly, but clock drift is underestimated before and after the jump. (b) Clock error curves for the case of missing samples, where 300 days into the deployment, a batch of samples corresponding to 0.4 s recording duration was not written to disk. The physical clock is always correct, and so are the measured skew value (star) and hence the estimated drift rates before and after 300 days (dashed dark blue and light blue curves). However, a constant erroneous timing offset of -0.4 s is assigned to all samples after the data gap so that the linear skew correction (red curve) removes all clock drift, but the constant offset remains for all times after the data gap. The estimated clock drift curves of RR07 and RR11 in Fig. 2.8(a) resemble the red curve in case (b) rather than in case (a).

synchronization was successful for RR07 and RR11). The reason is that the timing of the last sample would have been corrected properly according to the skew, regardless of any clock jumps that happened earlier, as explained by Fig. 2.13(a). However, such a return to zero drift is not indicated by the CCFs. Hence there must be another cause for the apparent clock jumps. We ruled out the possibility of artificially induced jumps due to an incorrect skew correction. Instead, we suspect that the apparent clock jumps are caused by missing samples, that is, that in each case the data logger failed to write a short chunk of about 50 samples to disk. If this failure went unnoticed by the logger, a wrong timing (too early) would have been assigned to all samples after the gap. Indeed, all apparent jumps have this sign towards early times (Table 2.1). Although this problem was rare (three occurrences across the entire OBS network and deployment period), the fact that it occurred in the exact same manner at three different stations (once per station) also points to a technical weakness of the data logger model used in the German DEPAS OBSs.

If the jumps are caused by data gaps, we would expect no clock drift before and after the jump for OBSs subjected to a linear skew correction (RR07 and RR11), as visualized in Fig. 2.13(b). Indeed we observe almost no clock drift after the jump in RR11. Clock drift in RR11 before the jump is rather high (see Table 2.1) but this can be explained by the short duration of this period (12 days), which allows no representative drift estimation. For RR07, the estimated clock drift rate before the jump tends to be slightly higher than the skew-derived rate, and after the jump it tends to be slightly lower (see Table 2.1), unlike the expectation for a physical clock jump (Fig. 2.13a). The small positive drift before the jump yields a total clock error as high as that produced by the small negative drift after the jump, but of opposite sign. This indicates that the linear skew approximation was applied properly but can be refined with our method. One could argue that this refinement is just a matter of measuring clock drift rates in different windows (before and after the jump). This is true, but the measured drifts appear sufficiently evident to reflect a real refinement of the skew correction. Observational evidence for RR07 is therefore compatible with our suggested explanation of missing samples.

For RR35, the OBS lacking a skew value, we would expect identical clock drift rates before and after the jump, irrespective of its cause. This is not strictly the case in Fig. 2.8(b), but again the different clock drift rates (see Table 2.1) may well be due to the short remaining recording duration (25 days) after the jump, which prevents a proper estimate of a long-term trend. *A priori* it is unlikely that the physical clock changed its drift rate in the presumed absence of sudden, significant seafloor temperature variations (only) around the time of the jump.

An undetected data gap manifests as an apparent clock jump and can only be distinguished from a true, physical clock jump when a measured skew value is available (Fig. 2.13a versus b). This is the case for RR07 and RR11, and their drift curves in Fig. 2.8(a) clearly support a data gap. In absence of a skew measurement, we cannot totally reject the possibility that a true clock jump occurred at RR35 (red curve in Fig. 2.8b) but the nearly identical jump magnitude of 1 s at all three OBSs reinforces our inference that the jump has the same technical cause in all three OBSs: a batch of 50 samples (probably corresponding to one data buffer worth of recordings) must not have been written to disk in each instance, given the sampling rate of 50 Hz.

The good news is that clock timing corrections using our CCF method succeed regardless of the nature and cause of these apparent clock jumps, whereas the problem would go undetected and uncorrected in the standard practice of simple, linear skew interpolation. Missing samples (and/or true clock jumps) may or may not be common across the wide variety of OBS recording systems that are in use. The method demonstrated here can shed light on this question both for past and future OBS systems.

2.5.5 Rapid clock drift immediately after deployment?

We have not considered the possibility of very rapid clock drift at the very beginning of the OBS recording period (amounting to a jump within our temporal resolution). Hypothetically, this could occur due to the very rapid temperature drop during sinking to the ocean bottom. Such a time offset can be detected by comparing the causal and the acausal parts of the CCFs because waves would arrive earlier in the causal part than in the acausal part, or vice versa (Gouédard et al., 2014). The time offset is then given by the half-shift between both parts. Applied to daily or 10-day stacked CCFs, this method would have been another possibility to estimate clock errors over time. A detailed method description is given by Gouédard et al. (2014). A big disadvantage of this technique is its requirement of highly symmetric CCFs, while for our presented method the stability of (symmetric or asymmetric) CCFs over time is sufficient (Fig. 2.2). To extract a potential shift that could have occurred at the beginning of the deployment, we correlate the causal and acausal parts of RCFs derived from time corrected data for all component pairs and station pairs (see example in Fig. 2.14). However, high variability in the results even between the different channel pairs of one station pair limits the significance of this test. We observe no large shifts (>0.5 s), while the method is not suited to reliably detect real tiny clock jumps, because highly symmetric CCFs of high SNR would be required.

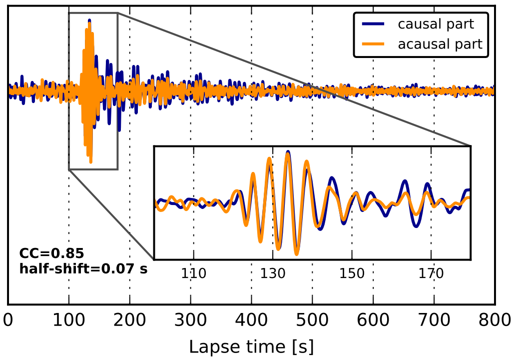


Figure 2.14: Attempt to detect a hypothesised clock jump in the first hours or days of the deployment through visual comparison of the causal and acausal (time-reversed) parts of the RCF of RR49-RR50 (OBS-to-OBS correlation, HH-component). Correlation of the two parts yields a half-shift of 0.07 s with a (high) correlation coefficient of 0.85, indicating no large time offset and hence no indication of abnormal clock behaviour, which might have resulted from the rapid temperature drop during OBS sinking to the seafloor.

The following considerations show that OBSs are unlikely to accumulate a practically significant clock drift during the first hours or days of (non-stationary) operation. Sinking from surface to ocean bottom occurred within roughly four hours, entailing a temperature change from ∼22 °C (at surface) to 4 °C (at ocean bottom). Presumably, this is accompanied by a proportional change in clock drift rate, since clock frequency is primarily controlled by temperature. In this short thermal equilibration span of hours or even a day, undetected clock drift that is significant (compared to total operation over one year) could only be incurred if drift rates at ambient and transient conditions were one or more orders of magnitude larger than at the seafloor. Lab experiments under ambient conditions showed the opposite to be the case: clock drift were observed to be less severe under ambient conditions (see Appendix C and Table 2.3).

2.6 Conclusions

The cross-correlation of ambient noise is shown to be a routinely applicable method to detect and correct even small clock errors in seismograms and hydrograms of passive experiments with deployment times of several months or more. It allows for a time-continuous monitoring, in contrast to the comparison of earthquake arrival times (Anchieta et al., 2011) or burst events of persistent localised microseismic sources (Xia et al., 2015; Xie et al., 2018a). Our estimates of clock drifts as a function of absolute time were successful and quantifiable for all stations in the RHUM-RUM array that had recorded usable data, that is, 4 broad-band island seismometers and 52 broad-band OBSs in deep water.

We show that all three seismometer components and any hydrophone channels can and should be used for the clock error measurements. Cross-correlating seismograms with hydrophones yields stable and fully usable CCFs despite their different physical nature. The same is true for correlations between land stations and OBSs, which sample vastly different crustal conditions.

For four land stations on La Réunion, interstation distances ranged between 10 km and 43 km, for which CCF stacking over short one-day windows was sufficient. The clocks were affected by complex clock drifts and jumps. During linear drift episodes, typical drift rates ranged between -2.1 ms/day and +2.9 ms/day. The uncertainties σ on these estimates were 11.9-22.9 ms and did not vary with time of year.

Near-shore OBSs were correlated with land stations, over distances of 99-174 km, requiring 10-day CCF stacks, with σ between 8.4 ms and 12.8 ms. OBS-to-OBS distances ranged between 16 km and 374 km, also requiring 10-day time averaging and yielding σ of 10.2-43.9 ms. For all OBSs, we confirm the *a priori* expectation of overwhelmingly linear clock drift, at absolute rates of 0.2-8.8 ms/day. For the 29 OBSs where a skew value had been obtained upon recovery, the implied linear drifts are consistent with our direct drift estimates.

Typical measurement uncertainties σ for the clock error at any time are thus 20 ms. This corresponds to 1-2 samples (at 50-100 Hz sampling rate); or to 0.3% of the Rayleigh-wave traveltime for station distances of typically 20 km on land; or to 0.03% of Rayleigh-wave traveltime for typical OBS-to-OBS distances of 200 km (assuming a velocity of 3 km/s).

Errors on clock drift estimates seem to be caused by imperfect convergence of the CCFs to a function that is perfectly stationary over the recording period. The method reaches its limits when interstation distances increase and/or CCF stacking time windows are shorter than desirable (due to a trade-off with the temporal resolution required to capture actual clock drifting pattern) (Stehly et al., 2007; Sens-Schönfelder, 2008). Within these limiting systematics, the accuracy of clock error estimates can be improved several-fold (typically 3-4 fold) by averaging estimates from all seismogram and hydrophone components and from multiple station pairs, where available.

This method is a powerful tool to estimate and correct clock drifts in OBSs and land stations that are beyond the reach of a GPS signal, and to verify the common linear skew correction applied to OBS data. The method can diagnose actual, physical clock drift (e.g. caused by changes in water temperature, which affect clock oscillator frequency), as well as apparent clock jumps, which seem to be due to a data logger's failure to write a batch of samples to disk.

Our method's robust and accurate performance on very heterogeneous seismogram and hydrophone data suggests additional new applications. The timing quality of existing temporary and permanent networks could be assessed and corrected, for ongoing and historical recordings. This would render many problematic data sets first usable, or better usable, for the purpose of structural imaging of mantle and crust, and is particularly pertinent to the expensive but challenging seismograms and hydrograms from OBS deployments. The (re)use of such data will provide valuable contributions to future imaging studies. A different and new imaging application will be the generation of long-range noise correlation functions for the structural imaging of crust and uppermost mantle, most importantly in ocean basins. The routine use of interstation distances exceeding 200 km could previously not be envisaged.

2.7 Appendix A: Hydrophone response change

For 12 ocean bottom stations, our analyses uncovered poor performance of hydrophones during their first few months of deployment, followed by settling into high-quality operation for the remainder of the experiment. The 12 affected stations are all German DEPAS instruments, specifically RR01, RR13, RR17, RR18, RR26, RR27, RR35, RR41, RR42, RR54, RR55, RR57. (The French INSU stations used differential pressure gauges instead of hydrophones.) Investigation showed that the instrument response of these 12 hydrophones changed markedly over time, and in each case transitioning rapidly from an initial state to a final state, several weeks or months into the deployment. Changes in instrument response causes changes to the recorded waveforms and hence CCFs. The transition event would have caused an apparent time shift in the CCFs, which is indistinguishable from a clock error in a single CCF involving a hydrophone, but is unmasked by comparison to CCFs that involve only seismometer channels.

Fig. 2.15(a) shows an example for station pair RR03-RR26. The timing of both stations has already been corrected according to their measured skews, so that we do not expect large clock errors in either station. Nevertheless, clock errors of up to -0.3 s can be observed solely in component pairs including the H-component of RR26 (1H, 2H, HH,

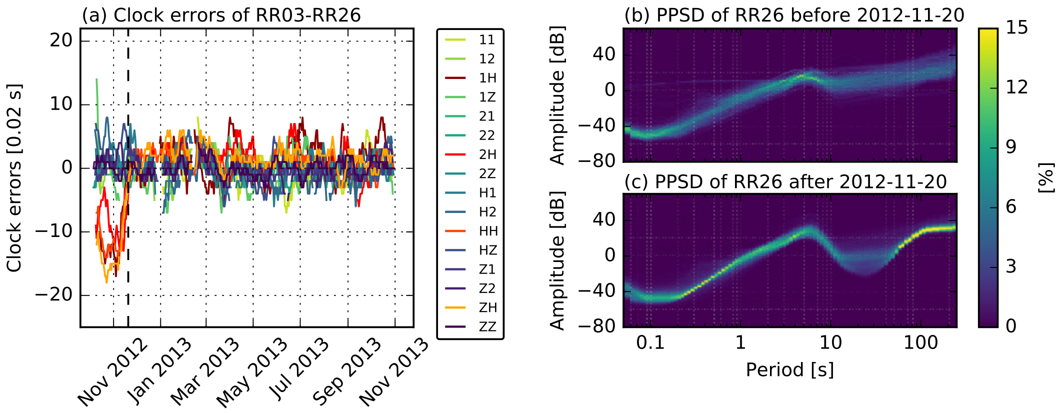


Figure 2.15: Diagnosis of hydrophone malfunctioning at some stations during the first few months of deployment, on the example of station RR26. (a) Apparent clock drift curves for the 16 individual component pairs of station pair RR03-RR26 (OBS-to-OBS correlation). Up until about 2012-11-20 (dashed vertical line), four clock error curves (in reddish line colours) are seen to deviate from all others. These are the four component pairs involving the hydrophone channel of station RR26 (1H, 2H, HH, ZH). (b) Probabilistic power spectral density plot of the H-channel at RR26 for the time period before 2012-11-20. (c) Same as (b) but for the time period after 2012-11-20. Evidently the hydrophone response changed rapidly and markedly around this date, several weeks into the deployment, for reasons that are unclear. By changing the shapes of the hydrograms and their CCFs, this behaviour translates into apparent clock drifts over time.

ZH), which indicates that the hydrophone channel is responsible for the clock deviations. Probabilistic power spectral densities (PPSD) calculated on the H-channel indeed reveal a marked spectral change around 2012-11-20 (Figs 2.15b and c), which coincides with the date around which the suspicious CCFs settled into concordant behaviour. More detailed investigation of the hydrophone PPSDs showed that the change in the instrument response (in the period range between ∼7 s and ∼100 s) occurred gradually over a few days, starting on 2012-11-15 and reaching their final state on 2012-11-20. It is not surprising that the CCFs are sensitive to this noise level change, given that they are computed in the period sub-band of 2-20 s. For example, Zhan et al. (2013) showed that a sudden change in frequency content can cause a temporal dilation of CCFs.

We made similar findings for the other 11 OBSs, where the problematic operation interval ranged from one to three months after deployment. The cause for the initially different instrument responses remains unclear, as does the reason or occasion for their settling into normal operation. Seasonal variations of the noise intensity due to storms (Davy et al., 2015) seem to play no role because the issue is limited to 12 out of 52 stations, does not affect the seismometer components, and the dates of the noise transitions differ across affected stations. Moreover, seasonal variations would mainly affect secondary microseisms (∼7 s), whereas the hydrophone noise changes also occur at significantly longer periods. From the temporal evolution of the noise, we speculate that certain internal hydrophone components took excessively long to settle into normal working mode. The hydrophones consist of a piezo element that accumulates charge on a capacitor that might need a few hours or even a day to discharge completely, a requirement for proper functioning. Still this process should not take several months. Several hydrophones are of an older but otherwise identical model (HTI-01-PCA instead of HTI-04-PCA in RR03, RR04, RR06, RR11, RR13, RR18, RR20, RR23, RR24, RR35, RR41, RR53). The older model is slightly over-represented among the faulty ones, but not by much. Hannemann et al. (2014) detected also hydrophone malfunctioning at 3 DEPAS OBSs in the beginning of their operation period, which could be associated with small waveform amplitudes that tend to recover over time. Similarly, we found in some of our erroneous stations that the hydrophone malfunctioning seems to coincide with an amplitude reduction. However, a general correlation with reduced amplitude and stepwise recovery as described by Hannemann et al., 2014 cannot be clearly confirmed. Hence further investigation would be needed to clarify the causes of these initially faulty instrument responses in a subset of the German hydrophones.

2.8 Appendix B: Relationships between SNR, CC, distance and standard deviation σ

For OBS-to-OBS correlations, we are dealing with a large data set comprising 87 station pairs with interstation distances ranging from 16 km to 374 km. For each station pair, we are able to use up to 16 component pairs, yielding a total of 1076 station and component pairs, an unprecedented large data set for a clock error study. Fig. 2.16 explores the relationships between the four most significant parameters characterising each correlation pair: interstation distance, SNR, CC and standard deviation σ. SNR and CC values are determined as averages over the individual values of the 10-day stacks. The σ estimates, our measure of estimation accuracy, are calculated from the scatter of the residual clock errors around zero.

Fig. 2.16(a) shows that high SNRs are associated with high CCs. The relationship can be approximately quantified by fitting a logarithmic curve (orange curve):

$$CC = 0.32 + 0.15 \cdot \ln(SNR - 3.48). \tag{2.6}$$

The colouring indicates that high SNRs and high CCs tend to be achieved for small interstation distances.

Fig. 2.16(b) shows CC values as a function of interstation distance. For each station pair, the CC values of up to 16 component pairs are plotted in a column-like way. Again, high CCs are seen to coincide with small interstation distances. The σ values are lower (and accuracy is higher) for shorter interstation distances. For a given distance, σ estimates tend to be lower for higher CC values. The relationship between σ and CC is presented in Fig. 2.12. Fig. 2.16(c) gives the standard deviation σ as a function of distance with CC values represented by the colouring. This plot supports our findings from Fig. 2.16(b). Both Figs 2.16(b) and (c) indicate a quasi-linear relationship between CC and interstation distance, as well as between standard deviation σ and interstation distance.

2.9 Appendix C: Comparison with laboratory experiments

For OBSs without skew measurements, Stähler et al. (2016) attempted to obtain clock drift estimates by re-running their data loggers in the laboratory. Short of specialized equipment, these tests were run under ambient pressure and temperature conditions.

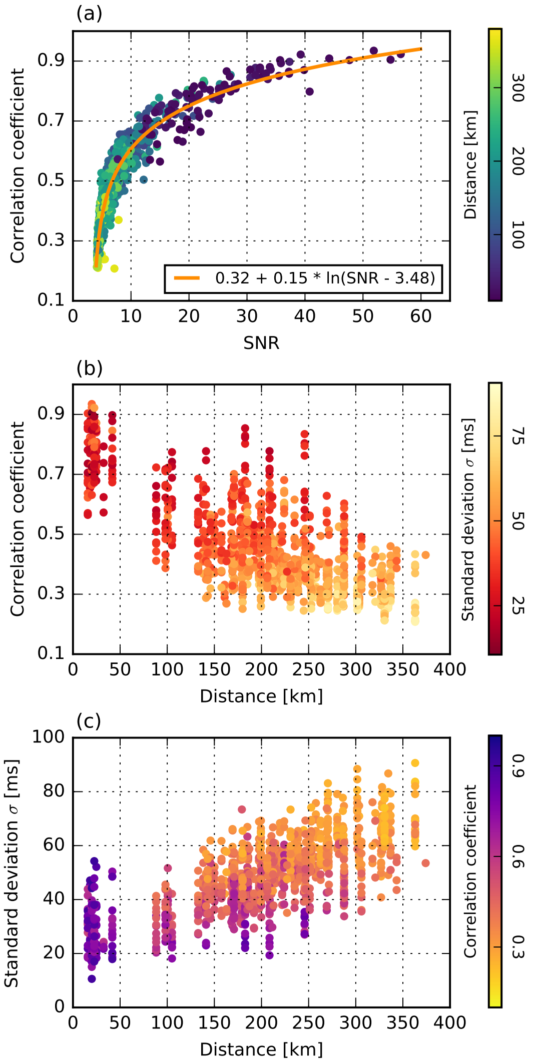


Figure 2.16: Relationships between SNR, CC, distance, and standard deviation σ of the 1076 component pairs from OBS-to-OBS correlations. For details see Appendix B. The relationship between σ value and correlation coefficient is shown in Fig. 2.12.

They ran two tests of 7 days and 33 days duration, on the seven afflicted stations that had not already been redeployed at the time (RR06, RR11, RR41, RR43, RR44, RR45, RR55). The measured skew values were linearly extrapolated to derive a hypothetical skew value for a 365-day long experiment. For direct comparison, we also calculated hypothetical cumulative clock errors after 365 days of deployment from our estimated drift rates (Table 2.1) for these seven OBSs. The values are given in Table 2.3, together with the results of the two lab runs.

Stähler et al. (2016) doubted that their laboratory results yielded reasonable approximations of clock drift during the actual ocean bottom deployment. First, their skew values for station RR44 differed by >1 s between the two runs. Second, the lab-predicted skew

Table 2.3: Comparison of our CCF-derived clock drift estimates of Table 2.1 with direct laboratory measurements of clock skews, as reported by Stähler et al. (2016). Seven of the RHUM-RUM OBS data loggers were put through two test runs of 7 days and 33 days duration, under ambient temperature and pressure conditions at the facilities of the DEPAS OBS instrument pool at Bremerhaven. For both methods, the estimated/measured clock drift rates are linearly extrapolated to a hypothetical deployment duration of 365 days. The direct skew measurements are inconsistent with our results, presumably because the test runs were not conducted at the low temperatures that define deep sea conditions.

OBS	This study	Lab run (7 days)	Lab run (33 days)
RR06	0.39 ± 0.008 s	0.15 s	0.13 s
RR11	0.57 ± 0.010 s	-0.15 s	-0.21 s
RR41	3.20 ± 0.016 s	0.30 s	0.23 s
RR43	0.51 ± 0.013 s	0.00 s	0.033 s
RR44	0.26 ± 0.021 s	-0.50 s	0.55 s
RR45	0.08 ± 0.019 s	0.045 s	-0.05 s
RR55	0.30 ± 0.018 s	0.0015 s	-0.03 s

for RR11 differed by ∼0.8 s from the measured skew. RR11 was the only station among the seven for which a skew had been obtained upon recovery. Our own, noise-based drift estimate for RR11 is consistent with the skew measurement obtained aboard the ship. Third, the skews of both lab runs are generally small in magnitude, compared to skews measured on the ship (for stations other than the seven in question). Consistent with this, our noise-based drift estimates are larger in magnitude than those measured in the lab (except for RR44).

Hence we concur with Stähler et al. (2016) that reliable OBS clock drift estimates cannot be gained from laboratory experiments without simulating at least realistic temperature conditions on the ocean bottom, of relatively constant 4 °C. (Our earlier findings of almost linear OBS clock drifts throughout the deployment indicate that temperature transients upon deployment (sudden decrease from ∼22 °C to 4 °C) and recovery (sudden increase to ∼22 °C) are too rapid to accumulate a noticeable clock error.

Chapter 3

Crustal tomography of the western Indian Ocean

This chapter will be submitted in March 2019 to *Geophysical Journal International* under the title *Crustal tomography of the western Indian Ocean from noise cross-correlations of land and ocean bottom seismometers* (Hable et al., 2019a).

Abstract

We use seismic noise cross-correlations to obtain a 3-D tomography model of S-wave velocities beneath the western Indian Ocean, in the depth range of the oceanic crust and uppermost mantle. The study area covers 2000×2000 km^2 between Madagascar and the three spreading ridges of the Indian Ocean, centred on the volcanic hotspot of La Réunion. We use seismograms from 38 ocean bottom seismometers (OBSs) deployed by the RHUM-RUM project and 10 island stations on La Réunion, Madagascar, Mauritius, Rodrigues, and Tromelin. Phase cross-correlations are calculated for 1119 OBS-to-OBS, land-to-OBS, and land-to-land station pairs, and a phase-weighted stacking algorithm yields robust group velocity measurements in the period range of 3-50 s. We demonstrate that OBS correlations across large interstation distances of >2000 km are of sufficiently high quality for large-scale tomography of ocean basins. Many OBSs yielded similarly good group velocity measurements as land stations. Besides Rayleigh waves, the noise correlations contain a low-velocity wave type propagating at 0.8-1.5 km/s over distances exceeding 1000 km, presumably Scholte waves travelling through seafloor sediments. The 100 highest-quality group velocity curves are selected for tomographic inversion at crustal depths. The inversion is executed jointly with a data set of longer-period, Rayleigh-wave phase and group velocity measurements from earthquakes, which had previously yielded a 3-D model of Indian Ocean lithosphere and asthenosphere. Robust resolution tests and plausible structural findings in the upper 30 km validate the use of noise-derived OBS correlations for adding crustal structure to earthquake-derived tomography of the oceanic mantle. Relative to crustal reference model CRUST1.0, our new shear-velocity model tends to enhance both slow and fast anomalies. It reveals slow anomalies at 20 km depth beneath La Réunion, Mauritius, Rodrigues Ridge, Madagascar Rise, and beneath the Central Indian spreading ridge. These structures are clearly associated with increased crustal thickness and/or volcanic activity. Portions of Madagascar Ridge crust are also slower than in the reference model. Thickened crust under La Réunion and Mauritius is matched by particularly fast and thick lithosphere, presumably reflecting magmatic depletion and dehydration of the deep source regions from which the crustal melts were extracted by hotspot activity. Our tomography model is available as electronic supplement.

3.1 Introduction

Cross-correlations of ambient seismic noise have become widely used to constrain crustal structure. The method is based on the principle that cross-correlation functions (CCFs) of seismic noise between two seismometers converge towards the Green's function of these stations, as shown by e.g. Shapiro & Campillo (2004) and Shapiro et al. (2005). Since the Green's function reflects wave propagation (and therefore subsurface structures) between the two stations, CCFs can be used for tomographic studies. As microseismic noise is mainly composed of surface waves with wave periods of 3-20 s (e.g. Friedrich et al., 1998), noise cross-correlations in this period band put constraints mainly on the earth's crust and uppermost mantle. Over the past 15 years, the number of noise tomography studies has exploded, ranging from local scale (e.g. Brenguier et al., 2007; Mordret et al., 2014, 2015), to regional and continental scale (e.g. Guo et al., 2013; Zigone et al., 2015; Goutorbe et al., 2015; Corela et al., 2017; Lu et al., 2018; Xie et al., 2018b) and up to global scale (e.g. Nishida et al., 2009; Haned et al., 2016). The majority of these studies used cross-correlations between land stations to investigate continental crust and/or mantle. For oceanic plates, structural details far away from continents are known only for a few areas because long-term seismological broadband arrays have rarely been deployed in the open oceans. Applied to relatively recent ocean bottom seismometer (OBS) arrays, the method of noise cross-correlations has the potential to increase our knowledge about ocean basins significantly. Only a few studies have used OBSs for noise tomography (e.g. Yao et al., 2011; Mordret et al., 2014; Zha et al., 2014; Ball et al., 2016; Corela et al., 2017; Ryberg et al., 2017; Tomar et al., 2018), and the interstation distances were relatively small, ranging from several metres (e.g. Mordret et al., 2014; Tomar et al., 2018) to at most hundreds of kilometres (e.g. Yao et al., 2011; Zha et al., 2014; Ball et al., 2016; Corela et al., 2017; Ryberg et al., 2017).

Here, we present an OBS noise tomography study of the western Indian Ocean with very large interstation distances of up to >2000 km. Due to very sparse instrumentation of this ocean, the few crustal tomographies that exist have low lateral resolution (e.g. Montagner & Jobert, 1988; Debayle & Lévêque, 1997; Ma & Dalton, 2017).

From 2011 to 2015, the Réunion Hotspot and Upper Mantle - Réunions Unterer Mantel (RHUM-RUM) experiment was conducted in the western Indian Ocean around the volcanic island of La Réunion, which is located 800 km east of Madagascar and 200 km southwest of Mauritius (Fig. 3.1a). The hotspot volcanism of La Réunion since 65 Ma, combined with a time-progressive hotspot track that leads to a large igneous province (the Deccan Traps in India and the Seychelles) point both to a deep mantle plume beneath the island (Barruol & Sigloch, 2013). RHUM-RUM's main objective was seismological imaging of the area from crust to core, in order to strengthen or reject the mantle plume hypothesis. A large array of 57 broadband OBSs was deployed over an area of 2000×2000 km^2 during 13 months in 2012-2013, and 37 temporary land stations were installed between 2011 and 2015 on La Réunion, Mauritius, Rodrigues Island, the Îles Éparses, southern Seychelles, and in southern Madagascar.

Mazzullo et al. (2017) used earthquake data recorded by the RHUM-RUM stations to retrieve Rayleigh-wave phase velocities for periods of 30-300 s and group velocities for 16-250 s. A three-dimensional, regional shear-velocity model was obtained by joint inversion of phase and group velocities, but the wave periods were too long to reliably constrain crustal depths, which were essentially filled in *a priori* by a crustal reference model, CRUST1.0 (Laske et al., 2013).

The scope of our study is to tomographically image the crust beneath the RHUM-RUM deployment, using cross-correlations of ambient seismic noise recorded by RHUM-RUM OBSs and island stations. We use phase cross-correlations (PCCs, Schimmel, 1999) which do not require special pre-processing, unlike classical CCF calculations. Subsurface structure is constrained by group velocity measurements, which are derived from the PCCs. Finally, our group velocity values are inverted jointly with the earthquake-generated Rayleigh-wave data previously measured and inverted by Mazzullo et al. (2017). We obtain a 3-D, regional shear-velocity model from the surface down to ~250 km depth, of which we discuss the upper 80 km, where the new ambient noise data appreciably add constraints.

Section 3.2 describes the seismic stations used in this study. Section 3.3 describes the calculation of the PCCs, the stacking procedure, and the estimation of group velocity curves from the PCC stacks. The results of the group velocity measurements are presented in Section 3.4, and Section 3.5 focuses on tomography based on these noise-derived group velocity data and on the earthquake-derived data of Mazzullo et al. (2017). The results are discussed in Section 3.6, followed by Conclusions in Section 3.7.

We used the open-source software toolboxes Python (Rossum, 1995), ObsPy (Beyreuther et al., 2010; Megies et al., 2011; Krischer et al., 2015), and obspyDMT (Hosseini & Sigloch, 2017) for data download and processing, and for plotting figures. The exceptions are Fig. 3.1, which was generated with the open-source mapping toolbox GMT (Wessel et al., 2013), and Figs 3.8, 3.9, and 3.10, which were created with Mathematica (Wolfram Research Inc., 2018).

3.2 Data description

The study area is shown in Fig. 3.1. We use continuous waveform data of 57 OBSs that were deployed in the course of the RHUM-RUM experiment around La Réunion for 13 months (October 2012 to November 2013) at water depths of 2200-5400 m (Barruol et al., 2017). The network comprised 44 wideband OBSs from the German DEPAS pool, 4 wideband OBSs from GEOMAR Kiel (identical in construction to the DEPAS OBSs), and 9 broadband OBSs from the French INSU pool. The DEPAS and GEOMAR OBSs were equipped with the OBS version of the Güralp CMG-40T sensor, named CMG-40T-OBS. Its corner period is usually 60 s, although 9 of the DEPAS OBSs were equipped with a prototype sensor of corner period 120 s (but these instruments failed to deliver data). The INSU OBSs were equipped with Nanometrics Trillium 240 sensors, featuring a corner period of 240 s. INSU OBSs are of the LCPO2000-BBOBS type, based on the Scripps Institution of Oceanography "L-CHEAPO" design. More details about OBS types and their performances during the RHUM-RUM deployment can be found in Stähler et al. (2016).

Our study is based on vertical component seismograms. Unfortunately, the seismometers of several OBS stations failed, mostly due to a stuck levelling mechanisms, which affected all nine 120 s-DEPAS instruments and a few others (Stähler et al., 2016). We included only one station (RR47) of a densely spaced sub-array of 8 OBSs on the Southwest Indian Ridge ("SWIR array" in Fig. 3.1), since ours was a large-distance study. In total, we could make use of 38 RHUM-RUM OBSs (26 DEPAS, 3 GEOMAR, 9 INSU), marked by red symbols in Fig. 3.1(a).

In order to enhance the wave path coverage of our study area, we use 10 land stations in addition to the 38 OBSs: three stations on La Réunion (MAID, PRO, RER, see Fig. 3.1b), four on Madagascar (LAHA, ANLA, RUM1, FOMA), and one station on each of the islands of Mauritius (MRIV), Rodrigues (ROCAM), and Tromelin (TROM). MAID, RUM1 and TROM were part of the temporary RHUM-RUM network 2011-2015 (Barruol et al., 2017); LAHA and ANLA were part of the temporary MACOMO experiment on Madagascar (Wysession et al., 2011); FOMA, RER and ROCAM are permanent GEOSCOPE stations (IPGP & EOST, 1982); PRO is a permanent station of Observatoire Volcanologique du Piton de la Fournaise (OVPF); and MRIV is a permanent station of the Mauritius Seismic Network. Each of these land stations operated for at least five months during the 13-month period of the OBS deployment.

A crucial requirement for our study is the correct timing of the seismic records. Unlike for land stations, a GPS signal acting as external reference clock cannot be received at the ocean bottom. Instead, the timing of OBS records has to rely on the OBS-internal clock, which is however prone to drift. The conventional procedure is to perform a plausible, but unverified timing correction after OBS recovery, the so-called skew correction: the timing of the seismograms is corrected linearly based on the skew, which is the measured time difference between OBS clock and GPS clock immediately after the recovery of the OBS. A GPS synchronization immediately before the OBS deployment ensures a correct timing at the beginning of the recordings. This procedure became problematic in the RHUM-RUM experiment, where almost half of the OBSs lacked the (second) GPS synchronization after recovery, due to faulty shut-down of the OBS clocks prior to OBS recovery (Stähler et al., 2016). Therefore, Hable et al. (2018) conducted an extensive clock error study to estimate and correct clock drifts for OBSs without skew, and to verify commonly assumed (linear) clock drift patterns used to justify the skew correction procedure. We demonstrated that noise cross-correlations can be used to measure clock drifts for large interstation distances (>300 km) with a high temporal resolution of 2 days. We demonstrated that averaging over multiple station and component pairs improves the method's timing accuracy severalfold, to $\sim$20 ms. We verified that the RHUM-RUM OBS clocks indeed drifted almost linearly. For all OBSs (except one very noisy station not included in the present study), a linear clock drift value could be measured from the correlation functions (or verified, if a skew measurement existed).

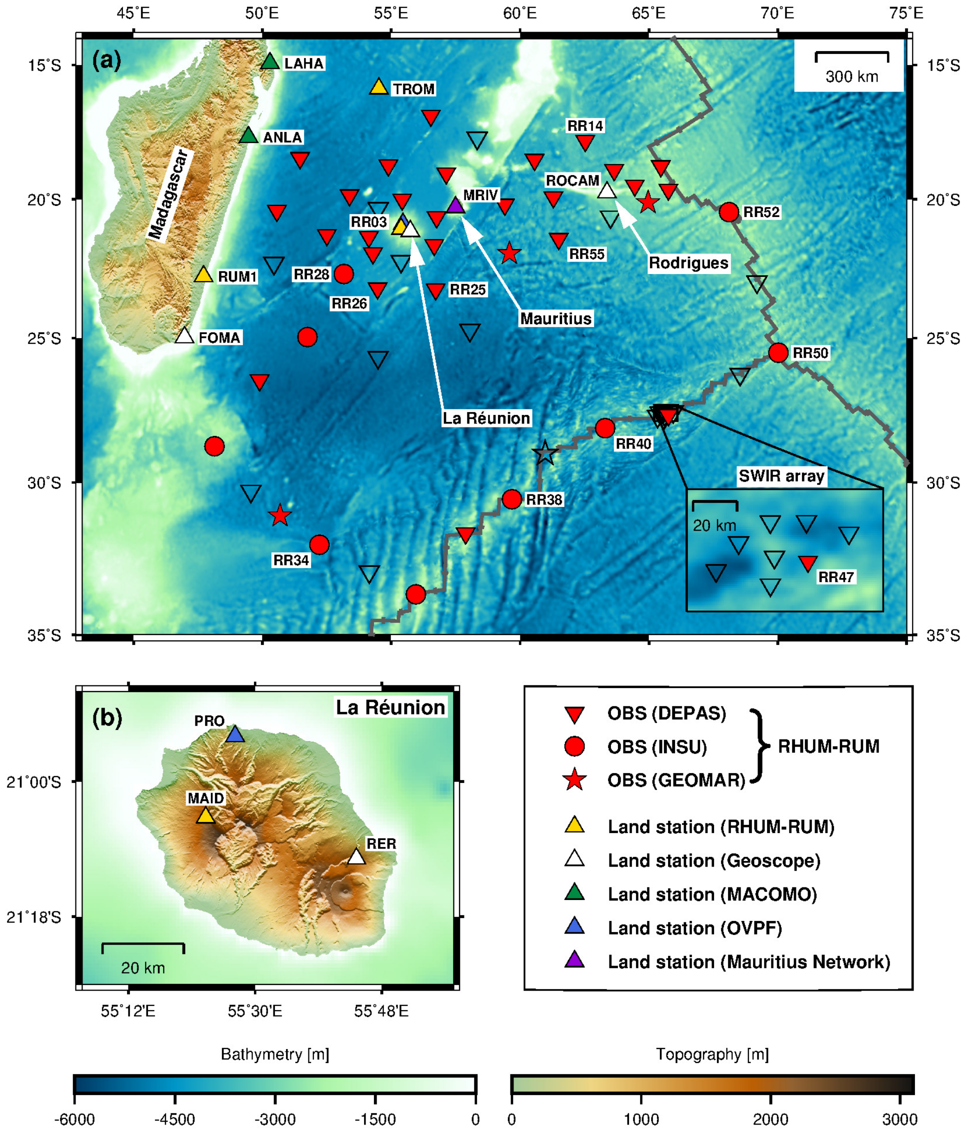


Figure 3.1: Maps of the study area in the western Indian Ocean, showing bathymetry, topography, plate boundaries and seismometer deployments. (a) 57 ocean-bottom seismometers installed by the RHUM-RUM experiment (Barruol et al., 2017) are shown as red symbols (stations that worked) or unfilled symbols (stations that failed). Different red symbols mark different OBS types (DEPAS, INSU or GEOMAR), which are discussed in the text. Unfilled stations in a local sub-array ("SWIR array" in the inset map) worked but were too closely spaced to be useful. Upright triangles mark land stations that were deployed contemporaneously by RHUM-RUM (fill colour yellow); by the permanent GEOSCOPE network (white) (IPGP & EOST, 1982); by the temporary MACOMO experiment (green) (Wysession et al., 2011); by the Observatoire Volcanologique du Piton de la Fournaise (blue), and by the Mauritius Seismic Network (purple). Stations are labelled by their names if they are explicitly mentioned in the text. Grey lines indicate the area's three mid-ocean spreading ridges: Central Indian Ridge (striking N-S), Southeast Indian Ridge (NW-SE) and Southwest Indian Ridge (SW-NE) plate boundaries. (b) Map zoom shows the island of La Réunion and three land stations that are used.

Drift magnitudes ranged between 0.2 ms/day and 8.8 ms/day. Additionally, we could detect that three OBSs were affected by apparent clock jumps of ∼1 s, caused by (very rare) failure to successfully write batches of digital samples to disk.

Land stations should not be affected by clock errors, because frequent synchronization with a GPS signal can usually be ensured on land. An exception was station MAID on La Réunion, which suffered from large clock errors of several minutes due to a data logger failure. The timing problem of MAID (and three other land stations not part of our study) was also detected and successfully corrected by Hable et al. (2018).

For this study, we used time-corrected data for all OBSs and for land station MAID according to Hable et al. (2018).

3.3 Methodology

3.3.1 PCC calculation

The subsurface properties between two points are reflected by the Green's function. This function represents the impulse response, i.e. the seismogram measured at one point if a delta pulse acts as input at the other point. Since cross-correlation functions of ambient seismic noise recorded at two seismometers (s_1 and s_2) converge towards the Green's function between the two stations, CCFs are a powerful tool to investigate subsurface structure, and more specifically crustal structure, because the periods that dominate in ambient microseismic noise correspond to Rayleigh waves propagating at crustal depths (e.g. Shapiro & Campillo, 2004; Sabra et al., 2005b; Shapiro et al., 2005). In practice, CCFs are calculated on relatively short time series (e.g. one day), and are subsequently stacked over a longer duration (e.g. one year) in order to achieve the best possible convergence towards the Green's function (Bensen et al., 2007; Schimmel et al., 2011). The stacking procedure enhances the signal part common to all daily CCFs, and suppresses the random noise components.

CCFs comprise a causal part and an acausal part. The causal part represents seismic waves recorded at station s_2 as a response to an impulse at the location of s_1. Conversely, the acausal part represents waves propagating from station s_2 to station s_1. Ideally, both sides of a CCF should be symmetric because the waves travel through the same medium. In practice, CCFs are usually not symmetric because unevenly distributed noise sources lead to different amounts of wave energies travelling in either direction (Stehly et al., 2007). Since CCF asymmetry is mainly limited to unequal amplitudes of the causal and acausal parts, stacking over both CCF parts is an appropriate technique to enhance the signal-to-noise ratio of the CCF stack.

The classical method of CCF requires pre-processing to reduce the effect of high-energy, non-random signals like earthquakes, which could deteriorate the CCFs (Bensen et al., 2007). Here we use cross-correlations based on the alignment of the phases in seismograms from two stations, termed phase cross correlations by Schimmel (1999). Schimmel et al. (2011) showed that these PCCs can improve convergence to the Green's function. PCCs are insensitive to amplitudes, in contrast to classical CCFs which are based on the alignment of high amplitudes (Schimmel et al., 2011). Hence no time-domain normalization (such as a 1-bit normalization) described by Bensen et al. (2007) is required for the PCC pre-processing. A detailed comparison of PCCs with classical CCFs is given by Schimmel et al. (2011).

The PCC definition has its roots in analytical signal theory. The analytic signal $s(t)$ of a real time series $u(t)$ can be described in the complex number domain by following equation (Schimmel et al., 2011):

$$s(t) = u(t) + iH(u(t)) = a(t) \cdot e^{i\phi(t)}, \tag{3.1}$$

where $H(u(t))$ denotes the Hilbert transform of $u(t)$; $a(t)$ represents the envelope and $\phi(t)$ the instantaneous phase of the waveform. Based on this representation, the following definition of PCC calculations was introduced by Schimmel (1999) and further investigated by Schimmel et al. (2011):

$$PCC(t) = \frac{1}{2T} \sum_{\tau=\tau_0}^{\tau_0+T} \left(\left|e^{i\phi_1(t+\tau)} + e^{i\phi_2(\tau)}\right|^{\nu} - \left|e^{i\phi_1(t+\tau)} - e^{i\phi_2(\tau)}\right|^{\nu} \right), \tag{3.2}$$

where ϕ_1 and ϕ_2 are the instantaneous phases of two stations s_1 and s_2; t is the lapse time of the PCC; τ_0 is the start time of the PCC; and T is the length of its correlation window. The sensitivity of the PCC is determined by ν; we follow Schimmel et al. (2011) and use ν=1.

Our study uses the Z-component of day-long seismograms. The pre-processing consists of only three steps: (1) downsampling of the time series to 2 Hz, from originally 50 Hz, 62.5 Hz or 100 Hz; (2) correction for the instrument response; and (3) removal of the mean and linear trend. Downsampling renders the PCC calculation less time-consuming, and instrument correction is particularly important because we are cross-correlating data from different seismometer types (Stähler et al., 2016).

Before the PCCs are calculated, we split each pre-processed, day-long time series into four 6-hour long traces in order to increase the number of PCCs available for stacking. Next the traces are filtered by three zerophase bandpass filters of periods of 3-10 s, 10-20 s, and 20-50 s. Use of a single, widebanded filter (e.g. 3-50 s) is not recommended because this tends to hamper robust group velocity estimation, as shown by Haned et al. (2016). We demonstrate in Section 3.3.3 that group velocity estimation can indeed be improved by using the three narrower sub-bands.

We noticed that PCCs correlating two INSU OBSs were afflicted by almost monochromatic frequency artefacts at multiples of 0.05 Hz (0.05 Hz, 0.1 Hz, and 0.15 Hz), i.e. a “ringing” at these frequencies across all lapse times of the PCC can be observed. This instrument problem necessitated the selective removal of very narrow frequency bands around these three frequencies in all INSU OBS seismograms (although only specific stations might actually have been affected). This was successfully achieved by applying three narrow bandstop filters centred on 0.05 Hz, 0.1 Hz, and 0.15 Hz; the outcome is visualized in Fig. 3.2. Although the problem affected only INSU-INSU correlations, it could not be observed at all INSU-INSU station pairs, probably because only the individual stations RR38, RR40, RR50, and RR52 were affected. Figs 3.2(a)-(c) analyses a ten-day long record of RR38. The spectrogram in Fig. 3.2(b) reveals two monochromatic peaks at 0.05 Hz and 0.1 Hz as horizontal lines throughout the recording period. Expected features are the two diffuse, high-energy bands centred on 14 s and 7 s, which correspond to the primary and secondary microseismic noise bands, respectively. In order to diagnose further, we calculated the power spectrum of each of the 10 days and performed an average over these 10 power spectra in Fig. 3.2(c). The frequency artefacts appear as sharp peaks at 0.05 Hz and 0.1 Hz in the averaged daily power spectrum. Investigation of other INSU stations identified 0.15 Hz as a third potentially contaminating frequency. Their origin remains unclear, but they are almost certainly instrument artefacts due to their unnatural, narrow-banded nature. Since the PCC definition is based on phase coherence, phase correlations of two records affected by the same persistent frequencies are dominated by exactly these frequencies. This is shown in Fig. 3.2(d) for INSU-INSU station pair RR38-RR50. After removal of the frequency artefacts by bandstop filtering, Rayleigh-wave arrivals emerge clearly in the PCC stack (Fig. 3.2e), enabling a robust group velocity measurement (Fig. 3.2f).

For all station pairs (OBS-to-OBS, OBS-to-land, land-to-land), we calculate PCCs of 6-hour long traces according to eq. (3.2), with lapse times running from -1000 s to +1000 s. It is empirically known that proper group velocity estimation requires interstation distances of at least two or three wavelengths of the longest occurring wave period (e.g. Haned et al., 2016). Hence we skip the computation of PCCs in the 20-50 s

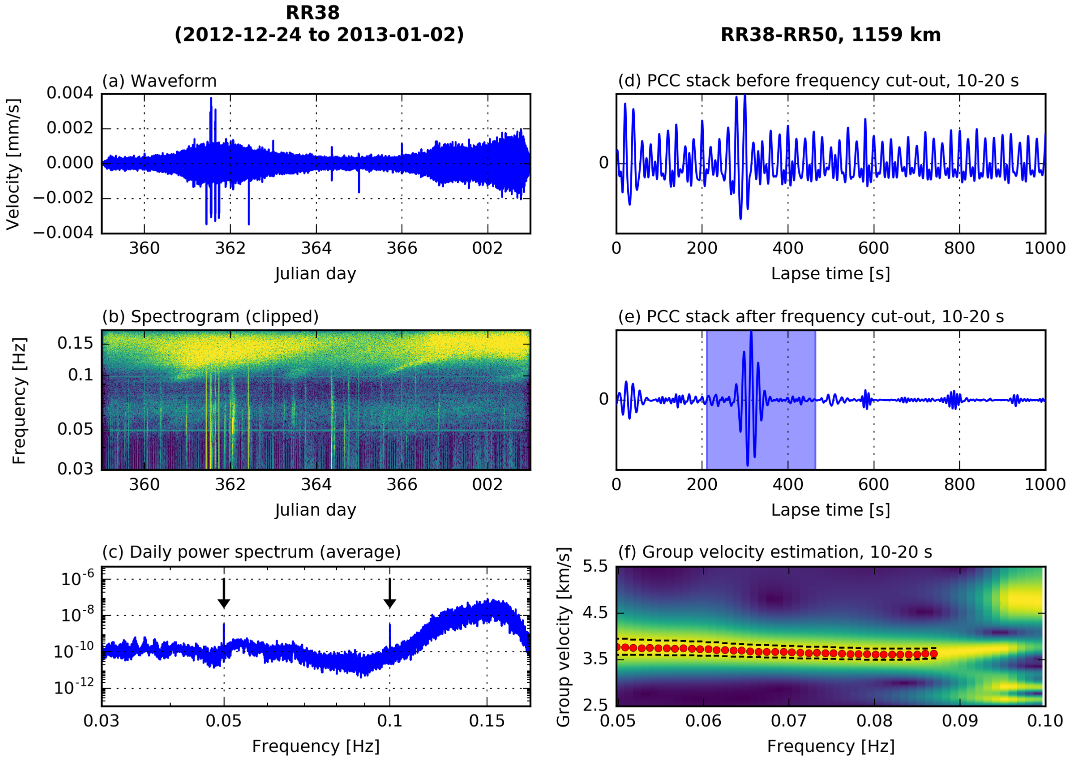


Figure 3.2: Visualization of measurement procedures. This includes the removal of a harmonic frequency artefact, which occurred in one type of OBS that was of the highest quality otherwise. At least four stations were affected, including RR38, for which data are shown here. (a) Continuous raw waveform data for 10 days (2012-12-24 to 2013-01-02). (b) Spectrogram for the same time series. Monochromatic frequency artefacts affecting RR38 stand out as horizontal lines at 0.05 Hz and 0.1 Hz. (c) Power density spectrum obtained by subdividing the time series into 10 day-long segments and averaging the 10 individual power spectra. Monochromatic frequency artefacts are visible as vertical peaks at 0.05 Hz and 0.1 Hz (see arrows). (d) Phase cross-correlation (PCC) stack of two INSU stations RR38 and RR50 (interstation distance 1159 km). Both stations were affected by the artefact, and hence the PPC is dominated by oscillations of 20 s and 10 s period. (e) The same PCC as in (d), but after removal of the artefact by prior bandstop-filtering of both time series around 0.05 Hz and 0.01 Hz. Rayleigh-wave arrivals are now readily identifiable within the *a priori* plausible lag time window (shaded blue). (f) Group velocity curve measured on the PCC in (e) for period band 10-20 s. The colouring of the group velocity plot is explained in Section 3.3.3 and Fig. 3.4.

period band for station pairs spaced by less than 450 km. Next we split the PCCs into a causal and a time-reversed acausal part, which yields 8 PCCs (4 causal and 4 acausal) per day for each station pair. The resulting large number of more than 3000 PCCs for the 13-month long recording period (8 PCCs × 30 days × 13 months) should ensure a high signal-to-noise ratio of the PCC stack.

3.3.2 Phase-weighted stacking

In contrast to the conventional linear stacking procedure, we apply a time-frequency domain phase-weighted stacking (tf-PWS), which was developed by Schimmel et al. (2011) as an extension of the phase-weighted stacking introduced by Schimmel & Paulssen (1997). The purpose of tf-PWS is to downweight the incoherent parts of the linear stack in the time-frequency domain (τ, f), which is achieved by the time-frequency phase coherence $c_{ps}(\tau, f)$ (Schimmel et al., 2011):

$$c_{ps}(\tau, f) = \left| \frac{1}{N} \sum_{j=1}^{N} \frac{S_j(\tau, f) e^{i2\pi f\tau}}{|S_j(\tau, f)|} \right|^{\nu}, \tag{3.3}$$

where $S_j(\tau, f)$ denotes the S-transform of the jth individual PCC and N is the number of individual PCCs. The S-transform is used to transform the PCCs in the time-frequency domain (Stockwell et al., 1996). As before, we follow Schimmel et al. (2011) who suggested to use ν=2 for eq. (3.3). The tf-PWS is then defined as the product of the phase coherence $c_{ps}(\tau, f)$ and the S-transform of the linear stack $S_{ls}(\tau, f)$ of all PCCs (Schimmel et al., 2011):

$$S_{pws}(\tau, f) = c_{ps}(\tau, f) \cdot S_{ls}(\tau, f). \tag{3.4}$$

Finally, an inverse S-transform is performed to transform the PCC stack from the time-frequency domain $(S_{pws}(\tau, f))$ back to the time domain $(s_{pws}(t))$ (Schimmel et al., 2011). Schimmel et al. (2011) and Corela et al. (2017) demonstrated in detail the gain in clearer wave arrivals when using tf-PWS instead of conventional linear stacking. This means that tf-PWS provides a faster convergence towards the Green's function and results in more robust group velocity measurements.

In this study, the PCC calculation as well as the tf-PWS are performed by the software package of Schimmel et al. (2011). We make use of all possible correlations that can be derived from our data set of 48 stations (38 OBSs and 10 land stations, see Section 3.2); a few exceptions are correlations between station pairs on La Réunion (like MAID-PRO)

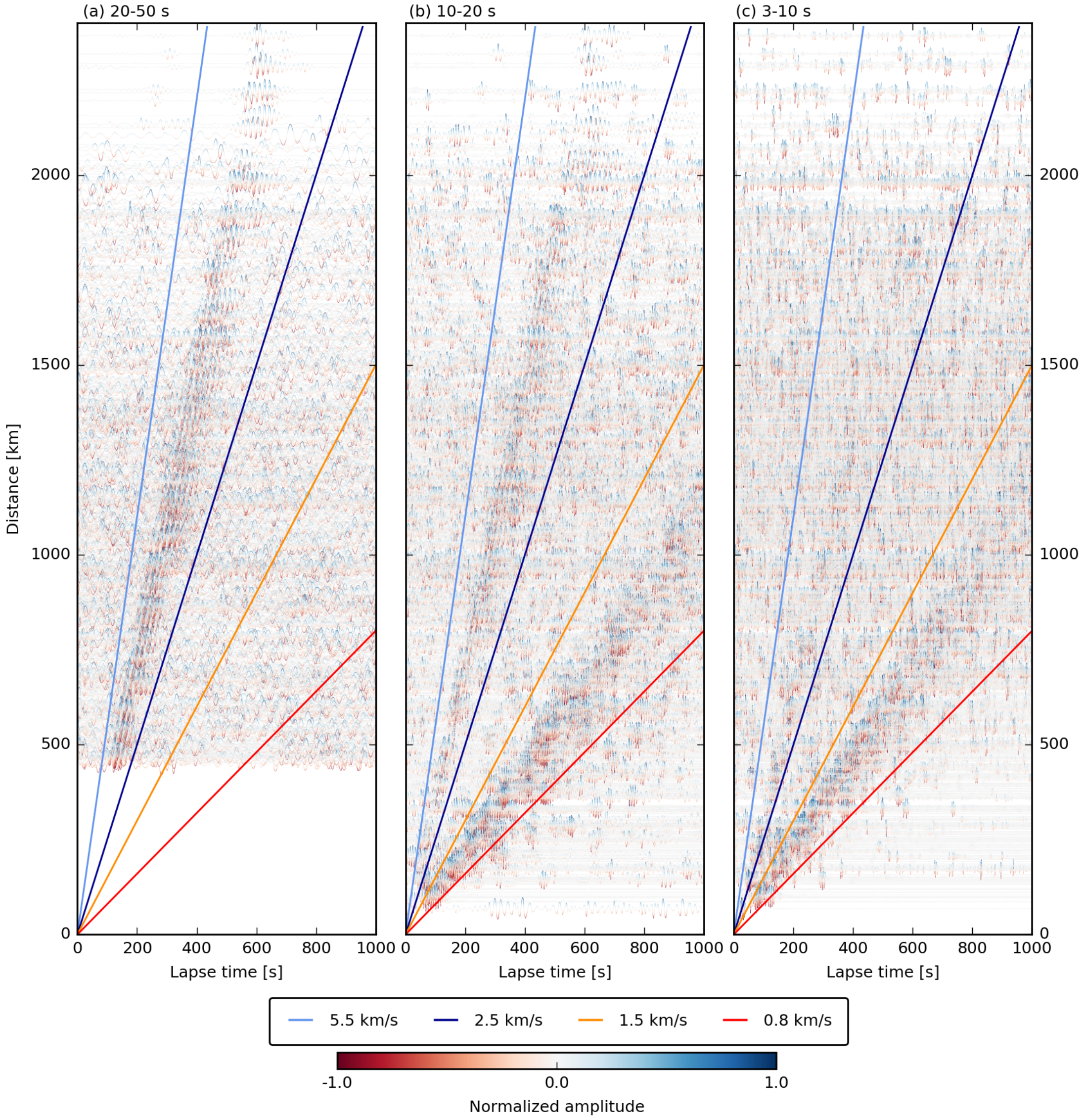


Figure 3.3: All 1119 PCC stacks obtained, sorted by interstation distance, for (a) the period band of 20-50 s; (b) 10-20 s; and (c) 3-10 s. Negative and positive wave amplitudes are coloured red and blue, respectively. The move-out times of plausible Rayleigh-wave arrivals are bounded by velocities of 5.5 km/s (light blue) and 2.5 km/s (dark blue). Rayleigh waves with ∼3.5 km/s are most prominent in the 20-50 s period band, still clear at 10-20 s, and weak at 3-10 s. PCCs in the 20-50 s band are not calculated for distances too short (<450 km) to yield useful group velocity values. A second type of arrival is observed in the period bands of 10-20 s and 3-10 s, travelling at low wave velocities between 0.8 km/s (red line) and 1.5 km/s (orange line). These are interpreted as Scholte waves propagating through seafloor sediments, evidently over distances exceeding 1000 km.

or on Madagascar (like ANLA-FOMA), which cannot provide information about the western Indian Ocean. (Investigation of the crustal structure beneath La Réunion from 23 island stations is the subject of a forthcoming study.) The number of possible station pairs n is determined by following equation:

$$n = \frac{k(k-1)}{2}, \tag{3.5}$$

where k is the number of stations used. This yields 1119 station pairs processed in this study, considering that we excluded 9 stations pairs on La Réunion and on Madagascar. Fig. 3.3 shows the 1119 PCC stacks calculated according to eq. (3.4) for each of the 1119 station pairs and sorted by interstation distance. Since we are working with ZZ correlations, we expect the PCCs to consist mainly of Rayleigh waves. Indeed, the 20-50 s and 10-20 s period bands show prominent wave package arrivals corresponding to propagation velocities of roughly 3.5 km/s (Figs 3.3a and b), which can be associated with Rayleigh waves. However, the 3-10 s period band is dominated by a wave package of significantly lower velocity, in the range of 0.8-1.5 km/s, while Rayleigh wave trains are almost absent (Fig. 3.3c). The same low-velocity wave package is visible in the 10-20 s period band. We think that these slow waves correspond to near-surface waves that propagate through sediments deposited on top of the oceanic crust. This notion is supported by the slow wave velocities and by the observation that the slow waves are visible mainly at shorter periods, which only penetrate the shallowest subsurface. These waves can probably be assigned to a group of interface waves called Scholte waves (Scholte., 1947). Scholte waves of similar velocities were also observed by Le et al. (2018) in data of eleven OBSs deployed in the South China Sea (interstation distances of 60-270 km).

3.3.3 Group velocity calculation

For measuring group velocities, we use the algorithm developed by Schimmel et al. (2017) (but without their data resampling approach). Within a given velocity band, the algorithm localises the amplitude maximum of a PCC stack in the time-frequency domain as a function of frequency f. The arrival time $t(f)$ of the maximum can be transformed to the group velocity $v_g(f)$ by taking the distance d between the cross-correlated stations into account:

$$v_g(f) = \frac{d}{t(f)}. \tag{3.6}$$

The algorithm starts with determining the maximum of the lowest frequency and gradually advances to higher frequencies. For each frequency (except the first), four maxima are calculated, and the one that shows the smallest velocity jump to the preceding group velocity value is chosen. In some cases, no velocity value is determined, e.g. when the velocity jump is too large or the determined maximum is below a certain amplitude threshold.

For extracting group velocity values that can be associated with Rayleigh waves, we specify the velocity range of 2.5-5.5 km/s (cf. Fig. 3.3). In all PCC plots, this velocity range is indicated by a blue-shaded area (e.g. Fig. 3.2e). Group velocity values are illustrated as amplitude spectra, which are vertically normalized (e.g. Fig. 3.2f). This means that each frequency column contains the maximum value of 1, which is represented by yellow colour; smaller amplitude values are shown in green and blue, with dark blue corresponding to 0. If the algorithm could retrieve a maximum, the associated group velocity value is represented by a black dot. Group velocity values whose amplitudes are within 95% of the maximum amplitude are marked by dashes.

At this point, we want to emphasize the benefit of using several narrower frequency bands instead of one wide band (see Fig. 3.4). We calculate PCC stacks for RR40-RR50 for a wide band of 3-50 s (Fig. 3.4a), and for its three narrower sub-bands of 20-50 s, 10-20 s, and 3-10 s (Figs 3.4b-d), which we actually proceeded to use in our study. In each of the PCC stacks, clear Rayleigh-wave arrivals are observed, regardless of the width of the band. For each of the period bands, we calculate group velocity values between 2.5 km/s and 5.5 km/s (blue-shaded areas in Figs 3.4a-d). The group velocity plots are shown in Figs 3.4(e)-(h), the colouring corresponds to the description given above. The scaling of the frequency axis of Figs 3.4(f)-(h) is adjusted such that it matches the frequency axis of Fig. 3.4(e). Group velocity values selected by the algorithm in the 3-10 s period band are very similar to the values selected in the 3-50 s band for the same frequencies. By contrast, no reliable group velocity measurement can be retrieved in the 3-50 s band for frequencies below ~0.1 Hz, as the yellow ridges of high-amplitude region become oscillatory and/or multivalued. For a reliable result, we would expect the high amplitudes gradually approach to one maximum value for each frequency, forming a smooth group velocity curve as function of frequency. Such benign behaviour is indeed observed in the narrower sub-bands in Figs 3.4(f) and (g). Similar observations were made by Haned et al. (2016). We concluded that the three narrow sub-bands should be adopted to obtain more numerous and more robust group velocity measurements.

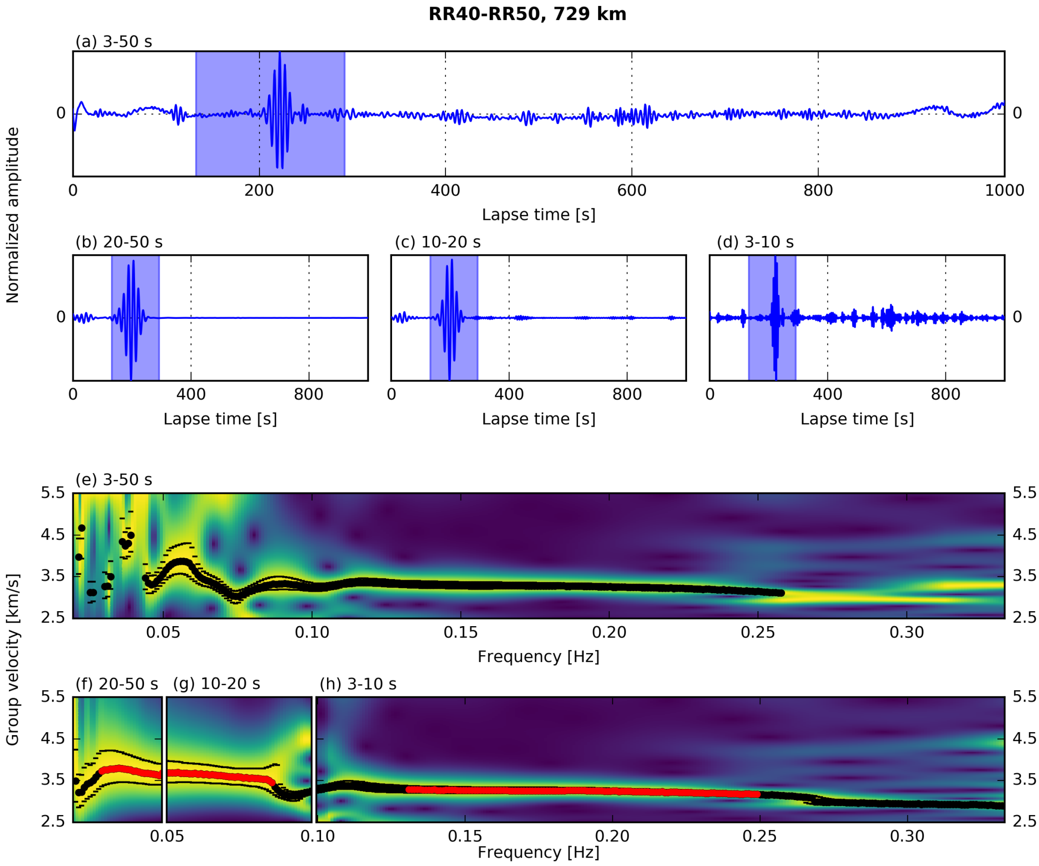


Figure 3.4: Measurement of PCC stacks and group velocities for station pair RR40-RR50, two OBSs spaced by 729 km. PCC stacks for (a) the wide period band of 3-50 s, versus sub-bands of (b) 20-50 s, (c) 10-20 s, and (d) 3-10 s. Blue-shaded areas bracket the arrival times of wave velocities between 2.5 km/s and 5.5 km/s. For this velocity range, estimated group velocities are shown in (e) for the wide band, and in (f)-(h) for the three sub-bands. Yellow and blue colours indicate high and low amplitudes, respectively, and amplitude is normalized so that its maximum value (yellow) is identical at each frequency. Black dots mark the group velocity values selected by the algorithm of Schimmel et al. (2017), and black bars denote their 95 % confidence ranges. Red dots replacing black dots are group velocity values we subjectively accepted for tomography, based on the simple, smooth appearance of the sequence of black dot (see text for discussion). According to this criterion, no group velocity measurement in the 3-50 s period band would be acceptable at frequencies below ∼0.1 Hz, whereas subdivision into three narrower bands (20-50 s, 10-20 s, 3-10 s) allows confident group velocity estimation in all three bands.

3.4 Results: group velocity

Fig. 3.5 shows PCC stacks and group velocity plots for long-distance examples of the three kinds of station pairs correlated: OBS-to-OBS correlations exemplified by station pair RR34-RR50 in Fig. 3.5(a)-(f); land-to-OBS pairs exemplified by LAHA-RR52 in Fig. 3.5(g)-(l); and land-to-land pairs exemplified by MRIV-RUM1 in Fig. 3.5(m)-(r). The interstation distances for the pairs with an OBS are 1882 km and 1985 km, much larger than in any prior work, and Rayleigh-wave arrivals are observed very clearly in the 20-50 s and 10-20 s bands.

We visually evaluated all 1119 group velocity plots of the kind shown in Fig. 3.5 in order to decide on the subsets of automatically picked values (black dots) that seemed robust and plausible enough to be considered for further processing. This value extraction is to some extent subjective, but the desired curve shape is a smooth curve that is not distorted by blue holes. An example of blue holes deforming the velocity curve can be seen in Fig. 3.4(g) for frequencies above ∼0.09 Hz. Moreover, the existence of only one clear maximum per frequency points towards a reliable measurement. Frequencies in the 3-10 s period band are often characterized by several maxima of similar amplitudes, i.e. by several yellow bands in the group velocity plots.

We depict the manually selected group velocity values by red dots superimposed on the black dots (see Figs 3.2f, 3.4f-h, 3.5d and e, 3.5j and k, and 3.5p and q). The results of pairs RR34-RR50 and LAHA-RR52 demonstrate that reliable group velocity estimates can be derived from OBS seismograms for very large interstation distances of ∼2000 km.

We note that group velocity measurements on OBS-to-OBS correlations are particularly successful when at least one of the two stations is an INSU station. The reason for this becomes obvious when we compare the relative noise levels, quantified as probabilistic power spectral density plots (PPSD) (McNamara & Buland, 2004) of a DEPAS OBS and an INSU OBS (Fig. 3.6). The two sensors are situated only 150 km apart in the same abyssal plain at similar water depths, and hence presumably in similar ambient noise conditions. Yet DEPAS OBS RR26 features a dramatically higher noise level than INSU OBS RR28 for periods above ∼5 s, which is well within the period bounds of our study (3 s and 50 s, orange lines in Fig. 3.6). The noise levels of RR26 even exceed those of the New High Noise Model (NHNM) derived by Peterson et al. (1993) for terrestrial stations, whereas INSU OBS RR28 stays well below. Stähler et al. (2016) calculated PPSD plots for each RHUM-RUM OBS and showed that the noise level observations for RR26 and RR28 are generally valid for DEPAS/GEOMAR and INSU OBSs of the RHUM-RUM experiment. They concluded that the low power consumption of DEPAS seismometers

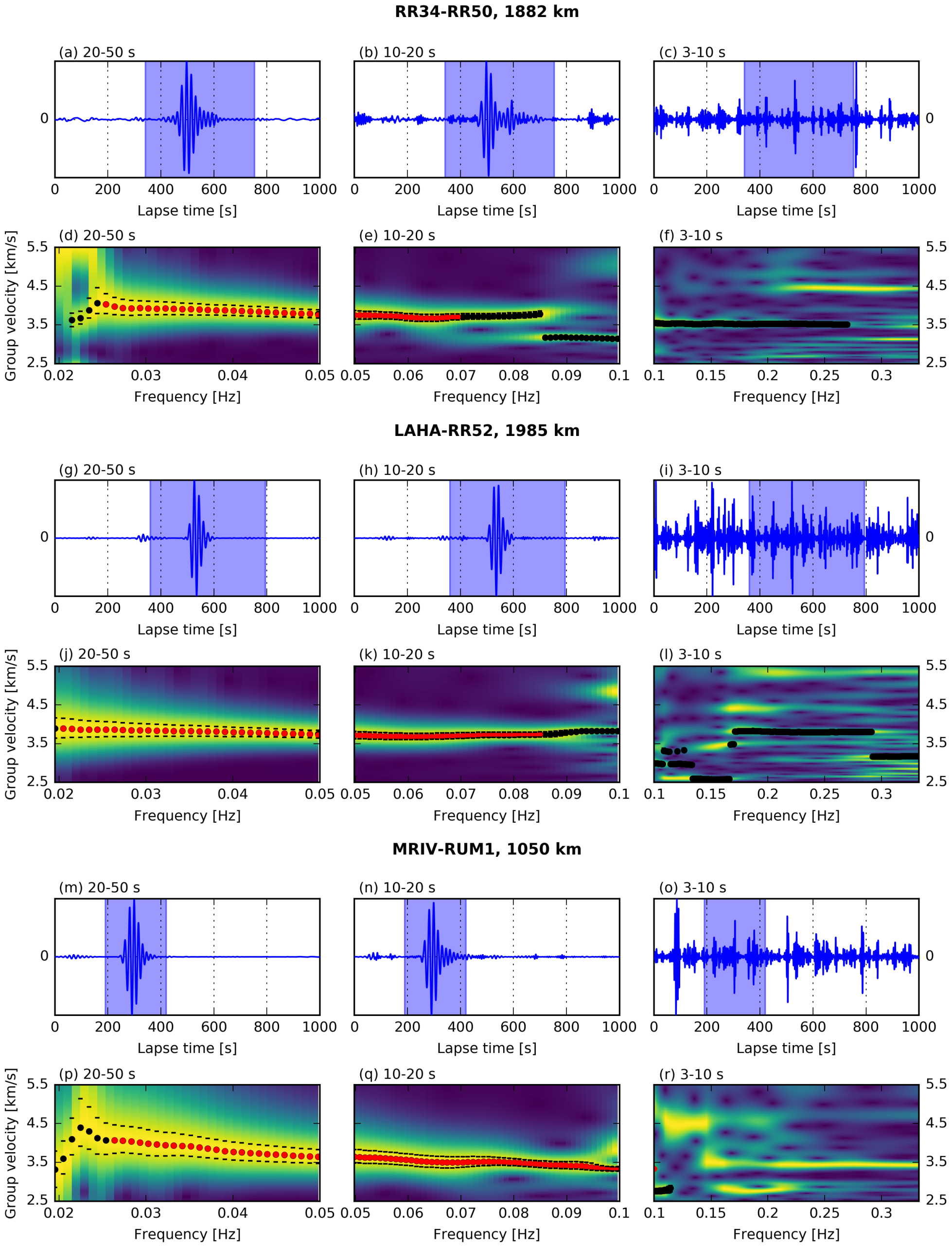


Figure 3.5: PCC stacks and group velocity estimation for three more station pairs: (a)-(f) correlation of two OBSs (RR34, RR50) spaced by 1882 km; (g)-(l) correlation of land station LAHA in Madagascar with OBS RR52, spaced by 1985 km; (m)-(r) correlation of two land stations (MRIV in Mauritius, RUM1 in southern Madagascar), spaced by 1050 km. Blue-shaded areas in (a)-(c), (g)-(i), and (m)-(o) denote the velocity range of 2.5-5.5 km/s, for which group velocities are estimated in (d)-(f), (j)-(l), and (p)-(r). See caption of Fig. 3.4 for complete explanation of colours and symbols used.

(Güralp CMG-40T-OBS) probably comes at the cost of very high instrumental self-noise (above NHNM) of the vertical component. This problem was further investigated by Stähler et al. (2018), who recommended to use CMG-40T-OBS sensors only for short-period applications. The presence of this very high self-noise explains the difficulty for DEPAS-to-DEPAS PCCs to converge to the Green's function, which hampers group velocity measurements.

Fig. 3.5 depicts that group velocity estimates of INSU-to-INSU OBS pairs are of comparably high quality as for land-to-OBS pairs (especially land-to-INSU) and for land-to-land correlations. Figs 3.5(g)-(l) illustrate a successful noise correlation between a land station and an OBS, an achievement that has been reported in only a few prior studies (e.g. Corela et al., 2017; Hable et al., 2018). All three station pairs presented in Fig. 3.5 lack a clear wave package signal in the 3-10 s period band, both in the Rayleigh-wave arrival window (shaded blue) and in the entire lag window. Considering the large interstation distances of 1000-2000 km, this observation is not surprising because the low-velocity (Scholte) wave trains of Fig. 3.3(c) would appear at later lapse times >1000 s.

From among the 1119 station pairs, we identify 628 station pairs that could contain group velocity values usable for tomography. These values are given in Fig. 3.7 by the

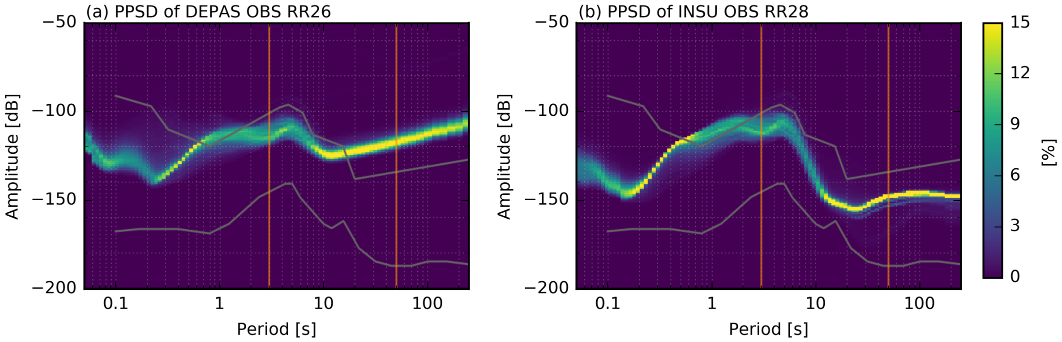


Figure 3.6: Comparison of noise spectra between the two types of OBS used. (a) Probabilistic power spectral density (PPSD) computed after McNamara & Buland (2004), for OBS RR26 from the German DEPAS pool. (b) PPSD for OBS RR28 from the French INSU pool, located at 150 km distance from RR26 in similar conditions. Colouring represents the occurrence of different noise levels, with yellow indicating frequent occurrence. Upper and lower grey lines mark the New High Noise Model and New Low Noise Model of Peterson et al. (1993). Orange vertical lines at 3 s and 50 s delimit the period range investigated here. At periods longer than ~5 s, the INSU OBS is much quieter than the DEPAS OBS. This comparison between the two OBS types holds true generally and seems to be due to high self-noise on the vertical components of the Güralp CMG-40T sensors in the DEPAS seismometers (Stähler et al., 2016). Hence station pairs comprising at least one INSU station are characterized by PCC stacks with clear Rayleigh-wave arrivals and group velocity measurements of accordingly higher confidence.

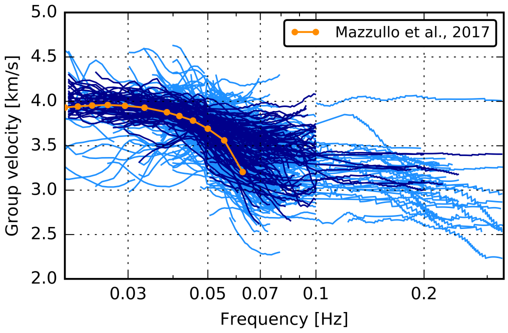


Figure 3.7: Selected group velocity curves that could contain usable information for tomography are shown in light blue (cf. red dots in Figs 3.2, 3.4, and 3.5). The best 100 group velocity curves used in this study are coloured dark blue. The orange curve denotes the average group velocities for the RHUM-RUM region, obtained by Mazzullo et al. (2017) from Rayleigh-wave tomography using earthquake sources.

light blue curves. The abrupt truncation of many curves at 0.1 Hz can be explained by the frequent absence of Rayleigh waves in the 3-10 s period band, which in most cases prevents reliable group velocity estimates. For tomography, we admit only the 100 best group velocity measurements, represented by the dark blue curves in Fig. 3.7. All group velocity plots represented in the figures represent one of these 100 best station pairs. 85 out of the 100 best station pairs are either land-to-OBS or OBS-to-OBS correlations. At least one INSU OBS is part of 82% of these 85 station pairs, clearly reflecting the superior data quality of INSU OBSs compared to DEPAS OBSs for the purpose of group velocity calculations.

3.5 Results: tomography

3.5.1 Tomographic inversion of the group velocity data

The scientific aim is to improve and complete the Rayleigh-wave tomography of Mazzullo et al. (2017), specifically by constraining 3-D crustal structures which were poorly resolved by the longer-period measurements on earthquake-generated seismograms. Mazzullo et al. (2017) used ∼300 earthquakes recorded by 130 seismic stations (mainly RHUM-RUM stations) to measure Rayleigh-wave group velocities in the period range of 16-250 s and phase velocities of 30-300 s. The sensitivity of these periods to the earth's upper ∼40 km is very limited. Our noise-derived group velocities between 3-50 s period are mainly sensitive to depths above ∼50 km. Hence the two data sets are complementary and are used in a joint tomographic inversion here.

First, we check if the ray coverage of our 100 retained, highest-quality measurements is sufficient for tomography (as an alternative might be to relax our quality control standards). For this purpose, we perform tomographic resolution (checkerboard) tests, as shown in Fig. 3.8. We forward-predict group velocity curves through a hypothetic crust consisting of squares of alternating high and low velocity, superimposed on the 3-D reference model, which is described below. The length of the squared velocity anomalies is 500 km; the velocity anomalies range from -4 % to +4 %. We use the same inversion parameters (correlation length 200 km, *a priori* error 0.1 km/s) as for the subsequent inversion of real data.

Fig. 3.8 shows the test input and the recovered, regionalised output model for a period of 16 s and the 92 wave paths accepted for this period. The checkerboard pattern is recovered well beneath the instrumented area. Recovery outside this area is not expected since noise correlations yield Green's functions and sensitivities between stations only. The good recovery of the checkerboard pattern validates our decision to admit only the very highest quality measurements (100 out of 1119), even though this means discarding over 90% of unique wave paths.

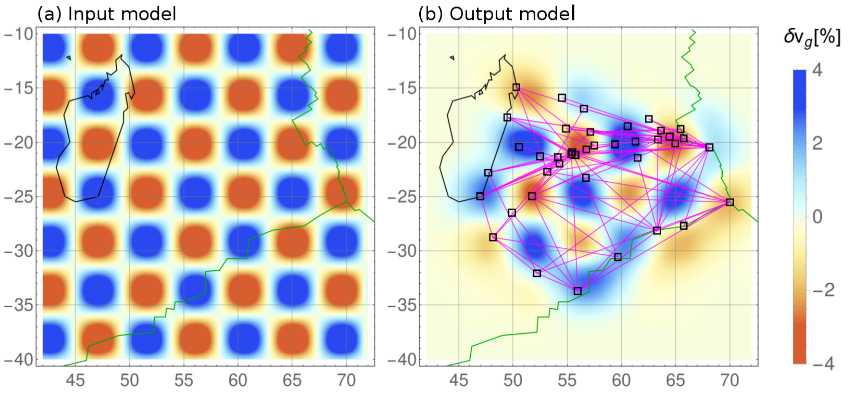


Figure 3.8: Resolution test for tomography, for wave period of 16 s (92 ray paths). (a) Checkerboard input model of alternating high- and low-velocity squares with a square length of 500 km. (b) Recovered model with black squares indicating the used stations and purple lines representing the 92 nominal ray paths. Good recovery within the instrumented area indicates that wave path coverage remains adequate even after the severe quality control step that accepted only the best 100 station pairs.

The joint inversion of both data sets follows the method described by Mazzullo et al. (2017). A locally modified (i.e. slightly smoothed) version of CRUST1.0 (Laske et al., 2013) is used as *a priori* 3-D crustal model, which is underlain by the spherically symmetric PREM reference model for the mantle (Dziewonski & Anderson, 1981), albeit with a modified (smoothed) 220 km discontinuity. CRUST1.0 can be interactively visualized in the *SubMachine* tomography web portal (Hosseini et al., 2018). The calculation of the S-wave velocity model is based on the *a priori* model and the weighted sum of B-spline basis functions:

$$V_s(z) = V_s^0(z) + \sum_{m=0}^{M-1} W_m N_{m,2}(z). \tag{3.7}$$

$W_m N_{m,2}(z)$ is the mth non-uniform quadratic B-spline basis function (De Boor, 1978), M is the number of B-spline basis functions, and W_m is the weighting coefficient. $V_s^0(z)$ is the *a priori* S-wave velocity reference model.

The transdimensional inversion is composed of two nested loops. The inner loop calculates the optimum model weighting coefficients W_m for a given spline basis $N_{m,2}$, by minimizing the data misfit function (χ_d^2) between measured and modelled velocities using the simulated annealing optimization algorithm (Press, 2007). The outer loop determines the optimum spline basis (the shapes and the number of splines M). It uses the golden section search method (Press, 2007) to minimize the expression $(\chi_d^2 + \chi_m^2)/2$ as a function of M, where χ_d^2 is the outcome of the inner-loop minimization and χ_m^2 is the model variance quantity defined by eq. (B3) in Haned et al. (2016). The procedure of the transdimensional inversion yields the optimal number of splines and their shapes, which reflects the best compromise between data fit and model smoothness. A more detailed method description of the inversion is given by Haned et al. (2016) and Mazzullo et al. (2017).

In practice, the inversion is performed in two steps. First, we use a correlation length of 800 km to regionalise our noise data set and the data set of Mazzullo et al. (2017). The inversion of the regionalised data is performed according to eq. (3.7) with the smoothed CRUST1.0 and smoothed PREM as *a priori* model. The obtained low-resolution model together with the original CRUST1.0 serves then as starting model for the mantle and the crust, respectively, for the second inversion, that uses a correlation length of 200 km to retrieve a model of higher resolution (presented in Section 3.5.2). This two-step inversion procedure is chosen to resolve large-scale velocity structures at deeper layers (from long-period data) as well as small-scale structures at shallower depths (from short-period data).

3.5.2 S-wave velocity model

We obtain a 3-D S-wave velocity model by inversion of our noise-derived, regionalised group velocity data jointly with the phase and group velocities of Mazzullo et al. (2017) according to eq. (3.7) and using the 3-D reference model described earlier. Fig. 3.9

compares the tomography results of Mazzullo et al. (2017) (left column) with those of our joint inversion (middle column) for depths of 20 km, 30 km, 40 km, 60 km, and 80 km. The last column of Fig. 3.9 highlights the differences between the two models by plotting the velocity variations of Mazzullo et al. (2017) subtracted from our joint inversion model. Each panel shows S-wave velocity variation (in per cent) from an average, absolute velocity value at this depth, which is given at the bottom right of each panel. For easier comparison, this average velocity value is chosen to be identical for the two tomography models at a given depth. Red and yellow colour shades indicate that seismic velocity is estimated to be lower than the average velocity at this depth, whereas blue shades indicate higher-than-average velocities. Fig. 3.9(p) shows the bathymetry of the studied region together with the main tectonic structures.

The two tomography models differ mainly at crustal depths (~20-30 km, Figs 3.9a-f), which are constrained primarily by our new noise data. The panels in Fig. 3.9 show the following, major differences:

1. Slower spreading ridges. Yellow and red shades beneath the Central Indian (CIR) and Southwest Indian Ridges (SWIR) in Fig. 3.9(c) indicate that the noise-derived data sense slower structure at 20 km depth than the earthquake-derived data of Mazzullo et al. (2017), which have good sampling coverage beneath the ridges, but not much sensitivity to these shallow depths. Our noise data set has very good sensitivity since these two mid-ocean ridges were actually instrumented by RHUM-RUM OBSs. As regions of mantle upwelling, decompression melting, and very thin lithosphere, mid-ocean ridges generally appear as localised bands of slow seismic anomalies in seismic tomography models. This is also true for the models of Mazzullo et al. (2017) at depths >30 km, where a velocity gradient perpendicularly away from a ridge is clearly evident (red to yellow to blue). Such a gradient is weak or absent at 20 km in the model of Mazzullo et al. (2017) (Fig. 3.9a), but is evident in the joint model (Fig. 3.9b) and in the difference plot (Fig. 3.9c). Note that below 40 km, the differences between the two tomographies are marginal as expected, given that our noise data set has negligible sensitivity to those depths. Below 40 km, the slowness of the spreading ridges is rendered by the earthquake data of Mazzullo et al. (2017); at shallower depths, the same is accomplished (only) by the noise-derived data.

2. Slower Madagascar Plateau. South of Madagascar at latitudes of ~27°and 30°-35°, two velocity anomalies that were already intensely slow in the inversion of Mazzullo et al. (2017) are rendered even slower by the joint inversion at 20 km. These anomalies coincide with the thickened oceanic crust of the Madagascar Plateau

(see also Fig. 3.1a), which is attributed to flood basalts from the plume head of the Marion hotspot (Storey et al., 1995). The crustal structure of Madagascar itself is not modified by our noise data, which basically do not sample beneath the continent, see Fig. 3.8. Madagascar is underlain by truly continental, 20-35 km thick crust (Rindraharisaona et al., 2017), which is built into the prior crustal model as very slow structure.

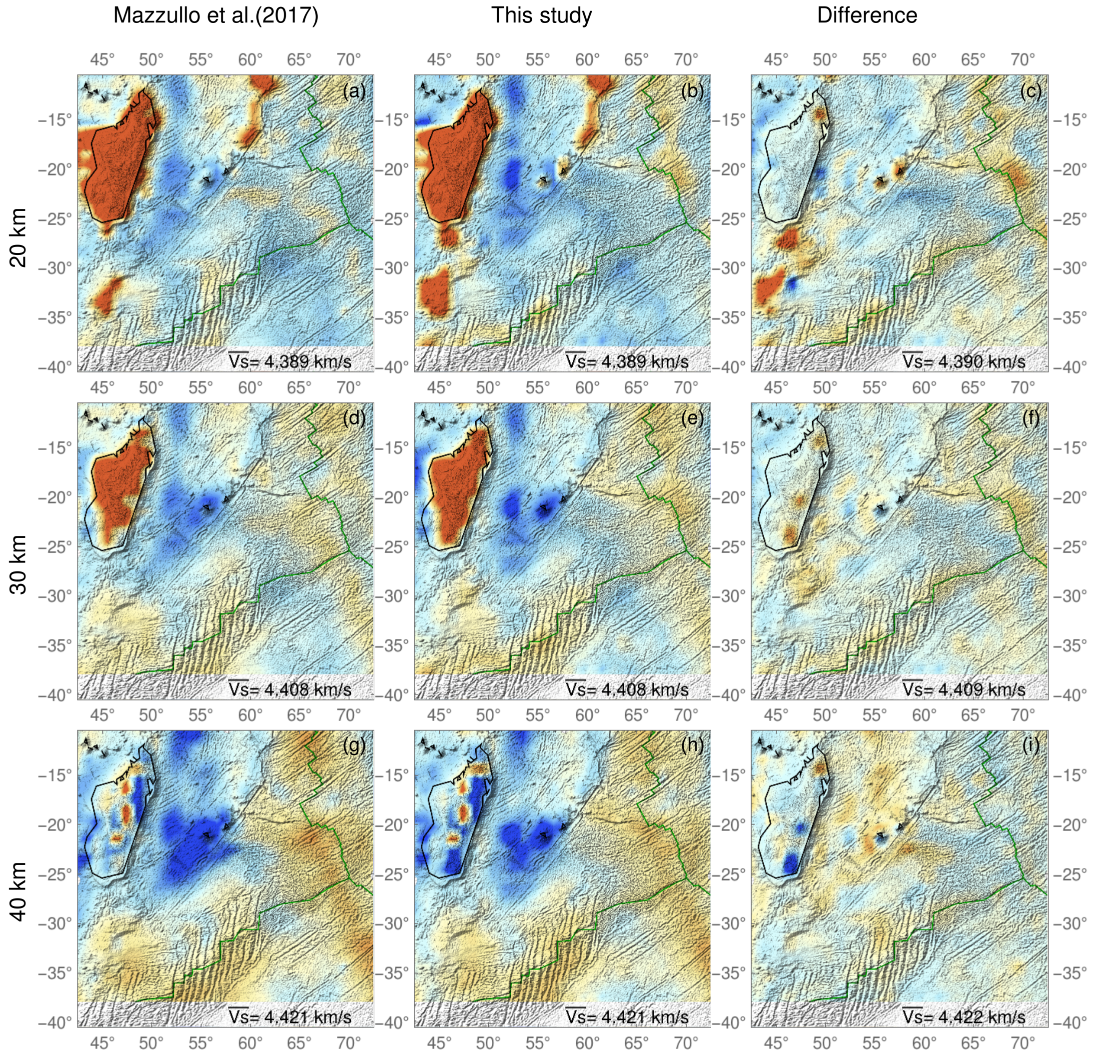


Figure 3.9: Tomography results. S-wave velocity models at the depths of 20 km in panels (a)-(c), 30 km in (d)-(f), 40 km in (g)-(i), 60 km in (j)-(l), and 80 km in (m)-(o). First column shows the velocity model of Mazzullo et al. (2017); second column resulted from the joint inversion of the 100 best group velocity curves obtained by this study, with the data set of Mazzullo et al. (2017). Third column represents the difference between the two results (model of Mazzullo et al. (2017) subtracted from the joint inversion). Velocity variations range from -6% (red) to +6% (blue) for 20 km and 30 km depth and from -4% (red) to +4% (blue) for 40-80 km depth. The variations refer to a depth-dependent average velocity value given at the bottom right of each panel. The main tectonic structures of the region are indicated in the bathymetry map in panel (p). Plate boundaries (spreading ridges) are indicated by green lines.

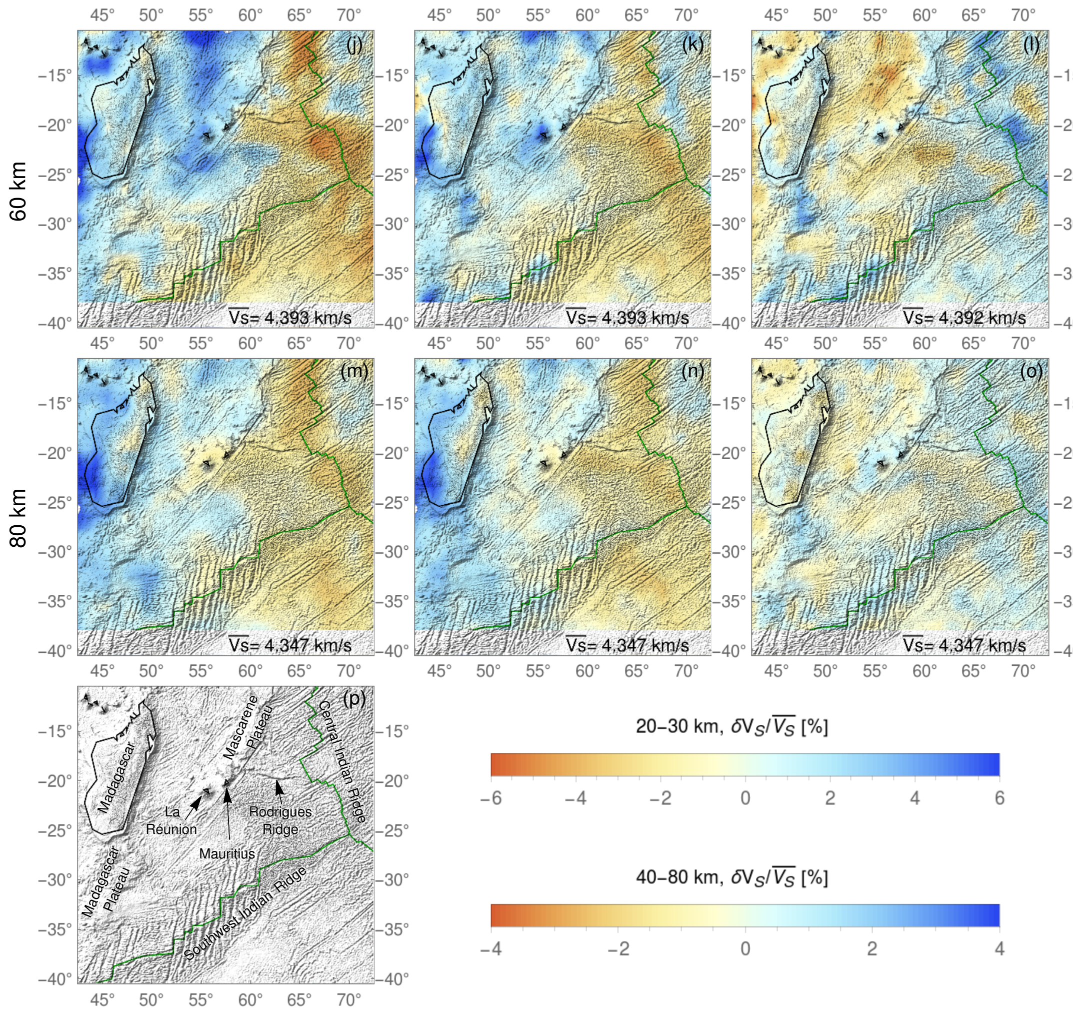


Figure 3.9: Continued.

3. Slower Mascarene Plateau. The Mascarene Plateau extends from Mauritius to the Seychelles and is characterized by very slow velocity anomalies at 20 km depth in Mazzullo et al. (2017) (Fig. 3.9a). It is rendered even slightly slower by our joint inversion (Figs 3.9b and c). The Mascarene Plateau is thought to represent part of the Réunion hotspot track, i.e. its oceanic crust would have been magmatically thickened by mantle plume activity.

4. Slow Rodrigues Ridge. Marked by a thickened, east-west striking ridge between Mauritius and the CIR, Rodrigues Ridge showed no slow-velocity anomaly at 20 km and 30 km depth in the model of Mazzullo et al. (2017), but it does in the joint inversion (Figs 3.9b and e). Rodrigues Ridge has long been hypothesized to represent a manifestation of hotspot-ridge interaction: the surface record of a leaky asthenospheric flow channel that transports plume heat and material to the spreading ridge (Morgan, 1978). Shallow slow anomalies would be expected here from

recent or ongoing volcanism and from crustal thickening (Morgan, 1978; Dyment et al., 2007).

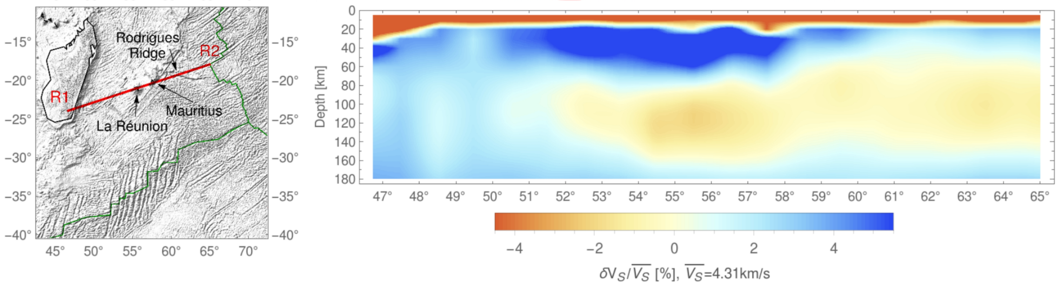
5. Faster old oceanic lithosphere. The high-velocity anomaly that extends beneath much of the Mascarene ocean basin between Madagascar and La Réunion at 20 km and 30 km depth is rendered even faster by the joint tomography compared to Mazzullo et al. (2017). This region hosts some of the oldest (hence fastest) crust in the Indian Ocean, and its seismic signature seems to have been underestimated by Mazzullo et al. (2017). More generally, the joint tomography indicates faster oceanic crust in almost all sampled regions that are not sites of recent volcanism or ancient crustal thickening.

6. Slower crust and faster mantle beneath the hotspot islands of La Réunion and Mauritius. At 20 km, these two islands show up as circular low-velocity anomalies of ∼100 km diameter, immersed in the seismically fast surroundings of Mascarene basin crust (Fig. 3.9b). These details are absent in the model of Mazzullo et al. (2017) (Fig. 3.9a), and Fig. 3.9(c) renders this difference very crisply. At 30 km, the situation is inverted in that the islands now appear as very fast dots in a moderately fast environment, which again is particularly clear in the difference plot Fig. 3.9(f). By comparison the model of Mazzullo et al. (2017) makes a blurred impression. At 80 km, the localised fast anomaly beneath La Réunion give way to a spatially broader, low-velocity anomaly (Fig. 3.9n), which can probably be attributed to high-temperature upwelling through the Réunion mantle plume (Mazzullo et al., 2017).

The situation of the hotspots is clarified by a cross section through La Réunion and Mauritius down to a depth of 180 km in Fig. 3.10(a). The oceanic crust appears as shallow slow (red) layer in the tomography. Its bottom, the Moho, undulates anticorrelated to the topography of the surface (red line above the cross section): where the two volcanic islands rise from the deep Mascarene Basin, the Moho is depressed by 5 km (beneath La Réunion) to 10 km (beneath Mauritius). This is broadly consistent with receiver functions indicating Moho depths of ∼12 km (Fontaine et al., 2015) for La Réunion (using one permanent station), and ∼10-20 km (Fontaine et al., 2015; Singh et al., 2016) for Mauritius (using a network of stations across the island). We concur with these authors that this very likely reflects the crustal thickening or magmatic underplating expected from hotspot activity, possibly enhanced by the presence of melt.

Below the crust, seismically fast lithosphere presents as a blue layer that mirrors the undulations of the Moho. The lithosphere in the left part of the plot (southwest of La Réunion with longitudes of 51°-54°) and between the two islands may be thought of

(a) This study



(b) *A priori* model

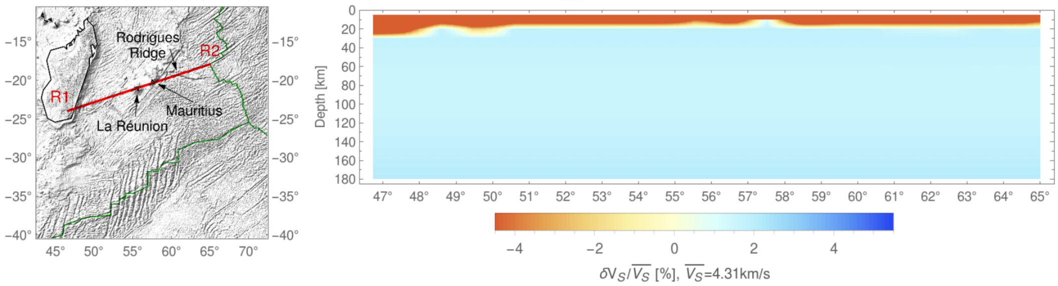


Figure 3.10: Cross section through the islands of La Réunion and Mauritius of (a) our joint tomography model and (b) the *a priori* model (smoothed CRUST1.0 and PREM). Location is given by the red line R1-R2 on the maps. Velocity variations range from -4.5% (red, slow) to +5.5% (blue, fast) relative to a velocity value of 4.31 km/s. Topography and bathymetry along the profiles are given by the red curve above the cross sections.

a "background" Mascarene Basin lithosphere; the bottom of this layer is depressed by about 10 km underneath the two hotspots. We think that this depression is the fast seismic signature of depleted and dehydrated mantle from which the hotspot sourced and sources the melt for the observed crustal thickening.

The *a priori* model of the first inversion step (smoothed CRUST1.0 and PREM) is given in Fig. 3.10(b) for the same cross section as shown in Fig. 3.10(a). The direct comparison of our joint model and the *a priori* model clearly reveals that the newly found undulations of the Moho and the lithosphere beneath the islands La Réunion and Mauritius are constraints added by our noise data set.

Aside from the two lithospheric depressions beneath the islands, there is a general thinning trend from southwest to northeast. Just northeast (right) of Mauritius, abrupt and pronounced thinning of the fast layer coincides with the transition at the surface between older seafloor produced by the Mascarene Basin paleo-spreading ridge and the significantly younger lithosphere produced by the (still-active) CIR. This general lithospheric thinning was also observed by Mazzullo et al. (2017) and is indeed mainly constrained by their data in our joint inversion. It agrees with receiver function results by Fontaine et al. (2015), who found thicker lithosphere beneath La Réunion (70 km) than beneath Mauritius (50 km).

Below the fast lithospheric layer, the slow asthenosphere lies, which is slowest beneath the hotspot area, presumably reflecting ongoing, hot plume upwelling (Mazzullo et al., 2017).

In summary, our new constraints from noise-derived group velocities act to render crustal velocity anomalies more pronounced (both in the fast and the slow directions), compared to the earthquake-derived constraints of Mazzullo et al. (2017). The latter model stayed closer to the crustal reference model, for lack of constraints to the contrary.

3.6 Discussion

We present the first crustal tomography of S-wave velocity of the western Indian Ocean between Madagascar and the three spreading ridges of the Indian Ocean, a region that is roughly centred on the hotspot of La Réunion. A new data set of 100 group velocity curves between 3 s and 50 s was inverted jointly with the Rayleigh-wave phase and group measurements obtained by Mazzullo et al. (2017) from earthquake sources and the same RHUM-RUM stations used here.

Interstation distances of up to 2000 km far exceed those measured and inverted in previous noise correlation studies in the oceans.

From our total data set of 1119 station pairs, we identify 628 station pairs that might contain group velocity measurements of suitably high confidence to be used in tomography. In the end, we decide to use only 100 measurements corresponding to our best group velocity curves to ensure the most robust tomography possible. The reason for this limited number is the high self-noise of the DEPAS OBS type (Güralp CMG-40T-OBS sensors), which hampers reliable group velocity estimation especially on DEPAS-to-DEPAS station pairs, which amount to 406 out of 1119 station pairs. INSU OBSs are not affected by such high self-noise and yield results of comparable quality as land stations.

Most of the structural anomalies detected by our joint inversion of noise-derived group velocities with the longer-period data of Mazzullo et al. (2017) reflect tectonic structures (Section 3.5.2). Known or suspected regions of thickened crust and/or volcanic activity are represented by low-velocity signals around 20 km depth. Several of these shallow anomalies, e.g. under Rodrigues Ridge and the CIR, were not picked up by the tomography of Mazzullo et al. (2017), which demonstrates the valuable contribution of noise-derived group velocity measurements to crustal imaging. As future work, group velocities of the 628 station pairs identified as potentially usable, could be used to investigate whether they can further refine the model of Fig. 3.9.

There is not much prior imaging work to compare to. Ma & Dalton (2017) investigated the lithosphere of Africa and the Indian Ocean with combined earthquake and ambient noise tomography, but limited to the sparse, permanent land and island stations and without the benefit of data from RHUM-RUM stations. Hence a reasonable comparison must be limited to large-scale anomalies in light. They derived phase velocity maps at a period of 30 s, which might allow for a rough comparison with our tomography models at 20-40 km depth, but a detailed comparison of phase velocity maps and a 3-D S-velocity model is not straightforward. In agreement with our results, Ma & Dalton (2017) found a slow-velocity anomaly beneath the CIR and a high-velocity signature between Madagascar and La Réunion (Mascarene Basin).

An interesting feature of Figs 3.9 and 3.10(a) is the high-velocity anomaly at ∼30-70 km beneath La Réunion. Similar velocity signatures have been reported beneath at least two other hotspot islands. Villagómez et al. (2007) discovered a high-velocity lid beneath the Galapagos archipelago at a depth of 40-70 km, above a slow-velocity plume signature in the asthenosphere. Schlömer et al. (2017) detected a circular high-velocity anomaly at a depth of 50 km beneath the Tristan da Cunha archipelago. The strikingly consistent depths of these observations suggest an identical mechanism for the high-velocity anomaly under La Réunion. Villagómez et al. (2007) suggested that the hot mantle plume caused a remelting of lithosphere, accompanied by depletion and dehydration, which would appear as high-velocity anomaly. Schlömer et al. (2017) speculated that the high-velocity body could be highly depleted plume material that has accreted to the lower lithosphere. The high-velocity anomaly beneath La Réunion had already been detected by Mazzullo et al. (2017), but due to the sparse data coverage in their model at shallow layers, it appears not as localised anomaly, but was blurred by the fast-velocity signature of the cold oceanic lithosphere. This reflects the important information gain due to the joint inversion.

In general our data acted to enhance, both in the seismically fast and slow directions, crustal anomalies that were already present in the *a priori* model, a smoothed version of CRUST1.0. This is highly plausible, since CRUST1.0 is an interpolation between different types of crustal constraints, acquired at very different length scales, and subsequently homogenized to the scale of 1 degree (∼100 km).

3.7 Conclusions

Group velocities between 3 s and 50 s period are calculated from OBS-to-OBS, land-to-OBS, and land-to-land correlations. The 100 very best measurements are inverted for 3-D S-velocity structure of crust and lithosphere, jointly with the earthquake-generated data set (periods >16 s) of Mazzullo et al. (2017). Our relatively shorter-period data are ideal for adding constraints on the the shallowest (i.e. crustal) depths of the S-wave velocity model of Mazzullo et al. (2017), who used phase and group velocities with periods of 30-300 s and 16-250 s, respectively. The joint inversion provides the first tomographic S-wave velocity image of the crust and the uppermost mantle of the western Indian Ocean between Madagascar and the three spreading ridges, centred on the hotspot island of La Réunion. The tomography model is made available as electronic supplement to this article.

We demonstrate that high-quality group velocity estimates can be obtained from OBS-to-OBS correlations with very large interstation distances of ∼2000 km, while prior studies had reported successful OBS group velocity estimates from distances of at most a few hundred kilometres (e.g. Yao et al., 2011; Zha et al., 2014; Ball et al., 2016; Corela et al., 2017; Ryberg et al., 2017). Our results mean that the noise-correlation method is applicable even for relatively sparsely instrumented areas, which includes most oceanic regions. In addition, we show that land-to-OBS correlations provide high-quality group velocity measurements, even though land stations and OBSs sample in different crustal conditions (e.g. Corela et al., 2017; Hable et al., 2018).

We newly image several slow-velocity signatures around 20 km depth, not visible in the model of Mazzullo et al. (2017). These signatures can clearly be associated with known tectonic structures of thickened crust and/or volcanic activity (La Réunion, Mauritius, Rodrigues Ridge, Central Indian Ridge, Madagascar Plateau). We also find a high-velocity body directly beneath La Réunion in a depth of 30-70 km. This fast anomaly is consistent with tomographic results for the same depths beneath some other hotspot islands (Galapagos, Tristan da Cunha) and can probably be attributed to highly depleted and dehydrated material, either of remelted lithosphere or of underplated plume material. This is the first time that such a high-velocity anomaly is described beneath La Réunion.

Chapter 4

Group velocity estimations of the island La Réunion

This chapter is being prepared for submission under the title *Noise tomography of the island La Réunion* (Hable et al., 2019b).

Abstract

We present the first seismological study that investigates underground structures of the entire island of La Réunion located in the western Indian Ocean. This volcanic hotspot island is characterized by the ancient volcano Piton des Neiges and the highly active volcano Piton de la Fournaise. Prior seismological studies have concentrated on the Piton de la Fournaise, whereas the Piton des Neiges is seismologically unexplored. We use phase cross-correlations of ambient seismic noise of 23 stations installed on La Réunion yielding a total data set of 253 station pairs. From these noise correlations, Rayleigh-wave group velocities with periods of 1-10 s are derived for each station pair. We identify 212 out of the 253 station pairs as high-quality group velocity estimates. Gravity and electrical resistivity studies proposed a hypovolcanic complex beneath the Piton des Neiges. Inspection of the group velocity curves indicates that waves that propagate through the proposed dense body are characterized by significantly increased wave speed. This observation provides the first seismological indication for the existence of a high-velocity body below the Piton des Neiges. As future work, an inversion of the selected high-quality group velocities is planned to retrieve a 3-D image of the island. Such a noise tomography will probably reinforce the existence of a high-velocity body beneath the Piton des Neiges

and could constrain its horizontal and vertical extent. Furthermore, a tomographic model of the island can be used for many future studies like local earthquake tomography and seismicity re-localisation.

4.1 Introduction

The volcanic island of La Réunion is located in the western Indian Ocean, 800 km east of Madagascar and 200 km southwest of Mauritius. The hotspot volcanism of the island stems probably from an underlying deep mantle plume, a hypothesis that is reinforced by a hotspot track that originates from the Deccan Traps in India and leads time-progressively to La Réunion (Courtillot et al., 2003).

The island is dominated by two large volcanic complexes: the dormant volcano Piton des Neiges in the northwest and the active volcano Piton de la Fournaise in the southeast of the island (Fig. 4.1). While the latest lavas of the Piton des Neiges are dated back to 12 ka (Deniel et al., 1992), the Piton de la Fournaise is a highly active volcano with one eruption every 10 months on average (Stieltjes & Moutou, 1989). This high activity is the reason why seismic studies have focused mainly on the Piton de la Fournaise so far (e.g Brenguier et al., 2007; Duputel et al., 2009; Rivet et al., 2014; Sens-Schönfelder et al., 2014; Nakata et al., 2016). In contrast to that, only a few seismic studies investigated the entire island of La Réunion (e.g. Gallart et al., 1999).

Information about the island's internal structure has gained either from geological surveys (e.g. Demange et al., 1989) or from other geophysical measurements like gravity, magnetic and electrical resistivity (e.g. Malengreau et al., 1999; Lénat et al., 2001; Gailler & Lénat, 2012). These studies showed that the Piton des Neiges is underlain by a dense hypovolcanic complex (Gailler & Lénat, 2012), a structure that has not been seismologically investigated so far. The extent of this complex correlates with three large overlying depressions (Cirques of Mafate, Cilaos, and Salazie) indicating that the formation of these Cirques is linked to the complex (Gailler & Lénat, 2012).

We present the first group velocity study of the entire island of La Réunion. We use cross-correlations of ambient seismic noise to reconstruct the Green's function between two seismic stations. The Green's function reflects the underground properties between both stations; convergence towards the Green's function is achieved by stacking the cross-correlation functions (CCFs). Instead of classical CCFs, we use phase cross-correlations (PCCs) that are based on the phase coherence of two signals (Schimmel et al., 2011). We calculate group velocity curves from stacked PCCs, which seem to support the findings of other geophysical studies like Gailler & Lénat (2012). Group velocity estimates retrieved with the same method for a 2000×2000 km^2 broad region around La Réunion are presented in Hable et al. (2019a).

Section 4.2 describes the data set used in this study. Section 4.3 gives a short overview over our methods of PCC calculation, stacking procedure and group velocity computation. In Section 4.4, we present the group velocity curves of La Réunion. Section 4.5 is a discussion, that includes the comparison with other studies. Our conclusions are given in Section 4.6.

We used the open-source toolboxes Python (Rossum, 1995), ObsPy (Beyreuther et al., 2010; Megies et al., 2011; Krischer et al., 2015), and obspyDMT (Hosseini & Sigloch, 2017) for data downloading, processing, and plotting all figures except Fig. 4.1, which was generated with the open-source mapping toolbox GMT (Wessel et al., 2013).

4.2 Data description

We use data of 23 temporary and permanent land stations installed on La Réunion (Fig. 4.1). 10 of these stations belong to the Réunion Hotspot and Upper Mantle - Réunions Unterer Mantel (RHUM-RUM) experiment, that was carried out between 2011 and 2015. In order to investigate the crust and the mantle beneath La Réunion, a large number of ocean bottom seismometers (OBSs) and land stations were deployed on and around the island in course of this project (Barruol & Sigloch, 2013; Stähler et al., 2016). The 10 RHUM-RUM land stations on La Réunion operated roughly three years, from mid-2012 to mid-2015 (Barruol et al., 2017), which defines our studied period. As this is the first time that the 10 RHUM-RUM stations are used for an imaging study, we evaluated the noise levels of these stations. For this purpose, we calculate probabilistic power spectral density (PPSD) plots for each component according to McNamara & Buland (2004). These plots are published as supplementary material to this article.

All stations show noise levels below but close to the New High Noise Model defined by Peterson et al. (1993). For further details, we refer to the supplement with Figs 4.5, 4.6, and 4.7.

12 stations of our study are part of the seismic network operated by the Observatoire Volcanologique du Piton de la Fournaise (OVPF). The stations of this seismic network concentrate on the Piton de la Fournaise, but the array has been expanded over the entire island at approximately the same time of the RHUM-RUM land station installation. Since we aim for imaging the entire island, we chose only 6 OVPF stations of the dense

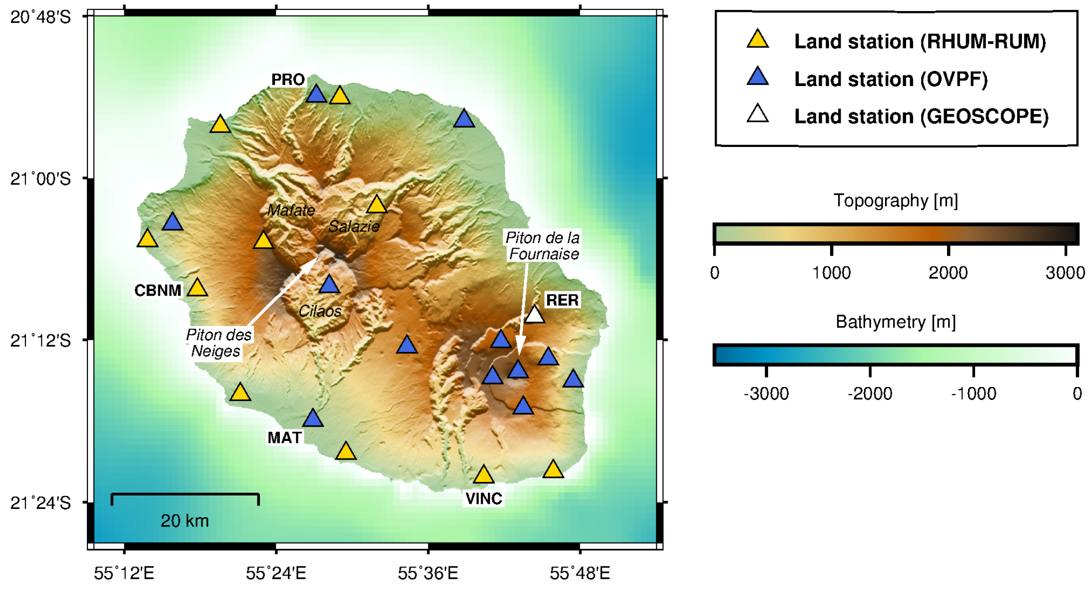


Figure 4.1: Map of the land stations on La Réunion used in this study. Yellow colour denotes stations installed during the RHUM-RUM experiment (Barruol et al., 2017), blue indicates stations operated by the OVPF, and white represents a permanent seismic station of the GEOSCOPE network (IPGP & EOST, 1982). Stations that are explicitly mentioned in the text are labelled with their station name. The northwest of the island is dominated by the large volcanic complex of the Piton des Neiges, whose summit is located in the middle of three large depressions (Cirques of Mafate, Cilaos, and Salazie). The active volcano Piton de la Fournaise is located in the southeast of the island.

Piton de la Fournaise array for our study. The remaining station of our study (RER) belongs to the permanently installed GEOSCOPE network (IPGP & EOST, 1982).

A group velocity study requires a correct timing of the used seismograms. For seismic land stations, a frequent synchronization with a GPS signal as reference clock should ensure a proper timing. However, Hable et al. (2018) detected that 4 of the 10 RHUM-RUM land stations on La Réunion were affected by clock errors due to a data logger failure. These clock errors were measured with noise cross-correlations of high temporal resolution of one day. Clock drifts (between -2.1 ms/day and +2.9 ms/day) as well as large clock jumps (up to 6 min) could be detected and successfully corrected. Hable et al. (2018) demonstrated that the accuracy of the cross-correlation method can be significantly improved to 10-20 ms by using an average over multiple component and station pairs. Moreover, the study of Hable et al. (2018) includes an extensive survey of the timing errors in RHUM-RUM OBS seismograms.

4.3 Methodology

Here, we provide a short method description, while a more comprehensive summary of the PCC calculation, stacking algorithm and group velocity estimation is given by Hable et al. (2019a), who used the same methods for retrieving group velocity results of an extended area around La Réunion. We used the software package of Schimmel et al. (2011) for both PCC computation and stacking. The group velocity calculation is based on the algorithm developed by Schimmel et al. (2017).

4.3.1 PCC calculation

Noise CCFs between two stations are a widely used method to explore crustal structures (e.g. Shapiro & Campillo, 2004; Sabra et al., 2005b; Shapiro et al., 2005). Its popularity comes from the convergence of the CCFs towards the Green's function, which represents the seismogram recorded at one station as response to a dirac delta function at another station. Thus, one can infer to the underground structures between the two points from the Green's function.

We use a cross-correlation technique that is based on the phase alignment of two signals (yielding PCCs) instead on the alignment of high amplitudes (yielding classical CCFs), since the usage of PCCs can enhance the convergence to the Green's function (Schimmel, 1999; Schimmel et al., 2011). Moreover, this correlation technique has the advantage that no complex pre-processing is required. For further information, we refer to Schimmel et al. (2011).

Our PCC calculation is based on day-long seismic records of the vertical component. Before the correlation process, we downsample the day-long traces to 4 Hz to diminish PCC computation time. After that, we correct for the instrument response and apply a demeaning and linear detrending. We calculate PCCs of daily traces that are filtered with two zerophase bandpass filters: 1-5 s and 5-10 s. Haned et al. (2016) and Hable et al. (2019a) showed that the usage of several smaller period bands instead of one broader band can enhance the group velocity estimation.

The PCC computation comes from the definition of the analytic signal $s(t)$ of a real time series. This analytic signal can be described by the envelope $a(t)$ and the instantaneous phase $\phi(t)$ of the waveform (Schimmel et al., 2011):

$$s(t) = a(t) \cdot e^{i\phi(t)}. \tag{4.1}$$

The PCC calculation is defined as follows (Schimmel, 1999; Schimmel et al., 2011):

$$PCC(t) = \frac{1}{2T} \sum_{\tau=\tau_0}^{\tau_0+T} \left(\left|e^{i\phi_1(t+\tau)} + e^{i\phi_2(\tau)}\right|^{\nu} - \left|e^{i\phi_1(t+\tau)} - e^{i\phi_2(\tau)}\right|^{\nu} \right), \tag{4.2}$$

with ϕ_1 and ϕ_2 representing the instantaneous phases of two stations (s_1 and s_2). The terms t, τ_0, and T indicate the lapse time, the start time and the correlation window length of the PCC, respectively. We use ν=1, as suggested by Schimmel et al. (2011).

According to eq. (4.2), we compute PCCs of day-long, pre-processed traces for each possible station pair and for the two period bands (1-5 s, 5-10 s). The PCC's lapse time goes from -100 s to +100 s. The acausal part of a CCF (negative lapse times) represents the seismogram recorded at station s_1 as response to an impulse at s_2. Corresponding to that, the causal part of a CCF (positive lapse times) reflects waves travelling in opposite direction (from s_1 to s_2). The stacking over both the causal CCF part and the time-reversed acausal part provides an enhancement of the signal-to-noise ratio of the resulting CCF stack. Thus, we split each calculated PCC in its causal and acausal part and include both parts in our stacking procedure. Since we use seismic records of a $\sim$3-year long period (from mid-2012 to mid-2015), we can make use of more than 2000 individual PCCs (2 PCCs × 365 days × 3 years) for a PCC stack, when both correlated stations record data during the entire period.

4.3.2 Phase-weighted stacking

For calculation of the PCC stacks, we use a time-frequency domain phase-weighted stacking (tf-PWS) developed by Schimmel et al. (2011) instead of conventional linear stacking technique. Schimmel et al. (2011) as well as Corela et al. (2017) demonstrated that the tf-PWS procedure yields stacks with significantly enhanced signal-to-noise ratio compared to linear stacks. The definition of tf-PWS is given by the multiplication of the phase coherence $c_{ps}(\tau, f)$ with the S-transform of the linear stack $S_{ls}(\tau, f)$ of all PCCs (Schimmel et al., 2011):

$$S_{pws}(\tau, f) = c_{ps}(\tau, f) \cdot S_{ls}(\tau, f) = \left| \frac{1}{N} \sum_{j=1}^{N} \frac{S_j(\tau, f) e^{i2\pi f \tau}}{|S_j(\tau, f)|} \right|^{\nu} \cdot S_{ls}(\tau, f). \tag{4.3}$$

N indicates the number of all PCCs; $S_j(\tau, f)$ describes the S-transform of an individual PCC j. The S-transform is used to transform the PCCs in time-frequency domain (τ, f)

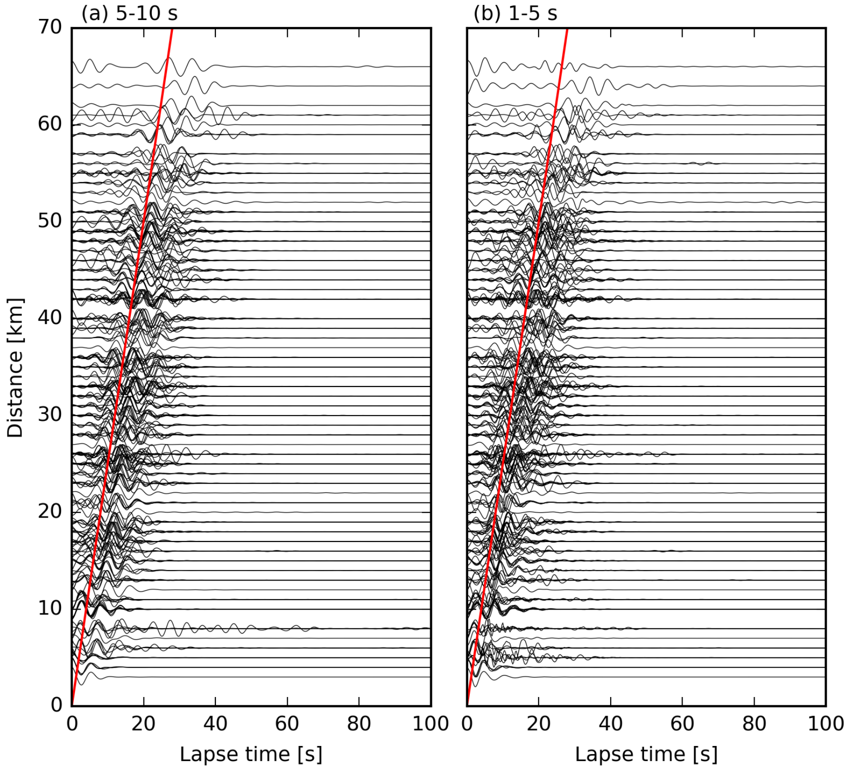


Figure 4.2: PCC stacks as a function of interstation distance for the period bands (a) 5-10 s and (b) 1-5 s. Rayleigh waves with a velocity of roughly 2.5 km/s indicated by the red line are clearly visible in both bands.

(Stockwell et al., 1996). We follow Schimmel et al. (2017), who retrieved best results for ν=2.

The number of all possible station pairs n is determined by the number of used stations k:

$$n = \frac{k(k-1)}{2}. \tag{4.4}$$

In our study, we yield 253 station pairs considering that we make use of the data of 23 seismic stations. Fig. 4.2 shows the PCC stacks calculated according eq. (4.3) for each of the 253 station pairs as a function of interstation distance. Clear wave arrivals can be observed in the period band of 5-10 s (Fig. 4.2a) as well as in the band of 1-5 s (Fig. 4.2b). This wave package can be attributed to Rayleigh waves, as we are working with correlations of the vertical component. The velocity of this Rayleigh wave package is roughly 2.5 km/s (see red line in Fig. 4.2).

4.3.3 Group velocity calculation

Schimmel et al. (2017) developed an algorithm to measure group velocities from PCC stacks in time-frequency domain. This algorithm determines the arrival time of PCC amplitude maxima as a function of frequency f. The group velocity $v_g(f)$ can be inferred from the group arrival time $t(f)$ of the maxima using the interstation distance d (Schimmel et al., 2017):

$$v_g(f) = \frac{d}{t(f)}. \tag{4.5}$$

In some cases, the localisation of a maximum is not possible, for example due to spectral holes. We refer to Schimmel et al. (2017) for a detailed description of the algorithm.

We use the algorithm of Schimmel et al. (2017) to extract group velocity values in the velocity range of 0.5-5-5 km/s. Two station pairs (CBNM-VINC, MAT-PRO) are presented as examples in Fig. 4.3. The investigated velocity range is indicated by the blue-shaded area in the PCC plots, where the dashed red line indicates a velocity of 2.5 km/s (cf. red line in Fig. 4.2). The measured group velocities are presented as vertically normalized amplitude spectrum, where each frequency column contains the maximum value 1 indicated by the yellow colour. The black dots depict the velocity of the maximum amplitude, while black dashes define group velocity values with 95% of maximum amplitude.

4.4 Results: group velocity curves

We select those group velocity values that we consider to be of sufficiently high quality for tomographic imaging. These values are indicated in Fig. 4.3 by red dots (superimposed on the black ones). Only 41 station pairs are excluded from our total data set of 253 station pairs. In most cases, the reason is the too short interstation distance between the stations (e.g. very closely spaced stations on the Piton de la Fournaise), which prevents a proper group velocity estimation. Fig. 4.4 shows that the extracted velocity curves are characterized by a large variability showing velocity values between 1.3 km/s and 3 km/s.

Prior gravity and electrical resistivity studies showed a dense hypovolcanic complex below the Cirques of Mafate, Cilaos, and Salazie around the Piton des Neiges (e.g. Demange et al., 1989; Malengreau et al., 1999; Gailler & Lénat, 2012). Such a body would result in significantly higher wave velocities. Thus, we sort the group velocity curves

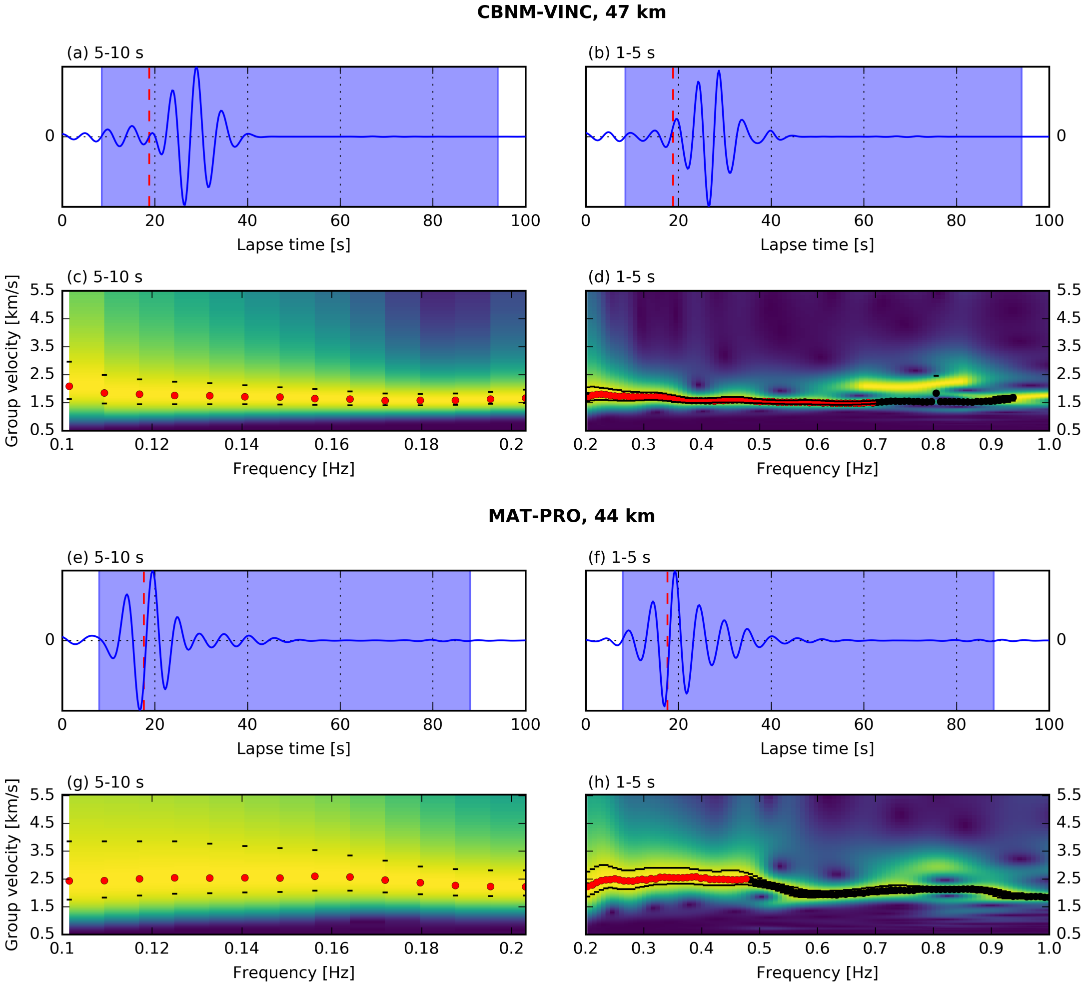


Figure 4.3: PCC stacks and group velocity estimations for (a)-(d) CBNM-VINC (distance 47 km), and for (e)-(h) MAT-PRO (distance 44 km). The blue-shaded areas in (a), (b), (e), and (f) denote the velocity range of 0.5-5.5 km/s, for which the group velocity is estimated in (c), (d), (g), and (h). The amplitude of the group velocity estimation is vertically (frequency) normalized (yellow and blue colours indicate high and low amplitudes, respectively). Black dots mark the group velocity selected by the algorithm of Schimmel et al. (2017), black bars denote the 95 % amplitude range of the selected values. Red dots (superimposed on the black ones) show the group velocity values we extracted for tomography.

into different groups. The first group contains all velocity values of station pairs where the waves propagate through the proposed high-density body (red colour in Fig. 4.4). An example is station pair MAT-PRO (Fig. 4.3e-h). The second group includes those station pairs where the wave propagation takes place far from the dense body (yellow colour in Fig. 4.4), like for CBNM-VINC (Fig. 4.3a-d). The last group comprises station pairs which can not clearly assigned to group one or two (orange colour in Fig. 4.4), as the waves travel probably close to the edges of the hypovolcanic complex. Fig. 4.4 demonstrates clearly that high group velocities (∼2-3 km/s) can be associated with station pairs of group one, while group two shows remarkably reduced velocity values

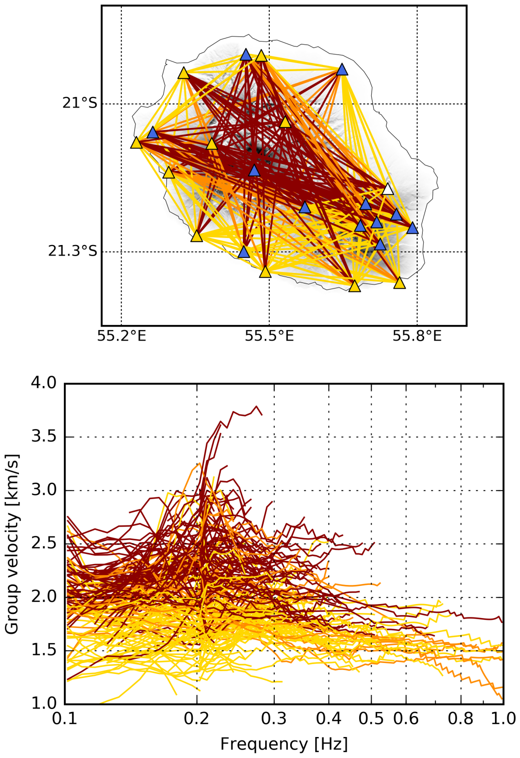


Figure 4.4: Selected group velocity curves that could contain usable information for tomography (cf. red dots in Fig 4.3). Group velocities of station pairs that intersect the proposed dense hypovolcanic complex below the Piton des Neiges are represented in red colour (e.g. MAT-PRO from Figs 4.3e-h), while station pairs that clearly do not intersect this complex are indicated by yellow (e.g. CBNM-VINC from Figs 4.3a-d). Station pairs which can not be associated clearly with one of these groups are given in orange. The wave paths with the corresponding colouring are displayed in the map above. The triangles represent the station locations, as described in Fig. 4.1. The high-velocity values of the red curves indicate that a high-velocity body exists below the Piton des Neiges.

(∼1.3-2 km/s). This observation provides the first seismological indication for the existence of a high-velocity body below the Piton des Neiges. Note that the high ray coverage in Fig. 4.4 illustrates that the entire island of La Réunion can be imaged by our data set.

4.5 Discussion

We present the first seismological survey of the entire island of La Réunion by providing group velocity curves between 1 s and 10 s based on PCCs between island stations. We identify 212 group velocity curves from our total data set of 253 station pairs as of sufficiently high quality for further seismic imaging.

A future tomographic inversion of these group velocity curves is planned to derive a 3-D tomographic model of the entire island. For now, we attempt to infer seismic structures by inspection of the group velocity curves.

A hypovolcanic complex of a diameter of ~20 km has been suggested by gravity and electrical resistivity studies (e.g. Demange et al., 1989; Malengreau et al., 1999; Gailler & Lénat, 2012) beneath the Piton des Neiges and surrounding depressions (Cirques of Mafate, Cilaos, and Salazie). Our study shows that ray paths between stations that intersect this proposed, dense body are generally characterized by high group velocities (above ~2 km/s), otherwise by rather low velocities (below ~2 km/s). The complex's composition of gabbros and ultrabasic cumulates of high density (Demange et al., 1989) accounts very likely for the high-velocity signature that we observe. Gailler & Lénat (2012) highlighted the strikingly consistent lateral extent of the hypovolcanic complex with the overlying topographic depressions, which argues for a potential genetic relation between the formation of the Cirques and the complex. A future seismic tomography model based on the derived group velocities could help to constrain the extent of the proposed high-velocity body beneath the Piton des Neiges especially in vertical direction.

Gallart et al. (1999) used lines from air-gun shots to derive a 2-D P-velocity model along a 340 km NE-SW transect through La Réunion. They found a high-velocity body in the northeast of the island, but it is unclear if this body can be associated with the hypovolcanic complex beneath the Piton des Neiges, as the investigated transect did not pass the proposed complex.

Brenguier et al. (2007) used noise data of a dense seismic network on the Piton de la Fournaise to model the interior of the volcano. They found an intrusive high-velocity body with a diameter of up to 3 km that extends from surface (~2000 m elevation) down to 1 km below sea level. This finding was supported by Mordret et al. (2015), who expanded the work of Brenguier et al. (2007) using an increased number of seismic stations for noise tomography below the Piton de la Fournaise. Such a small-scale structure can probably not be resolved with our study, as we concentrate on larger distances. The absence of a high-velocity body beneath the Piton de la Fournaise in our future model would be in agreement with large-scale gravity studies of Malengreau et al. (1999) and Gailler & Lénat (2012), who could not detect a large gravity anomaly below the volcano.

Instead, Malengreau et al. (1999) and Gailler & Lénat (2012) found a dense body at the east coast of the island. This structure can probably be associated with the buried ancient Les Alizés volcano (Lénat et al., 2012). Mordret et al. (2015) detected a high-velocity body at the margin of their velocity model that might potentially be part of this dense body. However, the location of this buried complex close to the coast of La Réunion had prevented a clear illumination with island stations. Likewise, the group velocity curves of our study will probably not be able to image the Les Alizés body thoroughly, as our data set is not suited to image structures outside the used seismic network.

4.6 Conclusions

The ancient volcano Piton des Neiges that dominates the northwestern part of the island La Réunion has not been seismologically imaged so far, as seismic studies concentrated on the active volcano Piton de la Fournaise in the southeast of the island. In order to image the internal structure of the entire island, we use cross-correlations of ambient seismic noise. Rayleigh-wave group velocities between 1 s and 10 s are calculated from noise cross-correlations of 253 station pairs. We identify 212 high-quality group velocity curves that can be used for a noise tomography.

These group velocity curves indicate an extended high-velocity body beneath the Piton des Neiges and the surrounding depressions (Cirques of Mafate, Cilaos, and Salazie). The extent of this body (~20 km in diameter) probably coincides with a dense, hypovolcanic complex that was found by gravity and electrical resistivity studies (Malengreau et al., 1999; Gailler & Lénat, 2012). This is the first indication of a hypovolcanic body beneath the Piton des Neiges by a seismological survey.

In future, an inversion of the group velocities for depth is planned to retrieve a high-resolution 3-D tomographic model of the entire island. A challenging aspect of the inversion may the rough topography (elevation of up to >3000 m) of the island. The effect of topography on noise cross-correlations is largely unknown, but may be measurable for small periods of <5 s (Köhler et al., 2012).

Such a noise tomography model of La Réunion can be used for many seismological applications. For instance, it can be used as starting model for a local earthquake tomography, which may refine small-scale structures. Furthermore, it may allow for the re-localisation of local events on La Réunion. A potential clustering of events may indicate known or even unknown faults, that have been undetected so far due to the widely unknown velocity structure of the northwestern part of the island.

4.7 Supplement

10 temporary land stations were installed on La Réunion during the RHUM-RUM experiment. They operated roughly three years, from mid-2012 to mid-2015. We calculate probabilistic power spectral density (PPSD) plots of these stations in order to assess their noise level. These plots are given in Figs 4.5, 4.6, and 4.7 for the vertical (HHZ), east (HHE), and north (HHN) component, respectively. An extensive description of PPSDs is given by McNamara & Buland (2004). In principle, the power spectral densities (PSDs) are calculated for individual segments and afterwards merged to a probabilistic plot. We split the data records into 1-hour long segments with a segment overlap of 50%. The number of used segments is given in the title of each PPSD plot together with the station's identity code and the start and end time of the recording period. The colouring of the PPSD plots indicates how frequent a noise level occurs among the individual PSDs, where yellow represents frequent occurrence. As reference curves, the New High Noise Model (NHNM) and New Low Noise Model defined by Peterson et al. (1993) are given as grey lines. The green bar below a PPSD plot represents the data availability, while the blue bar shows if the available data are used for the PPSD calculation. Thus, white colour represents gaps in the data.

Stations CBNM, MAID, POSS, and SALA are equipped with a Nanometrics Trillium 240 seismometer and a RefTek RT 130 data logger. Stations ETAN, SGIL, STPI, STPHI, and VINC are composed of a Nanometrics Trillium Compact seismometer and a Nanometrics Taurus data logger. Station RUN01 belongs to the Université de La Réunion and recorded data before and after the RHUM-RUM experiment with potentially different station settings. During the RHUM-RUM recording period (2013-04-09 to 2015-07-02), RUN01 is equipped with a Güralp CMG3-ESPC 120 s sensor and a Nanometrics Taurus data logger.

All stations show noise levels very close to the NHNM. Exceptions are the vertical component of CBNM, MAID, POSS, and SALA, which show very low noise levels for periods >10 s (Fig. 4.5). This can be attributed to the lower self-noise of the Trillium 240 sensors for longer periods compared to Trillium Compact instruments. In addition, the vertical channel of MAID shows a significantly reduced noise level below 0.5 s probably due to the station's location in a quiet, mountainous area, which is barely affected by high-frequency human noise sources.

For periods below ∼1 s and above ∼10 s a blurred noise level can be observed, which sometimes appears as two bands (e.g. ETAN in Fig. 4.5 or POSS in Fig. 4.7). For shorter periods (<1 s), this observation is caused by increased cultural noise during

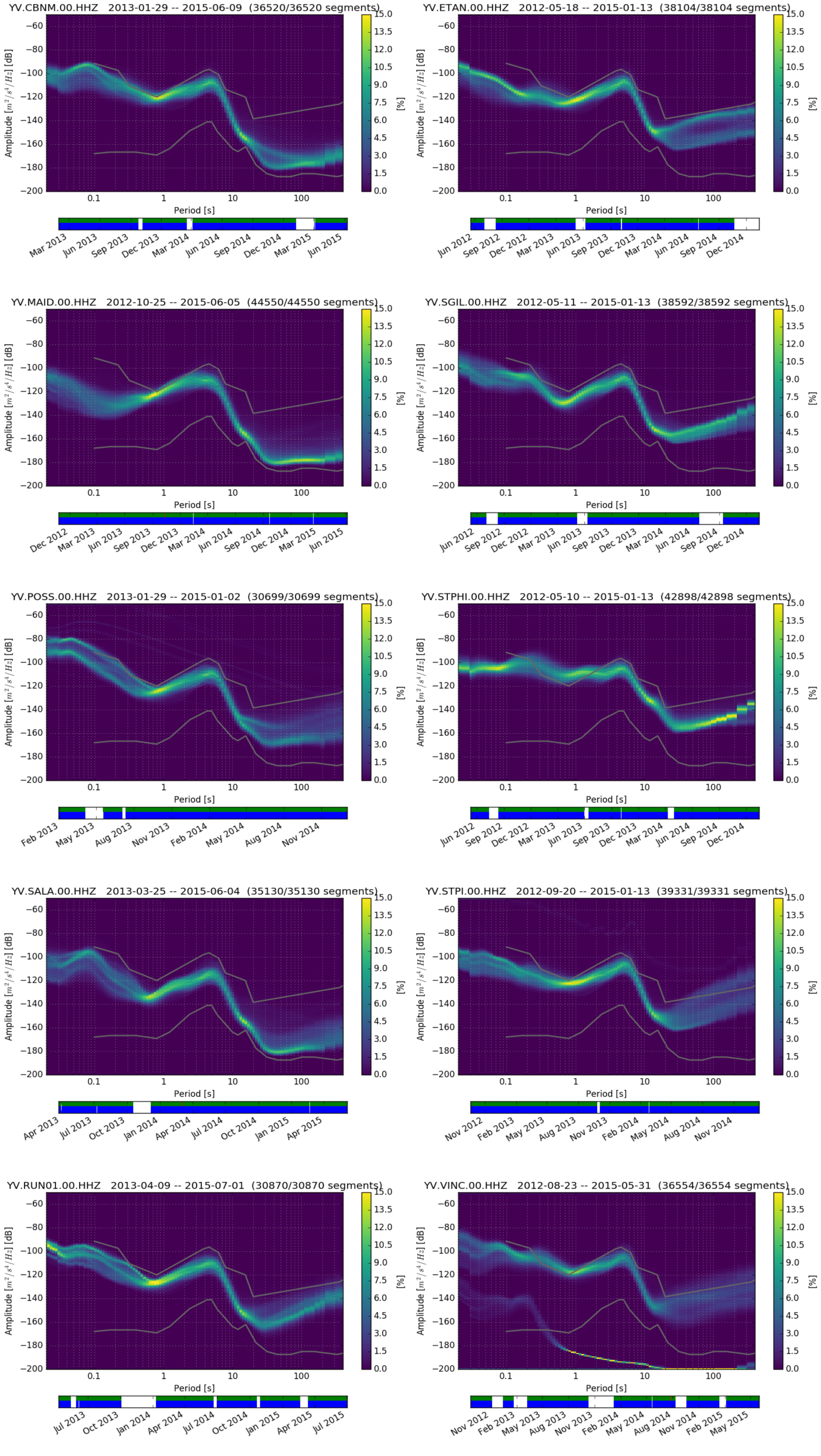


Figure 4.5: PPSD plots for 10 RHUM-RUM land stations (vertical component HHZ). See text for details.

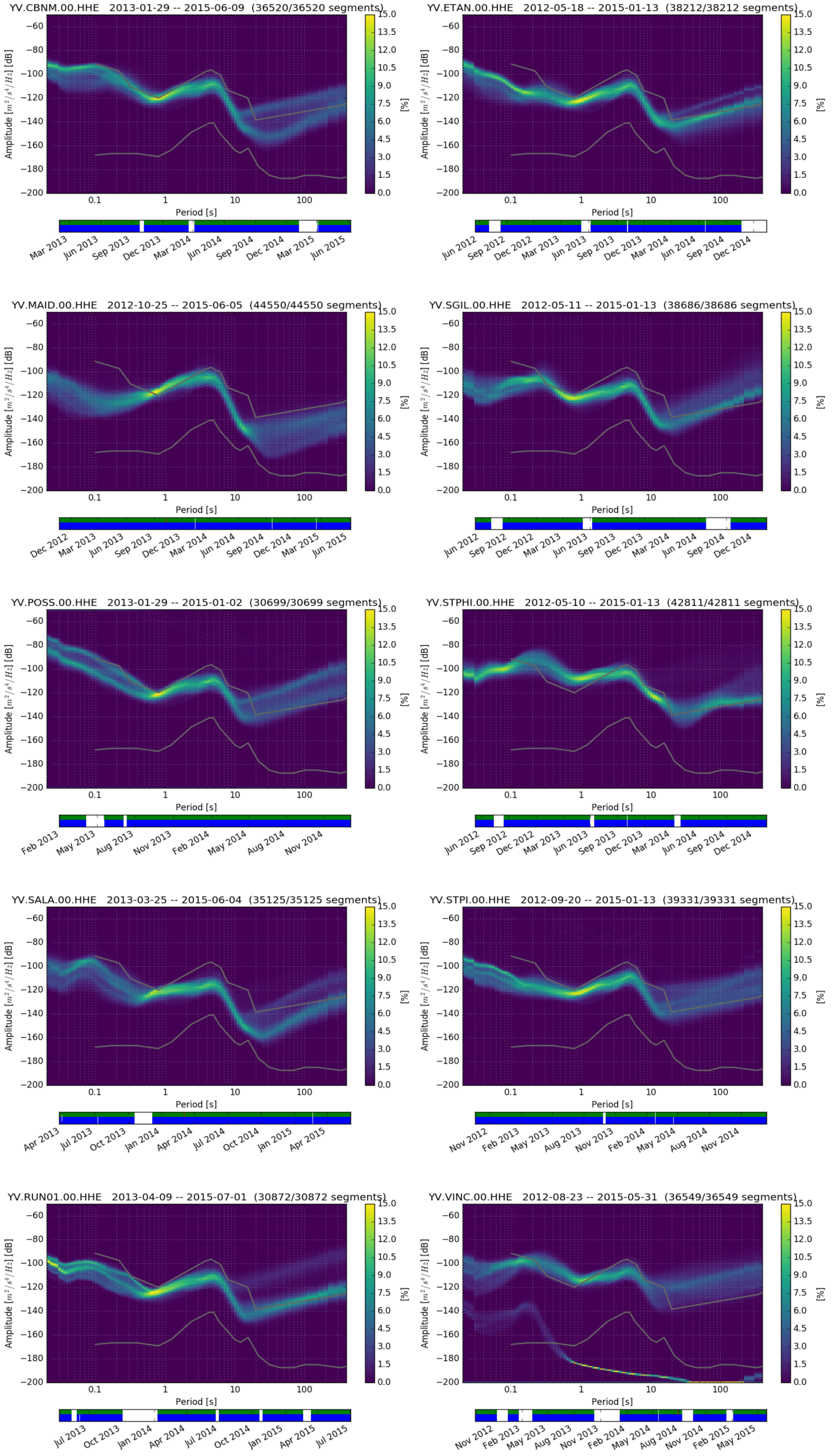


Figure 4.6: PPSD plots for 10 RHUM-RUM land stations (east component HHE). See text for details.

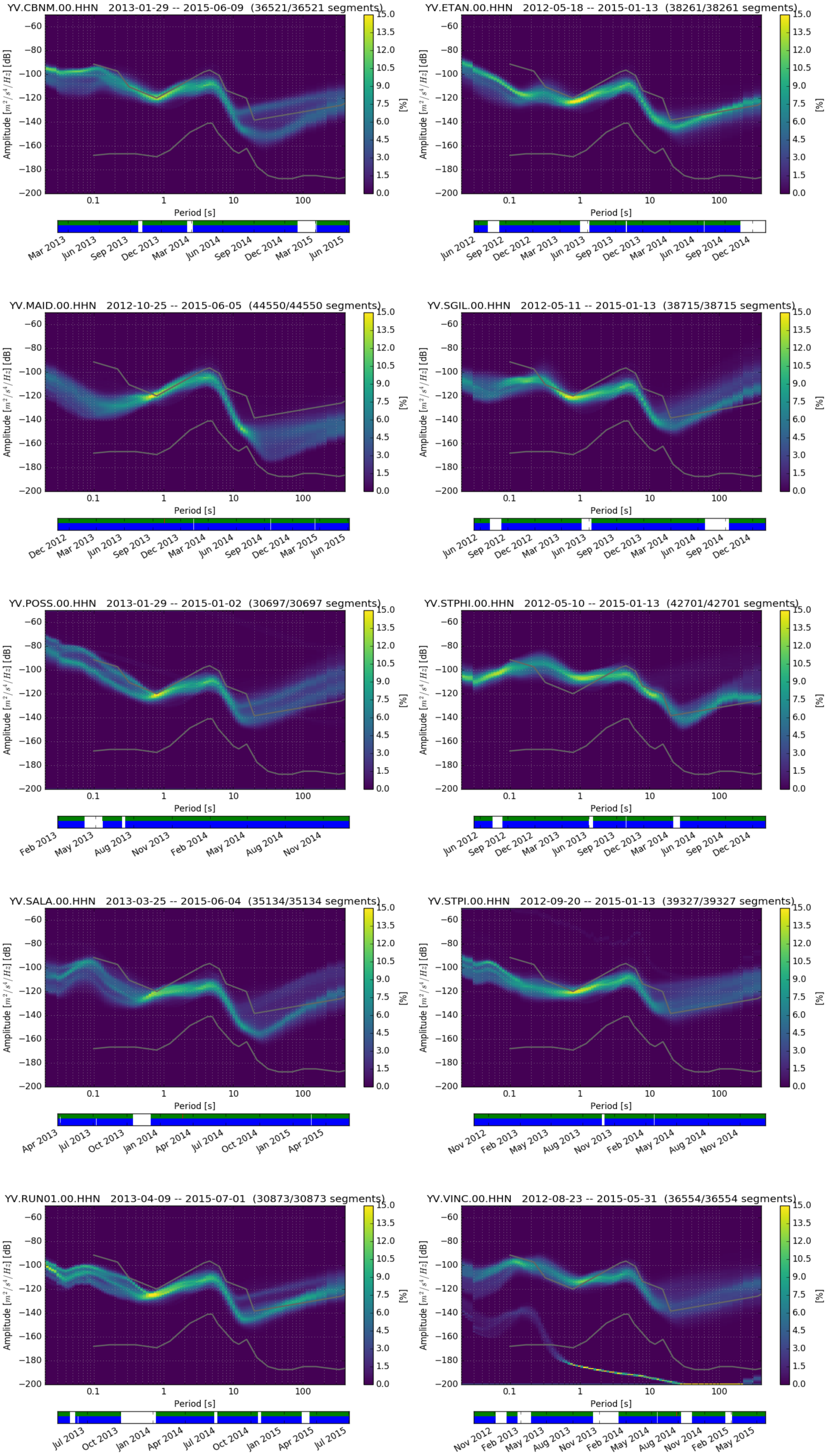


Figure 4.7: PPSD plots for 10 RHUM-RUM land stations (north component HHN). See text for details.

human working hours and decreased human noise during night. For longer periods (>10 s), the blurred signal can be most likely explained by diurnal temperature variations (McNamara & Buland, 2004).

Chapter 5

Conclusions & Outlook

Cross-correlations of ambient seismic noise provide a wide range of applications like clock error corrections and tomographic imaging of crustal structure. The scope of this thesis was to extend the limitations of prior clock error studies, where temporal resolution and timing accuracy of the estimated clock errors suffered from large interstation distances (∼200 km). Furthermore, this work intended to prove the usability of OBS records for noise tomography with large interstation distances (∼2000 km). Finally, the island La Réunion should be seismologically investigated by noise cross-correlations, as no seismic velocity model is available for the entire island so far.

Chapter 2 of this thesis demonstrated that the clock error detection highly benefits from using a multi-component procedure instead of a conventional single-component approach. The average over multiple components enhances the clock error accuracy severalfold to ∼20 ms standard deviation corresponding to one sample at a sampling rate of 50 Hz. Furthermore, the multi-component approach allows a clock error estimation with high temporal resolution (1-2 days) for interstation distances of up to ∼300 km. These findings will clearly enhance the usability of past and future data sets that are affected by timing problems. The presented clock error study allowed the usability of OBSs installed in the course of the RHUM-RUM project for structural imaging. This work was presented in Chapter 3. Furthermore, clock corrections of land stations that were a crucial part of the work presented in Chapter 4 were provided.

Chapter 3 of this thesis showed that OBS noise cross-correlations with interstation distances of up to ∼2000 km can be used to derive high-quality group velocity measurements. The joint inversion of these group velocities with the earthquake data set of Mazzullo et al. (2017) provided a high-resolution S-wave velocity model of the crust and the uppermost mantle of the western Indian Ocean. The imaging of crustal structure is significantly improved compared to the original model of Mazzullo et al. (2017), which indicates the important information gain due to the usage of noise group velocities. These results demonstrate that even large-scale velocity models of ocean basins can be envisaged with OBS noise tomography.

Chapter 4 of this thesis presented group velocities of the island La Réunion based on noise cross-correlations of island stations. The group velocity curves provide the first seismological indication of a large high-velocity body beneath the ancient volcano Piton des Neiges. This body can be associated with a basaltic, hypovolcanic complex proposed by gravity studies. Inversion of the group velocities will yield a high-resolution tomographic model of the island, which can be used as starting model for local earthquake tomography or for re-localisation of local earthquakes.

The data quality inspection by evaluating potential clock errors was a crucial requirement for the subsequent tomographic studies, which rely on an accurate timing of the seismic records. Thus, this study showed the high applicability of noise cross-correlations within the same data set. The detailed method descriptions of the clock error study as well as of the group velocity calculations provide a valuable guideline for future studies, that deal with land or OBS seismograms.

Bibliography

Aki, K., 1957. Space and time spectra of stationary stochastic waves, with special reference to microtremors, *Bull. Earthq. Res. Inst.*, **35**, 415-456.

Aki, K., 1965. A note on the use of microseisms in determining the shallow structures of the Earth's crust, *Geophysics*, **30**(4), 665-666, doi:10.1190/1.1439640.

Aki, K. & Richards, P.G., 1980. *Quantitative seismology: Theory and methods*, San Francisco: W.H. Freeman.

Anchieta, M.C., Wolfe, C.J., Pavlis, G.L., Vernon, F.L., Eakins, J.A., Solomon, S.C., Laske, G. & Collins, J.A., 2011. Seismicity around the Hawaiian Islands recorded by the PLUME seismometer networks: insight into faulting near Maui, Molokai, and Oahu, *Bull. Seismol. Soc. Am.*, **101**(4), 1742-1758, doi:10.1785/0120100271.

Ardhuin, F., Stutzmann, E., Schimmel, M. & Mangeney, A., 2011. Ocean wave sources of seismic noise, *J. Geophys. Res.*, **116**, C09004, doi:10.1029/2011JC006952.

Ball, J.S., Sheehan, A.F., Stachnik, J.C., Lin, F.-C., Yeck, W.L. & Collins, J.A., 2016. Lithospheric shear velocity structure of South Island, New Zealand, from amphibious Rayleigh wave tomography, *J. Geophys. Res. Solid Earth*, **121**, 3686-3702, doi:10.1002/2015JB012726.

Barruol, G. & Sigloch, K., 2013. Investigating La Réunion hot spot from crust to core, *Eos, Trans. Am. Geophys. Union*, **94**(23), 205-207, doi:10.1002/2013EO230002.

Barruol, G., Reymond, D. , Fontaine, F.R., Hyvernaud, O., Maurer, V. & Maamaatuaiahutapu, K., 2006. Characterizing swells in the southern Pacific from seismic and infrasonic noise analyses, *Geophys. J. Int.*, **164**(3), 516-542, doi:10.1111/j.1365-246X.2006.02871.x.

Barruol, G., Sigloch, K. & RHUM-RUM Group, 2017. RHUM-RUM experiment, 2011-2015, code YV (Réunion Hotspot and Upper Mantle – Réunion's Unterer Mantel) funded by ANR, DFG, CNRS-INSU, IPEV, TAAF, instrumented by DEPAS, INSU-OBS, AWI and the Universities of Muenster, Bonn, La Réunion. RESIF - Réseau Sismologique et géodésique Français, doi:10.15778/resif.yv2011.

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**(3), 1239-1260, doi:10.1111/j.1365-246X.2007.03374.x.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. ObsPy: a Python toolbox for seismology, *Seismol. Res. Lett.*, **81**(3), 530-533, doi:10.1785/gssrl.81.3.530.

Bonnefoy-Claudet, S., Cotton, F. & Bard, P.-Y., 2006. The nature of noise wavefield and its applications for site effects studies: a literature review, *Earth Sci. Rev.*, **79**(3), 205-227, doi:10.1016/j.earscirev.2006.07.004.

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, *Geophys. Res. Lett.*, **34**(2), L02305, doi:10.1029/2006GL028586.

Brenguier, F., Shapiro, N.M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O. & Nercessian, A., 2008a. Towards forecasting volcanic eruptions using seismic noise, *Nat. Geosci.*, **1**, 126-130, doi:10.1038/ngeo104.

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**(5895), 1478-1481, doi:10.1126/science.1160943.

Bromirski, P.D. & Duennebier, F.K., 2002. The near-coastal microseism spectrum: Spatial and temporal wave climate relationships, *J. Geophys. Res. Solid Earth*, **107**(B8), doi:10.1029/2001JB000265.

Bromirski, P.D., Stephen, R.A. & Gerstoft, P., 2013. Are deep-ocean-generated surface-wave microseisms observed on land?, *J. Geophys. Res. Solid Earth*, **118**(7), 3610-3629, doi:10.1002/jgrb.50268.

Capon, J., Greenfield, R.J. & Kolker, R.J., 1967. Multidimensional maximum-likelihood processing of a large-aperture seismic array, *Proc. IEEE*, **55**, 192-211, doi:10.1109/PROC.1967.5439.

Corela, C., Silveira, G., Matias, L., Schimmel, M. & Geissler, W.H., 2017. Ambient seismic noise tomography of SW Iberia integrating seafloor- and land-based data, *Tectonophysics*, **700**, 131-149, doi:10.1016/j.tecto.2017.02.012.

Courtillot, V., Davaille, A., Besse, J. & Stock, J., 2003. Three distinct types of hotspots in the Earth's mantle, *Earth Planet. Sci. Lett.*, **205**(3), 295-308, doi:10.1016/S0012-821X(02)01048-8.

Davy, C., Stutzmann, E., Barruol, G., Fontaine, F.R. & Schimmel, M., 2015. Sources of secondary microseisms in the Indian Ocean, *Geophys. J. Int.*, **202**(2), 1180-1189, doi:10.1093/gji/ggv221.

Debayle, E. & Lévêque, J.J., 1997. Upper mantle heterogeneities in the Indian Ocean from waveform inversion, *Geophys. Res. Lett.*, **24**(3), 245-248, doi:10.1029/96GL03954.

De Boor, C., 1978. *A practical guide to splines: Applied Mathematical Sciences*, Vol. 27, pp. 109-112, Springer-Verlag.

Demange, J., Chovelon, P. & Puvilland, P., 1989. Geothermal model of the Salazie Cirque (Reunion Island): Volcanic and structural implications, *J. Volcanol. Geotherm. Res.*, **36**(1), 153-176, doi:10.1016/0377-0273(89)90011-5.

Deniel, C., Kieffer, G. & Lecointre, J., 1992. New ^{230}Th-^{238}U and ^{14}C age determinations from Piton des Neiges volcano, Reunion — a revised chronology for the differentiated series, *J. Volcanol. Geotherm. Res.*, **51**(3), 253-267, doi:10.1016/0377-0273(92)90126-X.

Dewey, J. & Byerly, P., 1969. The early history of seismometry, *Bull. Seismol. Soc. Am.*, **59**(1), 183-227.

Duputel, Z., Ferrazzini, V., Brenguier, F., Shapiro, N., Campillo, M. & Nercessian, A., 2009. Real time monitoring of relative velocity changes using ambient seismic noise at the Piton de la Fournaise volcano (La Réunion) from January 2006 to June 2007, *J. Volcanol. Geotherm. Res.*, **184**(1), 164-173, doi:10.1016/j.jvolgeores.2008.11.024.

Dyment, J., Lin, J. & Baker, E.T., 2007. Ridge-hotspot interactions: what mid-ocean ridges tell us about deep Earth processes, *Oceanography*, **20**(1), 102-115, doi:10.5670/oceanog.2007.84.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, **25**(4), 297-356.

Ekström, G., 2001. Time domain analysis of Earth's long-period background seismic radiation, *J. Geophys. Res. Solid Earth*, **106**(B11), 26483-26493, doi:10.1029/2000JB000086.

Flores-Mendez, E., Carbajal-Romero, M., Flores-Guzmán, N., Sánchez-Martínez, R. & Rodríguez-Castellanos, A., 2012. Rayleigh's, Stoneley's, and Scholte's interface waves in elastic models using a boundary element method, *J. Appl. Math.*, doi:10.1155/2012/313207.

Fontaine, F.R., Barruol, G., Tkalčić, H, Wölbern, I., Rümpker, G., Bodin, T. & Haugmard, M., 2015. Crustal and uppermost mantle structure variation beneath La Réunion hotspot track, *Geophys. J. Int.*, **203**(1), 107-126, doi:10.1093/gji/ggv279.

Friedrich, A., Krüger, F. & Klinge, K., 1998. Ocean-generated microseismic noise located with the Gräfenberg array, *J. Seismolog.*, **2**(1), 47-64, doi:10.1023/A:1009788904007.

Gailler, L.-S. & Lénat, J.-F., 2012. Internal architecture of La Réunion (Indian Ocean) inferred from geophysical data, *J. Volcanol. Geotherm. Res.*, **221-222**, 83-98, doi:10.1016/j.jvolgeores.2012.01.015.

Gal, M., Reading, A.M., Ellingsen, S.P., Koper, K.D., Gibbons, S.J & Näsholm, S.P., 2014. Improved implementation of the fk and Capon methods for array analysis of seismic noise, *Geophys. J. Int.*, **198**(2), 1045-1054, doi:10.1093/gji/ggu183.

Gallart, J., Driad, L., Charvis, P., Sapin, M., Hirn, A., Diaz, J., de Voogd, B. & Sachpazi, M., 1999. Perturbation to the lithosphere along the hotspot track of La Réunion from an offshore-onshore seismic transect, *J. Geophys. Res.*, **104**(B2), 2895-2908, doi:10.1029/98JB02840.

Gouédard, P., Seher, T., McGuire, J.J., Collins, J.A. & van der Hilst, R.D., 2014. Correction of ocean-bottom seismometer instrumental clock errors using ambient seismic noise, *Bull. Seismol. Soc. Am.*, **104**(3), 1276-1288, doi:10.1785/0120130157.

Goutorbe, B., Coelho, D.L.d.O. & Drouet, S., 2015. Rayleigh wave group velocities at periods of 6-23 s across Brazil from ambient noise tomography, *Geophys. J. Int.*, **203**(2), 869-882, doi:10.1093/gji/ggv343.

Guo, Z., Gao, X., Shi, H. & Wang, W., 2013. Crustal and uppermost mantle S-wave velocity structure beneath the Japanese islands from seismic ambient noise tomography, *Geophys. J. Int.*, **193**(1), 394-406, doi:10.1093/gji/ggs121.

Gutenberg, B., 1911. *Die seismische Bodenunruhe*, Ph.D. thesis, University of Göttingen, Germany.

Gutenberg, B., 1951. Observation and theory of microseisms, in *Compendium of Meteorology*, pp. 1303-1311, Am. Meteorol. Soc., Boston, MA.

Hable, S., Sigloch, K., Barruol, G., Stähler, S.C. & Hadziioannou, C., 2018. Clock errors in land and ocean bottom seismograms: high-accuracy estimates from multiple-component noise cross-correlations, *Geophys. J. Int.*, **214**(3), 2014-2034, doi:10.1093/gji/ggy236.

Hable, S., Sigloch, K., Stutzmann, E., Kiselev, S. & Barruol, G., 2019a. Crustal tomography of the western Indian Ocean from noise cross-correlations of land and ocean bottom seismometers, *Geophys. J. Int.*, *to be submitted in March 2019.*

Hable, S., Sigloch, K., Stutzmann, E., Kiselev, S. & Barruol, G., 2019b. Noise tomography of the island La Réunion, *in prep.*

Haned, A., Stutzmann, E., Schimmel, M., Kiselev, S., Davaille, A. & Yelles-Chaouche, A., 2016. Global tomography using seismic hum, *Geophys. J. Int.*, **204**(2), 1222-1236, doi:10.1093/gji/ggv516.

Hannemann, K., Krüger, F. & Dahm, T., 2014. Measuring of clock drift rates and static time offsets of ocean bottom stations by means of ambient noise, *Geophys. J. Int.*, **196**(2), 1034-1042, doi:10.1093/gji/ggt434.

Hasselmann, K., 1963. A statistical analysis of the generation of microseisms, *Rev. Geophys.*, **1**(2), 177-210, doi:10.1029/RG001i002p00177.

Hobiger, M., Wegler, U., Shiomi, K. & Nakahara, H., 2012. Coseismic and postseismic elastic wave velocity variations caused by the 2008 Iwate-Miyagi Nairiku earthquake, Japan, *J. Geophys. Res.*, **117**(B9), B09313, doi:10.1029/2012JB009402.

Hosseini, K. & Sigloch, K., 2017. ObspyDMT: a Python toolbox for retrieving and processing of large seismological data sets, *Solid Earth*, **8**, 1047-1070, doi:10.5194/se-8-1047-2017.

Hosseini, K., Matthews, K.J., Sigloch, K., Shephard, G.E., Domeier, M. & Tsekhmistrenko, M., 2018. SubMachine: web-Based tools for exploring seismic tomography and other models of Earth's deep interior, *Geochem. Geophys. Geosyst.*, **19**, 1464–1483, doi:10.1029/2018GC007431.

Hughes, I.G. & Hase, T.P.A., 2010. *Measurements and their uncertainties: a practical guide to modern error analysis*, Oxford Univ. Press.

Institut de Physique du Globe de Paris (IPGP) & Ecole et Observatoire des Sciences de la Terre de Strasbourg (EOST), 1982. GEOSCOPE - French Global Network of broadband seismic stations, Institut de Physique du Globe de Paris (IPGP), doi:10.18715/GEOSCOPE.G.

Juretzek, C. & Hadziioannou, C., 2016. Where do ocean microseisms come from? A study of Love-to-Rayleigh wave ratios, *J. Geophys. Res. Solid Earth*, **121**, 6741-6756, doi:10.1002/2016JB013017.

Juretzek, C. & Hadziioannou, C., 2017. Linking source region and ocean wave parameters with the observed primary microseismic noise, *Geophys. J. Int.*, **211**(3), 1640-1654, doi:10.1093/gji/ggx388.

Köhler, A., Weidle, C. & Maupin, V., 2012. On the effect of topography on surface wave propagation in the ambient noise frequency range, *J. Seismolog.*, **16**(2), 221-231, 10.1007/s10950-011-9264-5.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Discov.*, **8**(1), 014003, doi:10.1088/1749-4699/8/1/014003.

Lacoss, R.T., Kelly, E.J. & Nafi, T.M., 1969. Estimation of seismic noise structure using arrays, *Geophysics*, **34**(1), 21-38, doi:10.1190/1.1439995.

Larose, E., Derode, A., Campillo, M. & Fink, M., 2004. Imaging from one-bit correlations of wideband diffuse wave fields, *J. Appl. Phys.*, **95**(12), 8393-8399, doi:10.1063/1.1739529.

Laske, G., Masters, G., Ma, Z. & Pasyanos, M., 2013. Update on CRUST1.0 - A 1-degree global model of Earth's crust, *Geophys. Res. Abstracts*, **15**, 2658.

Le, B.M., Yang, T., Chen, Y.J. & Yao, H., 2018. Correction of OBS clock errors using Scholte waves retrieved from cross-correlating hydrophone recordings, *Geophys. J. Int.*, **212**(2), 891-899, doi:10.1093/gji/ggx449.

Lénat, J.-F., Gibert-Malengreau, B. & Galdéano, A., 2001. A new model for the evolution of the volcanic island of Réunion (Indian Ocean), *J. Geophys. Res.*, **106**(B5), 8645-8663, doi:10.1029/2000JB900448.

Lénat, J.-F., Bachèlery, P. & Merle, O., 2012. Anatomy of Piton de la Fournaise volcano (La Réunion, Indian Ocean), *Bull. Volcanol.*, **74**(9), 1945-1961, doi:10.1007/s00445-012-0640-y.

Lobkis, O.I. & Weaver, R.L., 2001. On the emergence of the Green's function in the correlations of a diffuse field, *J. Acoust. Soc. Am.*, **110**(6), 3011-3017, doi:10.1121/1.1417528.

Longuet-Higgins, M.S., 1950. A theory of the origin of microseisms, *Philos. Trans. R. Soc. London, Ser. A*, **243**(a), 1-35, doi:10.1098/rsta.1950.0012.

Lu, Y., Stehly, L., Paul, A. & AlpArray Working Group, 2018. High-resolution surface wave tomography of the European crust and uppermost mantle from ambient seismic noise, *Geophys. J. Int.*, **214**(2), 1136-1150, doi:10.1093/gji/ggy188.

Ma, Z. & Dalton, C.A., 2017. Evolution of the lithosphere in the Indian Ocean from combined earthquake and ambient noise tomography, *J. Geophys. Res. Solid Earth*, **122**, 354-371, doi:10.1002/2016JB013516.

Malengreau, B., Lénat, J.-F. & Froger, J.-L., 1999. Structure of Réunion Island (Indian Ocean) inferred from the interpretation of gravity anomalies, *J. Volcanol. Geotherm. Res.*, **88**(3), 131-146, doi:10.1016/S0377-0273(98)00114-0.

Mazzullo, A., Stutzmann, E., Montagner, J.-P., Kiselev, S., Maurya, S., Barruol, G. & Sigloch, K., 2017. Anisotropic tomography around La Réunion Island from Rayleigh waves, *J. Geophys. Res. Solid Earth*, **122**, 9132-9148, doi:10.1002/2017JB014354.

McNamara, D.E. & Buland, R.P., 2004. Ambient noise levels in the continental United States, *Bull. Seismol. Soc. Am.*, **94**(4), 1517-1527, doi:10.1785/012003001.

Megies, T., Beyreuther, M., Barsch, R., Krischer, L. & Wassermann, J., 2011. ObsPy - what can it do for data centers and observatories?, *Ann. Geophys.*, **54**(1), 47-58, doi:10.4401/ag-4838.

Montagner, J.-P. & Jobert, N., 1988. Vectorial tomography—II. Application to the Indian Ocean, *Geophys. J. Int.*, **94**(2), 309-344, doi:10.1111/j.1365-246X.1988.tb05904.x.

Mordret, A., Landès, M., Shapiro, N.M., Singh, S.C. & Roux, P., 2014. Ambient noise surface wave tomography to determine the shallow shear velocity structure at Valhall: depth inversion with a Neighbourhood Algorithm, *Geophys. J. Int.*, **198**(1), 1514-1525, doi:10.1093/gji/ggu217.

Mordret, A., Rivet, D., Landès, M. & Shapiro, N.M., 2015. Three-dimensional shear velocity anisotropic model of Piton de la Fournaise Volcano (La Réunion Island) from ambient seismic noise, *J. Geophys. Res. Solid Earth*, **120**, 406-427, doi:10.1002/2014JB011654.

Morgan, W.J., 1978. Rodriguez, Darwin, Amsterdam, ..., a second type of Hotspot Island, *J. Geophys. Res. Solid Earth*, **83**(B11), 5355-5360, doi:10.1029/JB083iB11p05355.

Nakahara, H. 2012. Formulation of the spatial autocorrelation (SPAC) method in dissipative media, *Geophys. J. Int.*, **190**(3), 1777-1783, doi:10.1111/j.1365-246X.2012.05591.x.

Nakamura, Y. 1989. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface, *Quarterly Report Railway Tech. Res. Inst.*, **30**(1), 25-30.

Nakata, N., Boué, P., Brenguier, F., Roux, P., Ferrazzini, V. & Campillo, M., 2016. Body and surface wave reconstruction from seismic noise correlations between arrays at Piton de la Fournaise volcano, *Geophys. Res. Lett.*, **43**(3), 1047-1054, doi:10.1002/2015GL066997.

Nicolson, H., Curtis, A., Baptie, B. & Galetti, E., 2012. Seismic interferometry and ambient noise tomography in the British Isles, *Proc. Geol. Assoc.*, **123**(1), 74-86, doi:10.1016/j.pgeola.2011.04.002.

Nishida, K., Kawakatsu, H., Fukao, Y. & Obara, K., 2008. Background Love and Rayleigh waves simultaneously generated at the Pacific Ocean floors, *Geophys. Res. Lett.*, **35**, L16307, doi:10.1029/2008GL034753.

Nishida, K., Montagner, J.-P. & Kawakatsu, H., 2009. Global surface wave tomography using seismic hum, *Science*, **326**(5949), 112, doi:10.1126/science.1176389.

Nogoshi, M. & Igarashi, T., 1971. On the amplitude characteristics of microtremor (part 2), *J. Seismol. Soc. Jpn.*, **24**, 26-40, doi:10.4294/zisin1948.24.1_26.

Obrebski, M.J., Ardhuin, F., Stutzmann, E. & Schimmel, M., 2012. How moderate sea states can generate loud seismic noise in the deep ocean, *Geophys. Res. Lett.*, **39**(11), doi:10.1029/2012GL051896.

Peterson, J., 1993. Observations and modeling of seismic background noise, *U.S. Geol. Surv.*, Open-File Report, Albuquerque, New Mexico, doi:10.3133/ofr93322.

Press, W.H., 2007. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge Univ. Press.

Rindraharisaona, E.J., Tilmann, F., Rümpker, G., Giese, J., Rambolamanana, G. & Barruol, G., 2017. Crustal structure of southern Madagascar from receiver functions and ambient noise correlation: implications for crustal evolution, *J. Geophys. Res. Solid Earth*, **122**, 1179-1197, doi:10.1002/2016JB013565.

Rivet, D., Brenguier, F., Clarke, D., Shapiro, N.M. & Peltier, A., 2014. Long-term dynamics of Piton de la Fournaise volcano from 13 years of seismic velocity change measurements and GPS observations, *J. Geophys. Res. Solid Earth*, **119**, doi:10.1002/2014JB011307.

Rossum, G., 1995. Python Reference Manual, *CWI (Centre for Mathematics and Computer Science).*

Ryberg, T., Geissler, W.H., Jokat, W. & Pandey, S., 2017. Uppermost mantle and crustal structure at Tristan da Cunha derived from ambient seismic noise, *Earth Planet. Sci. Lett.*, **471**, 117-124, doi:10.1016/j.epsl.2017.04.049.

Sabra, K.G., Roux, P., Thode, A.M., D'Spain, G.L., Hodgkiss, W.S. & Kuperman, W.A., 2005a. Using ocean ambient noise for array self-localization and self-synchronization, *IEEE J. Oceanic Eng.*, **30**(2), 338-347, doi:10.1109/JOE.2005.850908.

Sabra, K.G., Gerstoft, P., Roux, P. & Kuperman, W.A., 2005b. Surface wave tomography from microseisms in Southern California, *Geophys. Res. Lett.*, **32**(14), L14311, doi:10.1029/2005GL023155.

Salvermoser, J., Hadziioannou, C. & Stähler, S.C., 2015. Structural monitoring of a highway bridge using passive noise recordings from street traffic, *J. Acoust. Soc. Am.*, **138**(6), 3864-3872, doi:10.1121/1.4937765.

Schimmel, M., 1999. Phase cross-correlations: design, comparisons, and applications, *Bull. Seismol. Soc. Am.*, **89**(5), 1366-1378.

Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.*, **130**(2), 497-505, doi:10.1111/j.1365-246X.1997.tb05664.x.

Schimmel, M., Stutzmann, E. & Gallart, J., 2011. Using instantaneous phase coherence for signal extraction from ambient noise data at a local to a global scale, *Geophys. J. Int.*, **184**(1), 494-506, doi:10.1111/j.1365-246X.2010.04861.x.

Schimmel, M., Stutzmann, E. & Ventosa, S., 2017. Measuring group velocity in seismic noise correlation studies based on phase coherence and resampling strategies, *IEEE Trans. Geosci. Remote Sens.*, **55**(4), 1928-1935, doi:10.1109/TGRS.2016.2631445.

Schlömer, A., Geissler, W.H., Jokat, W. & Jegen, M., 2017. Hunting for the Tristan mantle plume – an upper mantle tomography around the volcanic island of Tristan da Cunha, *Earth Planet. Sci. Lett.*, **462**, 122-131, doi:10.1016/j.epsl.2016.12.028.

Scholte, J.G., 1947. The range of existence of Rayleigh and Stoneley waves, *Geophys. J. Int.*, **5**(s5), 120-126, doi:10.1111/j.1365-246X.1947.tb00347.x.

Sens-Schönfelder, C., 2008. Synchronizing seismic networks with ambient noise, *Geophys. J. Int.*, **174**(3), 966-970, doi:10.1111/j.1365-246X.2008.03842.x.

Sens-Schönfelder, C. & Wegler, U., 2006. Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano, Indonesia, *Geophys. Res. Lett.*, **33**(21), L21302, doi:10.1029/2006GL027797.

Sens-Schönfelder, C., Pomponi, E. & Peltier, A., 2014. Dynamics of Piton de la Fournaise volcano observed by passive image interferometry with multiple references, *J. Volcanol. Geotherm. Res.*, **276**, 32-45, doi:10.1016/j.jvolgeores.2014.02.012.

Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**(7), L07614, doi:10.1029/2004GL019491.

Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615-1618, doi:10.1126/science.1108339.

Singh, M., Kaviani, A. & Rümpker, G., 2016. The crustal structure beneath Mauritius from teleseismic P-receiver functions - oceanic or continental?, *J. Geophys. Res. Solid Earth*, **43**, 9636–9643, doi:10.1002/2016GL070529.

Stähler, S.C., Sigloch, K., Hosseini, K., Crawford, W.C., Barruol, G., Schmidt-Aursch, M.C., Tsekhmistrenko, M., Scholz, J.-R., Mazzullo, A. & Deen, M., 2016. Performance report of the RHUM-RUM ocean bottom seismometer network around La Réunion, western Indian Ocean, *Adv. Geosci.*, **41**, 43-63, doi:10.5194/adgeo-41-43-2016.

Stähler, S.C., Schmidt-Aursch, M.C., Hein, G. & Mars, R., 2018. A self-noise model for the German DEPAS OBS Pool, *Seismol. Res. Lett.*, **89**(5), 1838-1845, doi:10.1785/0220180056.

Stehly, L., Campillo, M. & Shapiro, N.M., 2007. Traveltime measurements from noise correlation: stability and detection of instrumental time-shifts, *Geophys. J. Int.*, **171**(1), 223-230, doi:10.1111/j.1365-246X.2007.03492.x.

Stieltjes, L. & Moutou, P., 1989. A statistical and probabilistic study of the historic activity of Piton de la Fournaise, Reunion Island, Indian Ocean, *J. Volcanol. Geotherm. Res.*, **36**(1), 67-86, doi:10.1016/0377-0273(89)90006-1.

Stockwell, R.G., Mansinha, L. & Lowe, R.P., 1996. Localization of the complex spectrum: the S transform, *IEEE Trans. Signal Process.*, **44**(4), 998-1001, doi:10.1109/78.492555.

Storey, M., Mahoney, J.J., Saunders, A.D., Duncan, R.A., Kelley, S.P. & Coffin, M.F., 1995. Timing of hot spot—related volcanism and the breakup of Madagascar and India, *Science*, **267**(5199), 852-855, doi:10.1126/science.267.5199.852.

Tanimoto, T., Hadziioannou, C., Igel, H., Wassermann, J., Schreiber, U. & Gebauer, A., 2015. Estimate of Rayleigh-to-Love wave ratio in the secondary microseism by colocated ring laser and seismograph, *Geophys. Res. Lett.*, **42**, 2650–2655, doi:10.1002/2015GL063637.

Tian, Y. & Ritzwoller, M.H., 2017. Improving ambient noise cross-correlations in the noisy ocean bottom environment of the Juan de Fuca plate, *Geophys. J. Int.*, **210**(3), 1787-1805, doi:10.1093/gji/ggx281.

Tomar, G., Stutzmann, E., Mordret, A., Montagner, J.-P., Singh, S.C. & Shapiro, N.M., 2018. Joint inversion of the first overtone and fundamental mode for deep imaging at the Valhall oil field using ambient noise, *Geophys. J. Int.*, **214**(1), 122-132, doi:10.1093/gji/ggy122.

Villagómez, D.R., Toomey, D.R., Hooft, E.E.E. & Solomon, S.C., 2007. Upper mantle structure beneath the Galápagos Archipelago from surface wave tomography, *J. Geophys. Res. Solid Earth*, **112**, B07303, doi:10.1029/2006JB004672.

Wassermann, J., Wietek, A., Hadziioannou, C. & Igel, H., 2016. Toward a single-station approach for microzonation: using vertical rotation rate to estimate Love-wave-dispersion curves and direction finding, *Bull. Seismol. Soc. Am.*, **106**(3), 1316-1330, doi:10.1785/0120150250.

Wegler, U., Nakahara, H., Sens-Schönfelder, C., Korn, M. & Shiomi, K., 2009. Sudden drop of seismic velocity after the 2004 M_w 6.6 mid-Niigata earthquake, Japan, observed with Passive Image Interferometry, *J. Geophys. Res.*, **114**(B6), B06305, doi:10.1029/2008JB005869.

Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J.F. & Wobbe, F., 2013. Generic Mapping Tools: Improved version released, *Eos, Trans. Am. Geophys. Union*, **94**(45), 409-410, doi:10.1002/2013EO450001.

Wolfram Research Inc., 2018. *Mathematica, Version 11.3*, Champaign, Illinois.

Wysession, M., Wiens, D. & Nyblade, A., 2011. INVESTIGATION OF SOURCES OF INTRAPLATE VOLCANISM USING PASSCAL BROADBAND INSTRUMENTS

IN MADAGASCAR, THE COMORES, AND MOZAMBIQUE, International Federation of Digital Seismograph Networks, doi:10.7914/SN/XV_2011.

Xia, Y., Ni, S., Zeng, X., Xie, J., Wang, B. & Yuan, S., 2015. Synchronizing intercontinental seismic networks using the 26 s persistent localized microseismic source, *Bull. Seismol. Soc. Am.*, **105**(4), 2101-2108, doi:10.1785/0120140252.

Xie, J., Ni, S., Chu, R. & Xia, Y., 2018a. Assessing the short-term clock drift of early broadband stations with burst events of the 26 s persistent and localized microseism, *Geophys. J. Int.*, **212**(1), 324-332, doi:10.1093/gji/ggx401.

Xie, J., Chu, R. & Yang, Y., 2018b. 3-D upper-mantle shear velocity model beneath the contiguous United States based on broadband surface wave from ambient seismic noise, *Pure Appl. Geophys.*, **175**(10), 3403-3418, doi:10.1007/s00024-018-1881-2.

Yao, H., Gouédard, P., Collins, J.A., McGuire, J.J. & van der Hilst, R.D., 2011. Structure of young East Pacific Rise lithosphere from ambient noise correlation analysis of fundamental- and higher-mode Scholte-Rayleigh waves, *C.R. Geosci.*, **343**, 571-583, doi:10.1016/j.crte.2011.04.004.

Zha, Y., Webb, S.C., Wei, S.S., Wiens, D.A., Blackman, D.K., Menke, W., Dunn, R.A. & Conder, J.A., 2014. Seismological imaging of ridge-arc interaction beneath the Eastern Lau Spreading Center from OBS ambient noise tomography, *Earth Planet. Sci. Lett.*, **408**, 194-206, doi:10.1016/j.epsl.2014.10.019.

Zhan, Z., Tsai, V.C. & Clayton, R.W., 2013. Spurious velocity changes caused by temporal variations in ambient noise frequency content, *Geophys. J. Int.*, **194**(3), 1574-1581, doi:10.1093/gji/ggt170.

Ziane, D. & Hadziioannou, C., 2019. The contribution of multiple scattering to Love wave generation in the secondary microseism, *Geophys. J. Int.*, doi:10.1093/gji/ggz056.

Zigone, D., Ben-Zion, Y., Campillo, M. & Roux, P., 2015. Seismic tomography of the Southern California Plate Boundary region from noise-based Rayleigh and Love waves, *Pure Appl. Geophys.*, **172**(5), 1007-1032, doi:10.1007/s00024-014-0872-1.

Acknowledgements

This thesis has greatly benefited from the support of several people, to whom I would like to express my gratitude here.

First of all, I owe much to Karin Sigloch and her excellent supervision. She has always taken care of my work despite the large spatial distance between us. Her detailed comments and valuable suggestions highly improved this thesis. She enabled me to be part of the great RHUM-RUM community and to get in touch with excellent researchers.

I thank Heiner Igel for the opportunity to do my PhD in a great scientific environment at the institute of Munich and for his dynamic way of leading the seismology group.

I thank Eleonore Stutzmann, who gave me the chance to work on group velocities, which became the second part of my thesis. I really appreciate her scientific assistance during several meetings in Paris. Sergey Kiselev is thanked for his great help in tomographic inversion and their graphical visualization.

I am very thankful to Guilhem Barruol for sharing his comprehensive RHUM-RUM data knowledge and for organizing an amazing workshop on the beautiful island of La Réunion.

Great thanks go to Céline Hadziioannou and her enthusiasm for noise, which had an great impact on this thesis. Simon Stähler is thanked for useful discussions whenever unexpected features appeared in the data that I needed to explain. Further thanks go to Wayne Crawford for his expertise in any OBS data issues and for preparing timing-corrected data for RESIF.

Finally, I want to thank my parents for their constant support and encouragement throughout my whole life.

At the very end, special thanks go to Johannes, who has always been by my side during the last years and who will set off to a new adventure together with me.