UNIVERSITÀ degli STUDI di CATANIA

Dipartimento di Scienze
Biologiche, Geologiche e Ambientali

PhD Course in "Earth and Environmental Science"

Dottorato di Ricerca in "Scienze della Terra e dell'Ambiente"

XXXIII Cycle

UNRAVELLING THE RELATIONSHIP BETWEEN MICROSEISM AND SEA STATE BY STATISTICAL AND MACHINE LEARNING APPROACHES: ANTARCTICA AND EASTERN SICILY CASES OF STUDY

Salvatore Moschella

Advisor:
Prof. Stefano Gresta

Co-advisor:
Prof. Andrea Cannata
Dott. Giuseppe Di Grazia
Ing. Flavio Cannavò

Coordinator:
Prof. Agata Di Stefano

Ph. D. attended during 2017/2020

SUMMARY

ABSTRACT

The most continuous and ubiquitous seismic signal on Earth is the microseism, closely related to ocean wave energy coupling with the solid Earth. On the basis of the spectral content, it is possible to distinguish three types of microseism: primary (PM), secondary (SM), and short-period secondary microseism (SPSM). In this thesis, we characterise microseism in terms of amplitude and spectral content, and explore the relationship between microseism and sea state by using seismic signals recorded in two different areas: Antarctica and Eastern Sicily. Regarding the seismic signals, data recorded by 20 stations during 1993-2017 and by 6 stations during 2010-2017 were used in the former and latter areas, respectively. Concerning the sea state along the Antarctica coasts, we used sea ice concentration maps coming from satellite data in the whole Southern Ocean. On the other hand, regarding the Eastern Sicily, sea wave height data, provided by both buoys in the Ionian and Tyrrhenian Seas and hindcast maps of the Mediterranean Sea, were used. As for the microseism characterization, most of its energy turned out to be contained in the SPSM and SM bands, while the band 13–20 s (PM) showed a much weaker energy. Microseism exhibited seasonal variability in both Eastern Sicily and Antarctica. In the former area, the amplitude maxima were observed during winter and the minima during summer, as expected in all the temperate latitude regions. Conversely, in Antarctica the sea ice impedes the microseism generation and thus makes the temporal pattern of microseism amplitudes different with respect to the microseism recorded in temperate latitude regions. The results obtained by the correlation analysis showed clear anti-correlation patterns between sea ice and SPSM recorded in Antarctica and

high correlation values between SPSM and sea wave height in the Ionian and Tyrrhenian Seas. Finally, we presented algorithms, based on machine learning techniques, allowing to spatially and temporally reconstruct the sea state in both Antarctica and Eastern Sicily with fairly low error. This technique will allow inferring the sea state in periods when the satellite images or buoy data, routinely used for sea state monitoring, are not available, with wide applications in many fields, first of all climate studies. Combining the results coming from the correlation maps and the machine learning techniques, we showed a decrease in sea state sensitivity of microseism, due to the increasing distance from the station recording the seismic signal. This is particularly important for the future development of an experimental monitoring system of the sea state conditions based on microseism recordings.

Keywords: microseism, sea ice, sea wave height, correlation analysis, machine learning

Studio delle relazioni tra microseism e stato del mare mediante analisi statistiche e di machine learning: casi di studio in Antartide e Sicilia orientale

SINTESI IN ITALIANO

Il segnale sismico più continuo e onnipresente registrato sulla Terra è il microseism, generato dal trasferimento dell'energia delle onde oceaniche alla Terra solida. Sulla base del contenuto spettrale, è possibile distinguere tre tipi di microseism: primario (PM), secondario (SM) e secondario a corto periodo (SPSM). In questa tesi, il microseism è stato inizialmente caratterizzato in termini di ampiezza e contenuto spettrale, successivamente è stata studiata la relazione tra microseism e stato del mare utilizzando segnali sismici registrati in due diverse aree: l'Antartide e la Sicilia orientale. Per quanto riguarda i segnali sismici, sono stati utilizzati i dati registrati da 20 stazioni nel periodo 1993-2017 e da 6 stazioni nel periodo 2010-2017, rispettivamente nella prima e nella seconda area. Per quanto riguarda lo stato del mare lungo le coste dell'Antartide, abbiamo utilizzato mappe di concentrazione di ghiaccio marino provenienti da dati satellitari in tutto l'Oceano Antartico. Per quanto riguarda invece la Sicilia orientale, sono stati utilizzati i dati sull'altezza delle onde marine, forniti sia da boe installate nel Mar Ionio e nel Tirreno, sia dalle mappe di hindcast del Mar Mediterraneo. Per quanto riguarda la caratterizzazione del microseism, la maggior parte della sua energia è contenuta nelle bande SPSM e SM, mentre la banda PM ha mostrato un'energia molto più debole. Il microseism presenta una variabilità stagionale sia nella Sicilia orientale che in Antartide. Nella prima area, i massimi di ampiezza sono stati osservati durante l'inverno e i minimi durante l'estate, come

previsto in tutte le regioni di latitudine temperata. Al contrario, in Antartide il ghiaccio marino impedisce la generazione del microseism, dando luogo quindi ad una diversa variabilità temporale delle ampiezze del microseism rispetto a quanto osservato nelle regioni di latitudine temperata. I risultati ottenuti dall'analisi di correlazione hanno mostrato chiare anti-correlazioni tra ghiaccio marino e SPSM registrati in Antartide, e alti valori di correlazione tra SPSM e altezza delle onde del mare nel Mar Ionio e nel Tirreno. Infine, sono stati applicati algoritmi basati su tecniche di machine learning, che consentono di ricostruire spazialmente e temporalmente lo stato del mare sia in Antartide che in Sicilia orientale con un errore piuttosto basso. Questa tecnica permetterà di desumere lo stato del mare nei periodi in cui le immagini satellitari o i dati delle boe, abitualmente utilizzati per il monitoraggio dello stato del mare, non sono disponibili, con ampie applicazioni in molti campi, primo fra tutti lo studio sul clima. Combinando i risultati provenienti dalle mappe di correlazione e dalle tecniche di machine learning, è stata evidenziata una diminuzione dell’effetto dello stato del mare sul microseism, dovuta alla crescente distanza dalla stazione che registra il segnale sismico dal mare. Ciò è particolarmente importante per il futuro sviluppo di un sistema di monitoraggio sperimentale delle condizioni dello stato del mare basato su registrazioni di microseism.

Parole chiave: microseism, ghiaccio marino, altezza dell’onda marina, analisi di correlazione, machine learning

PREFACE

Modern seismology is able to obtain plenty of information by the analyses of signals, that until a couple of decades ago were considered to be noise, such as microseism. This is considered as the most continuous and ubiquitous seismic signal on Earth, generated by ocean wave energy coupling with the Earth's ground (Longuet-Higgins, 1950; Hasselmann, 1963; Ardhuin et al., 2015). On the basis of its source mechanism and spectral content, it is classified as: primary microseism (hereafter referred to as PM), secondary microseism (SM), and short-period secondary microseism (SPSM) (Haubrich & McCamy, 1969; Bromirski et al., 2005). Because of the microseism source nature, such a signal has been used to make inferences on climate changes (Grevemeyer et al., 2000; Aster et al., 2008; Stutzmann et al., 2009). Indeed, the microseism features closely depend on the sea state, and their link has been empirically and theoretically explored by several authors (e.g. Bromirski et al., 1999; Bromirski & Duennebier, 2002; Ardhuin et al., 2012; Ferretti et al., 2013, 2018).

Furthermore, microseism investigations and, more generally, seismological studies are currently undergoing a rapid increase in dataset volumes (e.g. Kong et al., 2018; Jiao & Alavi, 2019). For this reason, nowadays, applications of machine learning techniques (hereafter referred to as MLTs) on seismological data are increasing in number day by day.

In the following chapters of this thesis the results, obtained highlighting the relationship between microseism and sea state, will be presented. In particular, the Chapter 1 will deal with the fundamentals of microseism, an overview of microseism source location methods and the applications, with a final focus on machine learning applications in seismology. Successively, the thesis will consider the seismic signals

recorded in two different areas, Antarctica and Eastern Sicily. Chapter 2 describes the first study, where the link between microseism recorded around Antarctica coastline and sea ice will be investigated. Chapters 3 and 4, instead, will consider the second study, focusing on the link between microseism recorded on Eastern Sicily and the sea wave height data coming from both buoys and hindcast maps in the Mediterranean Sea. Specifically, each of the three chapters will be divided into sections concerning general introduction, methods of analysis, description of the results, discussion and conclusions. Finally, Chapter 5 will give an overview of the main conclusions obtained in the thesis.

The initial steps of both the above mentioned studies will concern the investigation of seismic signals recorded, focusing the attention on the microseism characterization in terms of frequency content and amplitude. The following work, on the other hand, will be articulated in two different types of analyses: i) correlation analyses between microseism and sea state time series, to highlight any relationship between them and improve the knowledge on microseism sources; ii) machine learning applications, with the final aim to reconstruct the spatial-temporal distribution of the sea state features starting from microseism, providing essential information on the seismic station locations and the microseism period bands to be used to achieve the best results.

The idea of the Antarctica study arises from the investigations performed by Anthony et al. (2017), who quantitatively studied the relationship between microseism and sea ice in the Antarctic Peninsula. However, to date no technique has been found to infer the sea ice distribution in the Southern Ocean, based on microseism recordings. By using 20 seismic stations in a time interval from 1993 to 2017, the quantitative

relationship between microseism, recorded along the Antarctic coasts, and sea ice concentration, will be discussed and the capabilities of machine learning-derived models to infer the sea ice distribution in the Southern Ocean, based on microseism recordings, will be explored.

The Eastern Sicily study will consider the quantitative relationship between microseism recorded on coastline of Eastern Sicily and significant sea wave height recorded in the Mediterranean Sea from 2010 to 2017. A previous study has already shown the spectral features of the microseism recorded in this area (De Caro et al., 2014), and Ferretti et al. (2013, 2018) found empirical relations to predict the significant wave height along the Ligurian coast (Italy). On the basis of the availability of seismic and buoy data in the Ionian and Tyrrhenian Seas and coastlines, together with the hindcast maps of sea wave height, the link between sea waves and microseism, as well as locations of microseism sources, will be highlighted for the first time in this area. Also in this case, regression models, obtained by up-to-date machine learning algorithms, will be constrained to infer sea wave height information from the microseism.

1. BACKGROUND AND STATE OF THE ART

1.1 SEISMIC AMBIENT NOISE AND MICROSEISM

The first seismic ambient noise study has been carried out in the late 19th century by Bertelli (1872, 1875, 1878) who recognized these small oscillations of ground motion recorded by a seismometer and related this signal to meteorological phenomena, as well as identified their seasonality.

Milne in 1883 observed these oscillations worldwide, correlating them to barometrical and tidal variations, even him indicating a natural origin.

These *Earth Pulsations*, as Milne called them, however, were no longer studied in detail until the second half of the twentieth century, primarily because the most important seismological studies were focused on earthquakes. As consequence, the continuous background signal recorded in seismic trace was discarded and considered as *noise.*

The continuous technology innovation in seismometers (for instance long period and broadband seismic stations), data storage and computational capacity therefore led to the first studies on ambient noise wave-field, its source mechanisms and locations (Longuet - Higgins, 1950; Hasselmann, 1963).

It has been widely observed (e.g. Gutenberg, 1958; Asten, 1978; Asten and Henstridge, 1984) that ambient noise sources have frequencies lower than 1 Hz; above this limit, sources are usually related to anthropic (cultural) sources including traffic and industrial operations that are related to economic conditions (Hong et al., 2020; Lecocq et al., 2020b). Also, the instrumental noise (the noise created by the equalizer, filter and amplifier circuitry), have frequencies lower than 1 Hz but prevails at frequencies outside of the seismometer's performance bandwidth, where the signal

is strongly attenuated (e.g. Peterson, 1993; Barzilai et al., 1998; Havskov & Alguacil, 2016).

Amplitudes of this ambient noise below 1 Hz were so overwhelming that, before the development of high-dynamic-range digital seismic instruments, seismologists recorded seismic waves separately for high-frequency range (above about 0.5 Hz) and for low-frequency range (below 0.1 Hz) in order to avoid this microseismic noise (Tanimoto et al., 2015). World-Wide Standard Seismograph Network had such separate (short-period and long-period) instruments. Modern instruments with high-dynamic range and digital recordings acquire seismic signals from about 1 mHz to 10 Hz using the same sensor causing that microseisms are now recorded continuously.

Peterson (1993) collected seismic data from a network of 75 worldwide seismograph stations in order to refine models of seismic background noise. The most important result of this study is the parameterized model called the new low and new high noise model (NLNM - NHNM). This model is referring only on vertical component of seismometers and it is represented by power spectral density amplitude (PSD) over frequency (Fig. 1.1). The Figure 1.1 shows different peaks and amplitude variations: from the low-frequency end, the decreasing amplitude trend with frequency up to about 2 mHz is clearly recognized (correlated with atmosphere effects), followed by a small peak at about 7–10 mHz (denoted as hum, correlated with infragravity waves), and two amplitude peaks at about 0.1–0.4 Hz (referred to as microseisms).

Sea waves are the primary source of noise observed on broadband seismometers at all Earth sites (Webb, 2002), while high frequency noise (> 1 Hz) is often correlated

to anthropogenic (cultural) sources (Hong et al., 2020; Lecocq et al., 2020b) or seismic coupling due to the wind.

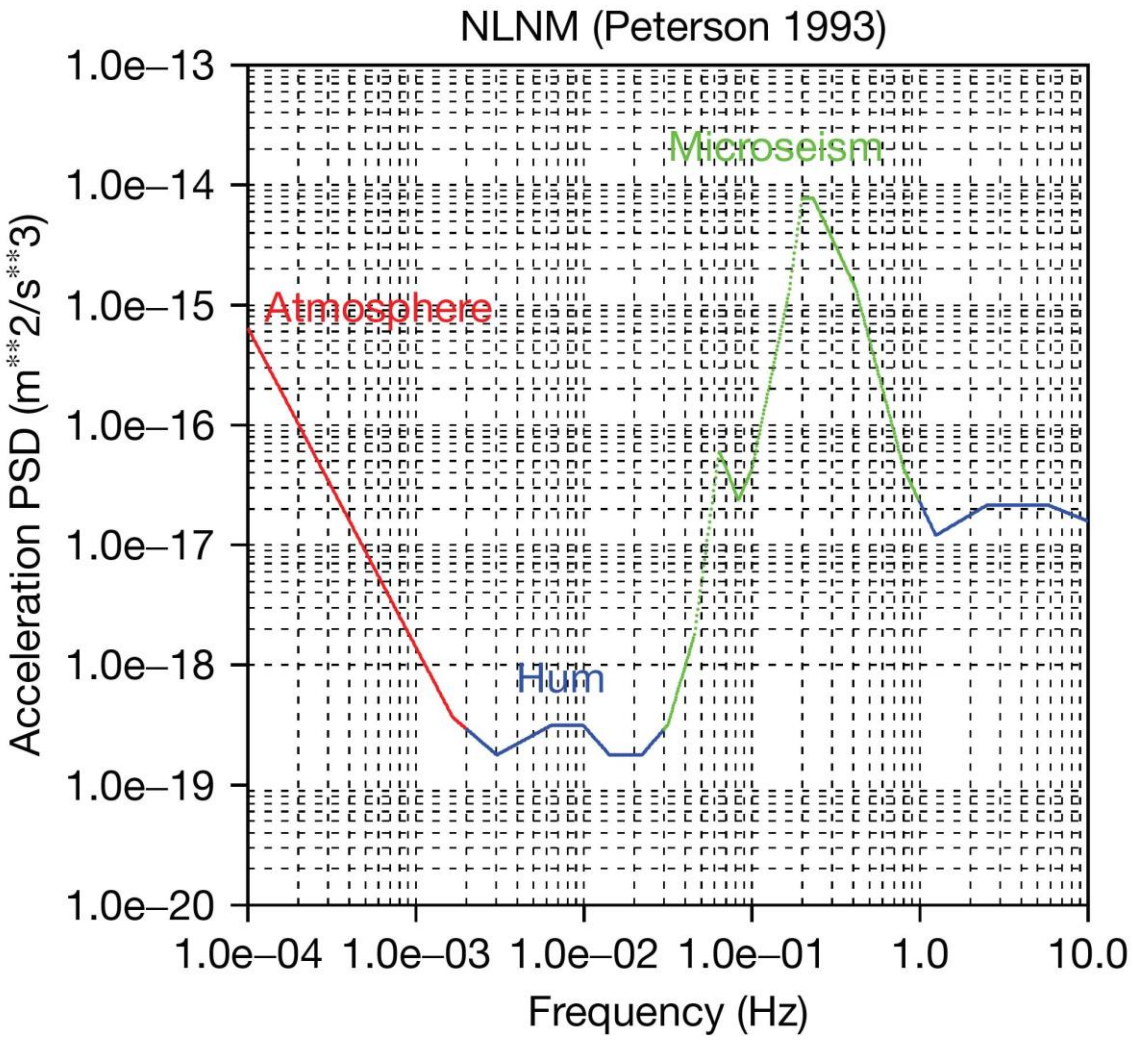

Figure 1.1: The new low-noise model (NLNM) by Peterson (1993). Power spectral density (PSD) in unit for acceleration is used for this plot. Frequency range is from 10^{-4} to 10 Hz (from Tanimoto et al., 2015).

The long-period band is controlled either by seawater currents or by the effect of the low-frequency sea waves, infragravity waves (Webb, 1998). The infragravity waves are long-wavelength surface gravity waves generated by a nonlinear superposition of wind-driven waves and swell (Webb et al., 1991). The hum is associated with ocean waves for periods shorter than 300 s (Tanimoto et al., 2005).

Microseisms arise from atmospheric energy converted to (primarily) Rayleigh waves via the intermediary of wind-driven oceanic swell (e.g. Haubrichet al., 1963). Their predominant frequency range varies from region to region, but it can be bracketed within the frequency band 0.05 - 0.4 Hz almost anywhere in the world (Tanimoto et al., 2007).

The strong relationship between microseism and sea waves led to correlate this kind of noise with the storm activity (e.g. Gutenberg, 1947; Ardhuin et al., 2012) and with the coastal sea wave interaction using buoy data and in-land seismometers (e.g. Bromirski et al., 1999, 2002; Ferretti et al., 2013, 2016, 2018; Cannata et al., 2020).

Microseism amplitudes at temperate latitudes in both northern and southern hemispheres show strong annual periodicity with maxima during the winter seasons, when the oceans are stormier, and minima during summers (e.g. Aster et al.,2008; Cannata et al., 2020). However, such a pattern is different in Antarctica where during the winter, because of the sea ice, the oceanic waves cannot efficiently excite seismic energy (e.g. Grob et al., 2011; Anthony et al., 2014, 2016; Cannata et al., 2019).

The wide spreading of seismic stations and the availability of historical seismic data, that can integrate the sparser buoy data, led to several studies on long-time microseisms records (e.g. Grevemeyer et al., 2000; Aster et al., 2010), attempting an interesting reconstruction of the wave climate comparing microseisms data with wave records and numerical wave model, as well as continuous monitoring climate change (e.g. Stutzmann et al., 2009; Lecocq et al., 2020a).

It is well known that microseism spectra have two peaks identified as primary (or single frequency) and secondary (or double frequency) microseisms (Haubrich et al., 1963; Haubrich & McCamy, 1969). These microseism peaks are evident in spectra

from any site, even far from the coast; this close similarity is a consequence of the efficient propagation of Rayleigh waves at periods larger than few seconds (Webb, 1998). Primary microseism (PM) has a spectral content equal to the ocean wave frequency (period between 13 and 20 s) and its source is associated with the energy transfer of ocean waves breaking/shoaling against the shoreline (Hasselmann, 1963). Secondary microseism (SM), with most energy between 5 and 10 s (roughly twice the frequency of ocean waves), is generally characterised by higher amplitude than primary microseism. According to the most accredited theories, secondary microseism is generated by interactions between waves of the same frequency travelling in opposite directions (Longuet and Higgins, 1950; Ardhuin et al., 2012). Finally, there is a short period secondary microseism (SPSM), characterized by period shorter than 5 s and sources generally linked to local sea state and wave activity, and influenced by local winds (Bromirski et al., 2005; Chen et al., 2011).

In Figure 1.2 the characteristics of microseism recorded in one of the seismic stations located on Eastern Sicily coast (EFIU station, part of INGV, Istituto Nazionale di Geofisica e Vulcanologia, seismic monitoring network) are shown.

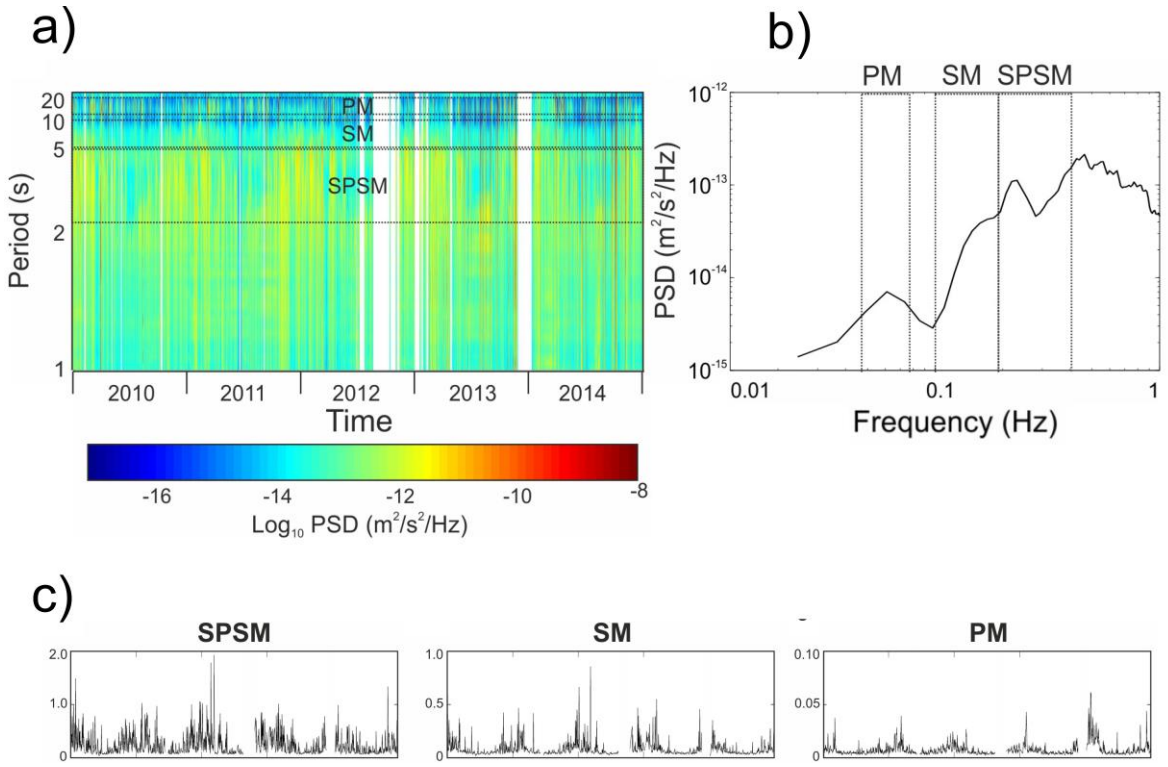

Figure 1.2: spectrogram of the seismic signal recorded by the vertical component (a), the corresponding median spectrum (b) and the RMS amplitude time series filtered in the three different microseism period bands (PM, SM and SPSM) (c) for EFIU seismic station, located in Eastern Sicily coast and part of the INGV seismic network.

1.1.1 PRIMARY MICROSEISM (PM)

Primary microseism (PM) is known to be generated in shallow water by nonlinear coupling of ocean gravity wave energy into seismic wave energy at the same frequency (Hasselmann, 1963; Haubrich & McCamy, 1969). Ardhuin et al. (2015) made a detailed quantitative estimate of seismic noise levels due to pressure variations generated by ocean gravity waves propagating in shallow water in the presence of seafloor slopes, as proposed by Hasselmann (1963). The resulting seismic waves have the same signal period as the ocean waves that cause them, typically between 10 and 20 s (see the weaker of the two peaks in the NLNM curve in Fig.

1.1). According to Airy first-order wave theory approximation, the first-order pressure variations associated with the orbital motion of water particles decrease exponentially with depth (Fig. 1.3).

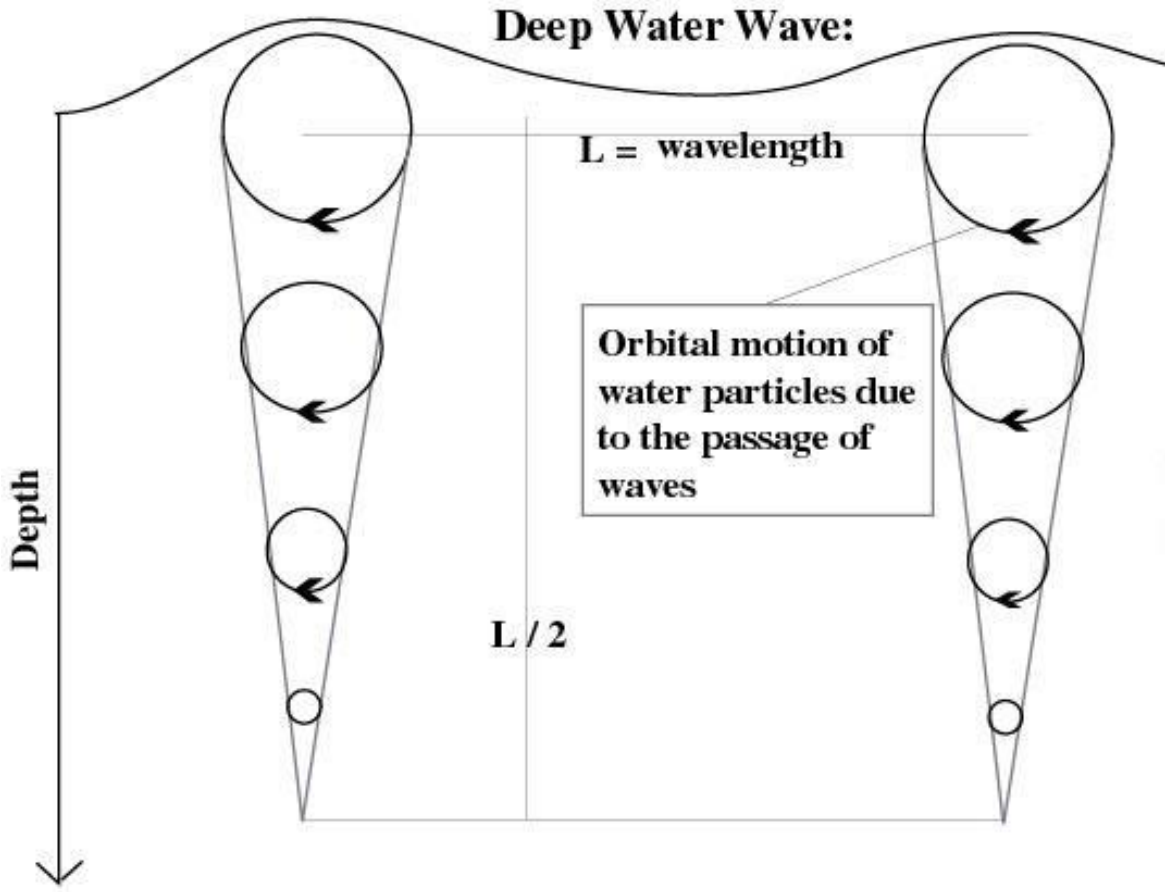

Figure 1.3: Airy's linear theory of orbital motion of water particles that decreases with depth (modified from Burley and Suddeth, 2013).

Gravity waves begin to interact appreciably with the seafloor when the water depth, *h*, is less than half the deep-water wavelength, *L*, determined using the Airy linear wave theory approximation as:

$$L = \frac{gT^2}{2\pi} \qquad (1.1)$$

where *g* is the gravitational acceleration and *T* is the wave period. Intermediate depths fall in the range $L/20 < h < L/4$, where $h < L/20$ is considered the shallow water zone

and $h > L/4$ is the deep-water zone. This suggests that most primary microseism generation probably occurs at water depths less than the $L/4$ bound (Bromirski et al., 2002).

In Figure 1.4, an illustration of primary microseism generation mechanism is shown.

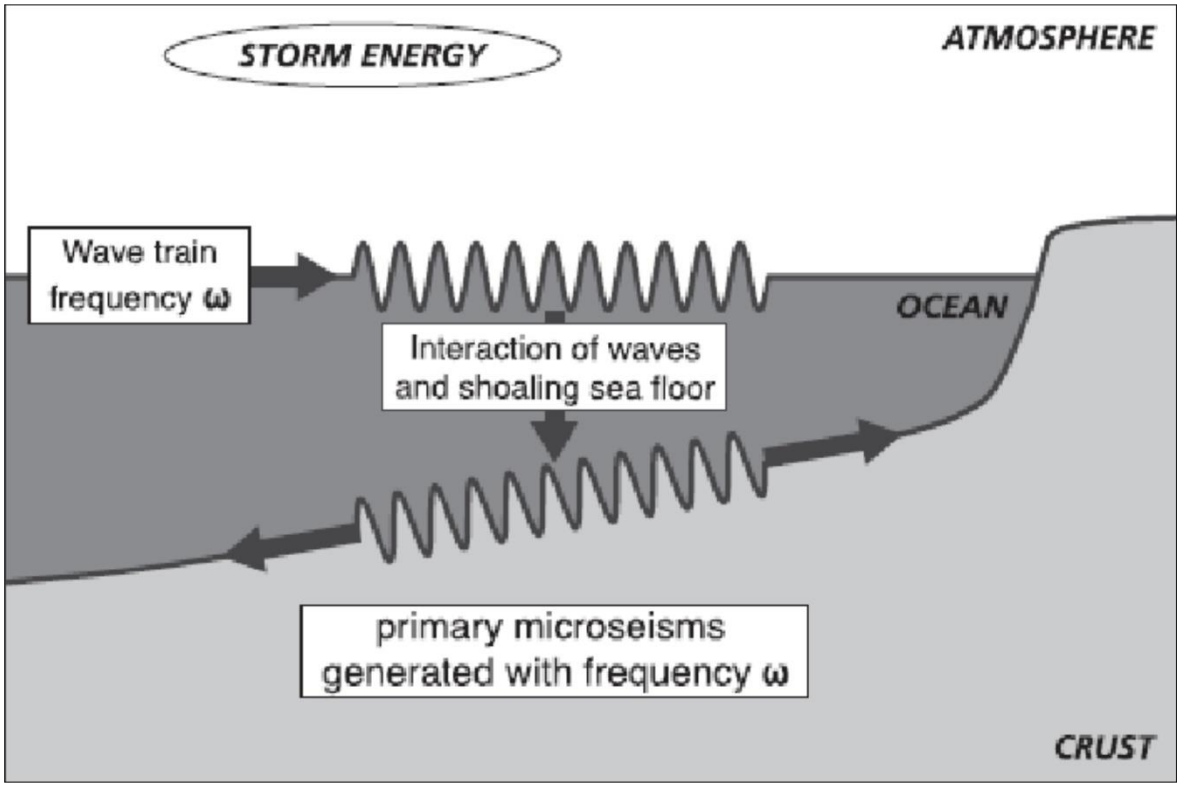

Figure 1.4: Primary microseisms are generated through nonlinear coupling of ocean wave energy into the seafloor by the shoaling action of waves in shallow water (Hasselmann, 1963). Primary microseism frequency ω is that of the ocean waves responsible for their generation, which may travel great distances before primary microseisms are generated (from Ebeling, 2012).

A theoretical framework of PM has mainly evolved for Rayleigh waves only. However, in previous studies, the contribution of Love waves to the PM wavefield was found to be comparable to or even larger than those of Rayleigh waves (e.g. Friedrich et al., 1998; Lin et al., 2008; Nishida et al. 2008; Juretzek et al., 2016) leading to a source mechanism more likely to be interactions of ocean waves with the

solid Earth in shallow oceans, thus producing horizontal forcing through topographic coupling (Saito, 2010). Ardhuin et al. (2015) highlighted that small-scale topographic features on the shallow part of continental shelves are the likely dominant sources of the PM peak.

PMs source areas are often spread out along the coast (Friedrich et al., 1998; Bromirski et al., 2001). Möllhoff and Bean (2016), for instance, detected source regions of PMs along west coast of Ireland in water shallower than 150 m, considering a PM average period peak of 14 s.

PM locations have shown a seasonal variability (e.g. Stehly et al., 2006; Stutzmann et al., 2009; Gulatieri et al., 2019): for the Northern Hemisphere, the sources are mainly located in the northern Oceans during winter, and southern Oceans during summer. This seasonality is explained by the seasonal increase of nearby oceanic storms (e.g. Aster et al., 2008). Gualtieri et al. (2019), for instance, showed that at low frequency (i.e. 0.05 Hz), the dominant source regions can be located thousands of kilometres away from the stations.

The link between microseism and ocean wave parameters (such as the significant wave height) has been quantitatively explored (e.g. Bromirski et al., 1999, 2002; Ferretti et al., 2013, 2016, 2018). However, such a pattern is different in Antarctica where during the winter, because of the sea ice, the oceanic waves cannot efficiently excite seismic energy (Aster et al., 2008; Anthony et al., 2015, 2017; Cannata et al., 2019; Turner et al., 2020).

1.1.2 SECONDARY MICROSEISM (SM)

The secondary (or double frequency) microseism (SM) is commonly generated by oceanic waves reflecting off the coastline and interacting with incoming waves to form standing pressure fluctuations of the oceanic wavefield (e.g. coastal reflections, storm–storm, or intrastorm wave interactions; Ardhuin et al., 2011) that oscillate and couple with the seafloor at half the period of the oceanic waves (e.g. Longuet-Higgins, 1950; Tanimoto, 2007).

SM with most energy between 5 and 10 s (Oliver & Page, 1963; Haubrich et al., 1963; Haubrich & McCamy, 1969) is generally characterised by higher amplitude than primary microseism (see the strongest peak in the NLNM curve in Fig. 1.1).

The wave–wave interaction is basically a vertical forcing, which excites Rayleigh waves much better than Love waves. This mechanism for the SM showed that interactions of two ocean waves in opposite direction can create the double-frequency microseisms (Fig.1.5) through the nonlinear (advection) term in the Navier–Stokes equation (Longuet-Higgins, 1950).

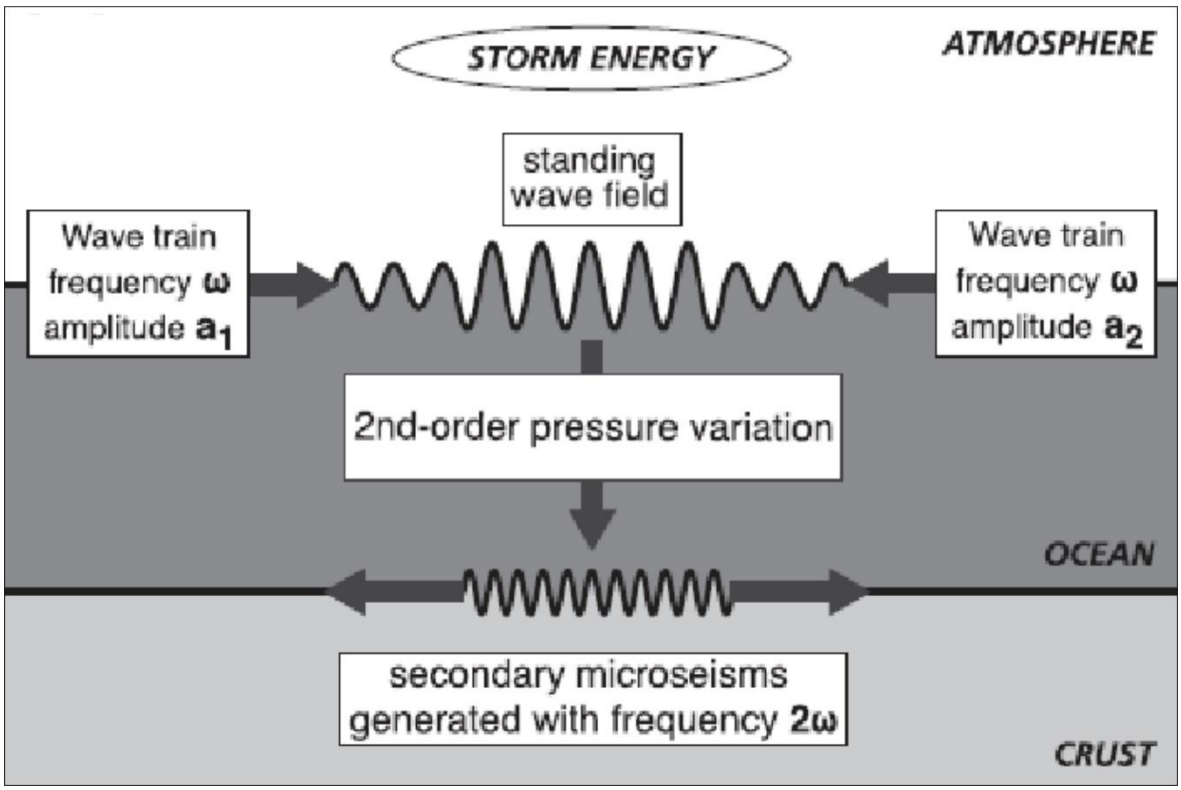

Figure 1.5: Secondary microseisms are generated only when ocean wave trains of the same frequency traveling in opposite directions meet. Under these conditions, a depth-independent second-order pressure variation arises, with amplitude proportional to the product of wave amplitudes (a_1 and a_2) and its frequency (2ω) double that of the ocean waves (from Ebeling, 2012).

Bromirski (2001) clearly showed that the dominant source area for the SM was not in the open ocean, where the highest waves occurred, but was near the coast. It seems therefore that standing ocean waves that occur near the coasts, even at the time of large low-pressure system, are the source of SM. Other studies suggested the existence of pelagic sources, based on a comparison between seismic data and hindcast ocean wave spectral data from the North Atlantic Ocean (e.g. Kedar et al., 2008). Möllhoff and Bean (2016), for instance, detected source regions of SM lying west and northwest of Ireland, suggesting that SM sources are frequently located in regions, that did not coincide with the area of strongest storm activity. Chevrot et al.

(2007), using array analysis, detected SM source areas in coastal regions where the swell reaches steep rocky coasts with normal incidence, in good agreement with the Longuet-Higgins model for the generation of secondary microseisms. In addition, they found evidence of occasional pelagic sources, which are closely related to moving storms, suggesting that nonlinear interaction between wave components can also generate secondary microseisms.

Stutzmann et al. (2009) showed that the secondary microseism amplitude variations are global and depend predominantly on the station latitude and season. Maximum amplitudes are observed during local autumn and winter for stations at high latitudes. Moreover, SM sources generated by ocean waves from the Southern Hemisphere highest wave areas can be recorded by Northern Hemisphere stations when local sources are weak, that is during local summer (Stutzmann et al., 2009). Finally, variations in SM power at specific stations in Antarctica are known to be strongly sensitive to both near-coastal storms and to wave state (e.g. MacAyeal et al., 2006) and are amplitude-modulated by the annual formation and breakup of sea ice (Aster et al., 2008, 2010; Grob et al., 2011, Anthony et al., 2014, 2016; Cannata et al., 2019). Finally, Bromirski et al. (2005) pointed out that the SM can often split in two different peaks: the first one, the long-period secondary microseim (LPSM), is found at ~ 0.085 – 0. 20 Hz; the second peak, short-period secondary microseism (SPSM), is found at 0.20 – 0.50 Hz, and it is associated with waves induced by local winds.

1.1.3 SHORT PERIOD SECONDARY MICROSEISM (SPSM)

Bromirski et al. (2005) highlighted that SPSM recorded at open-ocean seismic stations comes from local wave-wave interaction, strongly correlated with abrupt changes in wind direction. SPSM, recorded in the horizontal components of nearshore seismic stations, is produced nearby the coastlines by local storm activity and therefore by local wave activity, propagating inland as Rayleigh waves. Because of its high frequency range, SPSM does not propagate along long distance (Bromirski et al., 2005).

De Caro et al. (2014) observed a strong relationship between the increase in wind speed recorded by a Catania meteorological station and the increase in energy levels of the SPSM during the wintertime, recorded by an Ocean Bottom Seismometer (OBS) installed in Ionian deep sea.

Regarding source locations of SPSM, Chen et al. (2011), for instance, analysed microseism data collected in Taiwan and showed how stronger excitation in SPSM takes place in the narrow Taiwan Strait, where water depth is very shallow, while the excitations are relatively weak in the eastern off-shore area, an open sea with water depth increasing rapidly off the coast. Möllhoff and Bean (2016) detected source regions of SPSM coincident with the Ireland coastline where surf is generated and sea waves break. Other investigations found that SPSM could be effectively generated by resonance of the compression waves at favourable depth in deep oceans as well (Kedar et al., 2008).

1.2 OVERVIEW ON MICROSEISM SOURCE LOCATION METHODS

While the nonlinear interactions of ocean waves are necessary to generate microseism, there is some confusion as to exactly locate the excitation sources (Tanimoto et al., 2015). There has been a long debate about near-coastal sources versus pelagic sources (in the deep oceans). Near-coastal sources seem common but there seems to be a strong case for some pelagic sources (e.g. Cessaro, 1994; Kedar et al., 2008). What we know for sure is the fact that a source must be in a place where ocean-wave collisions occur.

Studies early in XX century proposed their association with meteorological storm systems in the ocean. Cessaro (1994) pointed out that, although pelagic storms can provide the source of microseismic wave energy, it is the interplay among (1) the pelagic storm parameters, such as tracking velocity, peak wind speed, location, effective area, and ocean surface pressure variation, (2) the resulting storm waves and their frequency distribution, (3) the direction of storm wave propagation, and (4) the near-shore and deep-ocean processes, that control the production of microseism.

Analysis of microseism recorded by seismic arrays has provided a better understanding of the nature of the low-frequency noise (seismic and acoustic) generated by large pelagic storms that propagate into the continental interior as microseism surface waves (Cessaro, 1994). Capon (1972) suggested the coincidence of an atmospheric low-pressure region with a microseism source determined from simultaneous frequency-wavenumber (f-k) analysis of two different array datasets. Cessaro and Chan (1989) and Cessaro (1994) found two nearshore source locations that were associated with nearly all the coherent primary microseism propagating across two arrays. Juretzek and Hadziioannou (2016, 2017), for instance, applied the

beamforming approach to three-component seismic data of different arrays distributed in Europe to investigate the directionality of PM ad SM sources and seasonal variations. From the perspective of a seismic array, at any given moment only the most energetic coherent portion of the noise field is detected by f-k analysis, i.e., a peak in the frequency-wavenumber power represents the most energetic coherent portion of the microseismic wave field at that instant (Cessaro, 1994).

Other source location methods use the polarizations analysis of the three components of the ground motion (e.g. Jurkevics, 1988). Stutzmann et al. (2009), for instance, determined the back azimuth of the SM as a function of time and frequency as well as the degree of polarization, showing that the source azimuths are predominantly toward the closest ocean, but with azimuth variations between winter and summer.

Another kind of approach, used to investigate the source locations of microseism, is the correlation with ocean wave data coming from buoys or hindcast maps. Bromirski (2001) computed the cross-correlation coefficients between wave spectra from widely separated buoy measurements in North America East Coast and corresponding SM spectra at near-coastal seismic station. The results showed that the dominant source area of SM during the intense storm activity is near the southern Massachusetts coast, not in the open ocean where the highest waves occurred. Essen et al. (2003) applied two methods to determine the generation areas. First, computed cross-correlation coefficients between square root of the microseism amplitudes and modelled ocean wave heights in the North Atlantic Ocean. Areas with maximum correlation coefficients, assumed as generation areas of microseism in the ocean, were determined. Second, an attempt is made to locate generation areas from the azimuthal direction of the microseism at different locations using array analysis. Both

techniques reveal that the main generation areas are in specific regions off the coast of Southwest Norway and North Scotland.

Other authors (e.g. Dham et at., 2006; Chevrot et al., 2007; Gal et al., 2015) used array analyses and correlation between microseism amplitude and sea wave height, obtaining distinct source location areas for different microseism bands.

Other approaches are based on characterizing or simulating the wavefield of microseisms (e.g. Tanimoto et al., 2006; Tanimoto, 2007; Brooks et al., 2009; Gualtieri et al., 2013, 2014, 2015) and modelling microseisms from observed and simulated ocean waves (e.g. Kedar et al., 2008; Graham et al., 2010; Ardhuin et al., 2011, 2012, 2014, 2015; Ardhuin & Roland, 2012; Stutzmann et al., 2012; Ardhuin & Herbers, 2013).

Another method, used to estimate the microseism source direction, is the cross-correlation functions (CCF) of seismic signal between station pairs. Compatible amplitudes for causal and acausal signals are less common, and it is because the source strengths of background noises are not spatially homogeneous (Chen et al., 2011). Because the causal and acausal signals of CCFs are mainly excited by the sources from opposite directions along the line of the station pair, the characteristics of CCF amplitudes has become a useful means to explore the source heterogeneities of ambient noises (e.g. Gu et al., 2007; Brzak et al., 2009; Chen et al., 2011).

Chen et al. (2011), for instance, using the temporal variations of the CCF for the observed SPSM indicate local sources generated from wind-driven ocean waves around offshore Taiwan.

Guo et al. (2020) applied a grid-search procedure to explore the frequency-dependent source locations of the PM and SM by the asymmetric cross-correlation function

(CCF) between sufficiently long ambient noise recordings at pairs of stations located in eastern North American margin.

1.3 MICROSEISM AND APPLICATIONS

One of the most important applications of microseism is the correlation with the ocean and the atmosphere parameters for the climate change studies. Already in the 19th century, Bertelli emphasized a correlation between the signal recorded in Florence by a Galileo pendulum (known as a tromometer) and barometric lows, and he suspected the influence of coastal sea waves (Bernard, 1990). The relation between microseismic noise and storms or hurricanes has been extensively investigated for the purpose of locating storms (Gilmore, 1946; Gutenberg, 1947; Gilmore and Hubert, 1948; Hjortenberg and Nikolaev, 1990) and the noise source area (e.g. Friedrich et al., 1998).

Microseism amplitudes show strong seasonal modulation, Stutzmann et al. (2000), for instance, observed an increase and a shift toward longer periods in autumn and winter than in spring and summer of the microseismic amplitude, explaining this phenomenon with the increase of the number and the amplitude of oceanic storms in autumn and winter. Indeed, at temperate latitudes, microseism shows periodicity, with maxima during the winter seasons, when the oceans are stormier, and minima during the summers (Aster et al., 2008). As above mentioned, this modulation is different along the coastlines of the Glacial Arctic Sea and the Southern Ocean where, during the winters, because of the sea ice, the oceanic waves cannot efficiently excite seismic energy (Aster et al., 2008; Stutzmann et al., 2009; Tsai and McNamara, 2011; Cannata et al., 2019).

One possible consequence of a climate change and of an anthropogenic global warming over the past several decades is an increase in ocean wave heights and power caused by upper-ocean warming (Reugero et al., 2019). The quantitative link between wave climate and seismic noise has been studied by Bromirski et al. (1999) for California, and Essen et al. (1999) for the North-East Atlantic. The analysis of many seismic records around the world indicate upward trends in seismic noise levels that was interpreted as an indication of the increase in ocean wave heights (Aster et al., 2010). Besides, Grevemayer et al. (2000) analysed a 40-year-long record of wintertime microseism and observed an increase in the number of monthly days with strong microseism activity, hence inferring an increase over time in surface air temperatures and storminess of the northeast Atlantic Ocean. Lecocq et al. (2020) used analog seismograms to improve the database of ocean storms before the 1980s by providing additional data for the quantitative validation of ocean wave modeling for extreme events.

The link between microseism amplitudes and the ocean wave height has been empirically explored by several authors (e.g. Bromirski et al., 1999; Bromirski and Duennebier, 2002; Ardhuin et al., 2012; Ferretti et al., 2013, 2016, 2018). For instance, Bromirski et al. (1999) determined site-specific seismic-to-wave transfer functions in the San Francisco Bay area (California). Ferretti et al. (2013, 2016, 2018) found empirical relations to predict wave parameters as period and the significant wave height along the Ligurian coast (Italy) as a function of the spectral energy-content of the microseism. In addition, other authors have derived physics-based models of the generation of the different kinds of microseism from the sea state (e.g. Gualtieri et al., 2013; Ardhuin et al., 2015; Gualtieri et al., 2019; Stopa et al., 2019).

Recent studies investigated the relationships between microseism and sea ice concentration in the polar regions (Stutzmann et al., 2009; Grob et al., 2011; Tsai and McNamara, 2011; Anthony et al., 2015, 2017; Cannata et al., 2017a, 2019).

Another application of microseism and seismic noise is its potential use for Earth structure study. Many techniques have been developed, each focusing on different aspects of data; examples include, in historical order, spatial variation of correlation amplitudes, array analysis for Rayleigh wave phase velocity measurements, use of horizontal–vertical amplitude ratios for Rayleigh wave signals, and the seismic interferometry that is two-station correlation technique to recover Green's functions between a pair of stations (Tanimoto et al., 2015). Recently, implementations of this technique used passive recordings at individual stations to detect changes of seismic velocity in the crust by phase cross-correlation of each component for each individual station with itself (e.g. De Plaen et al., 2019). Applications of seismic interferometry started with Campillo & Paul (2003) that recovered signals coming from cross-correlation of seismic coda between station pairs particularly in the microseismic frequency bands . The technique was soon applied from local- to regional-scale problems such as Southern California (Sabra et al., 2005; Shapiro et al., 2005). Typical seismic interferometry applications first recover Green's functions by cross correlation of seismograms at two stations. Since data are mainly from vertical components and are dominated by Rayleigh waves, traditional dispersion measurements for group velocity are often used to retrieve Earth structure (Tanimoto et al., 2015). Applications include the evaluation of building responses to seismic waves, tomography for crustal properties, changes in crustal properties over time, ground-roll removal from seismic data, and waveform modelling (Curtis et al., 2006).

Other applications on seismic interferometry in the microseism frequency range allows to investigate, for instance, the relationships between crustal velocity changes and volcano-tectonic activities at Mt. Etna volcano (Cannata et al., 2017b).

1.4 MACHINE LEARNING TECHNIQUES IN SEISMOLOGY

Machine Learning (ML) is the subfield of computer science that gives "computers the ability to learn without being explicitly programmed" (Samuel, 1959). Statistical (and/or mathematical) techniques are used in order to construct a model from observed data, rather than codifying a specific set of instructions that define the model for that data (Bishop, 2006). The era of big data calls for automated methods of data analysis which is what machine learning provides. In particular, machine learning is a set of methods able to automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (Murphy, 2012).

Most ML algorithms can be grouped into two main categories: supervised learning and unsupervised learning, depending on whether the data have target labels or not (Fig. 1.6).

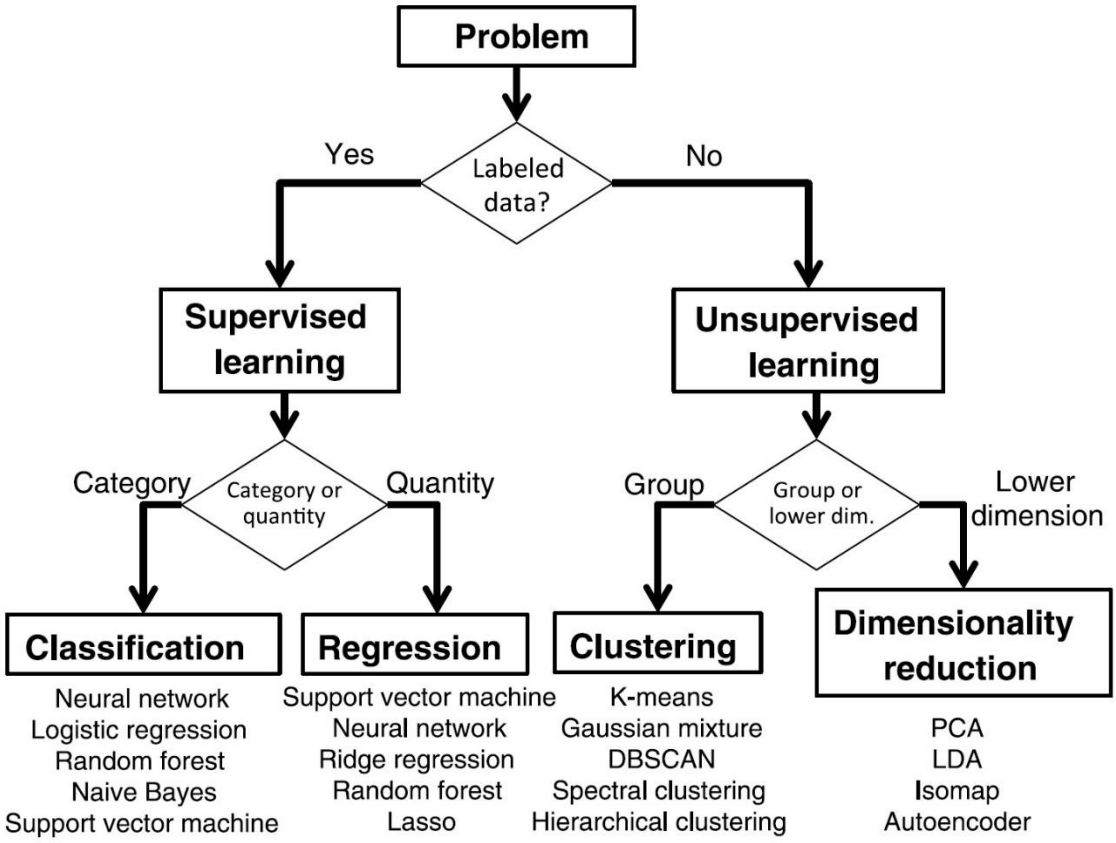

Figure 1.6: Types of machine learning (ML) algorithms. Supervised ML operates on labelled datasets with the objective to develop models that predict either categorical or quantitative target variables. Unsupervised ML operates on unlabeled datasets with the objective to group data by similarity or reduce the dimensionality of the input datasets. Some common ML algorithms are listed at the bottom for each category. (from Kong et al., 2018)

Supervised learning, which comprises predictive modeling and operates on labeled datasets, can be further subdivided into classification and regression algorithms based on whether the target outputs are categorical (classification) or quantitative (regression). Unsupervised learning is subdivided into clustering and dimensionality reduction, depending on whether we are interested in grouping data into categories based on similarity, or simply reducing the input data dimensions (Kong et al., 2018).

Furthermore, another type of ML approach is based on the concept of "reinforcement learning", where the machine learns its behaviour based on feedback from the "external environment".

ML algorithms, although diverse in their implementation, tend to follow a basic workflow that includes the following steps (Kong et al., 2018):

1. data collection: data are collected and partitioned into training and testing sets. A key aspect of ML is training the model on a random subset of the dataset, and then verifying the model on independent testing data;
2. preprocessing: data are cleaned and formatted, and missing data are removed or repaired. Feature extraction, which increases the performance of many ML algorithms by transforming the raw data into a more useful state for a given task, may also be performed;
3. model training: numerical optimization algorithms are used to iteratively tune the model parameters based on a cost function specific to the learning task of the problem;
4. model evaluation: model performance is evaluated on test data;
5. production: the finished ML model is applied in production mode to new data.

Nowadays, all these approaches find applications in many sectors including medical, social and financial disciplines. The number of applications of ML is growing impressively in geosciences too, including geophysics (Aminzadeh & de Groot, 2006). Examples of applications are seismic facies recognition and classification (e.g. Wrona et al., 2018), automatic interpretation of geophysical data, well log analysis and risk mitigation. Another field where ML can find useful applications is

integration of multidisciplinary information, like seismic, electromagnetic, gravity and magnetic data (Dell'Aversana, 2014).

Microseism investigations and, more generally, seismological studies are currently undergoing a rapid increase in dataset volumes (e.g. Kong et al., 2018; Jiao & Alavi, 2019). For this reason, nowadays, applications of machine learning techniques (hereafter referred to as MLTs) on seismological data are increasing in number day by day. Such techniques are used to extract information directly from data using well-defined optimization rules and help unravel hidden relationships between distinct parameters, as well as build predictive models (e.g. Kuhn & Johnson, 2013; Kong et al., 2018). Examples of the applications of MLTs to seismology include earthquake detection and phase picking (e.g. Wiszniowski et al., 2014), earthquake early warning (e.g. Kong et al., 2016), automatic waveform detection for amplitude-based microseismic imaging methods (e.g. Chen, 2018), improvement of Signal-to-Noise ratio (SNR) in processing of seismic reflection imaging (Chen et al., 2019), ground motion prediction for seismic hazard assessment (e.g. Trugman & Shearer, 2018), waveform correlation for detect repeating seismic events in an aftershock sequence (Bergman & Rabin, 2019), scanned images of old analog seismograms for earthquakes detection (e.g. Wang et al., 2019), improvement on seismic tomography and earthquake geodesy (Kong et al., 2018).

Focusing on supervised learning and in particular on regression analysis, we made use of the following MLTs to build predictive models that will be shown in chapters 3 and 4: (i) random forest (RF) regression; (ii) K-nearest neighbors (KNN) regression; (iii) linear regression; (iv) support vector machine (SVM) regression and (v) extremely randomized trees regression.

As for the RF technique, it is based on decision trees often used for classification and regression (Ho, 1995). One of the main problems with decision trees is the need to increase accuracy and avoid overfitting at the same time (Ho, 1998). RF overcomes such a limitation by generating many decision trees and outputting the mean prediction of the individual trees (Liaw & Wiener, 2002). Recently, RF has had many applications in geosciences, such as geochemical mapping (Kirkwood et al., 2016) and the lithological classification of underexplored areas by geophysical and remote sensing data (Kuhn et al., 2018).

K-nearest neighbors is a non-parametric technique applied for both classification and regression tasks (Altman, 1992). KNN regression simply predicts a new sample using the K-closest samples from the training set (Altman, 1992; Kuhn & Johnson, 2013). Hence, for a new input, the output is the average of the values of its K-nearest neighbors in the feature space of the training set. Such a method has been extensively used to classify remote sensing images (e.g. Li & Cheng, 2009; Noi & Kappas, 2018). Concerning linear regressions, relationships are modelled using linear predictor functions; that is, the relationship between predictors and responses falls along a hyperplane (Kuhn & Johnson, 2013). Similar to the two previous machine learning methods, linear regressions have been used in many fields of Earth Sciences, such as iron mineral resource potential mapping (Mansouri et al., 2018) and catchment-level base cation weathering rates (Povak et al., 2014).

SVMs are supervised learning models for both classification and regression analysis (e.g. Drucker et al., 1997; Kuhn & Johnson, 2013). As for regression, the SVM's goal is to find a function that deviates from each training point by a value no greater than a chosen constant, and at the same time is as flat as possible (e.g. Vapnik, 2000; Kuhn

& Johnson, 2013). Also, SVM has been applied in Earth Sciences, for instance to map landslide susceptibility (Reza Pourghasemi et al., 2013) and to classify remote sensing data (Jia et al., 2019).

Finally, extremely randomized trees regression is based on random forest and it applies a fully random selection to split data in the test nodes (Geurts et al., 2006).

Each of the above-mentioned techniques has its own advantages and disadvantages (e.g. Kuhn &Johnson, 2013; Yang et al., 2019). The main advantages of RF are its high accuracy and robustness to outliers and noise; also, RF parameter tuning does not have a drastic effect on performance. The disadvantages are the expensive training time and overfitting in the case of small datasets. KNN is effective and non-parametric, but it is not robust in the presence of noise and it is not easy to identify the best K value. As for linear regressions, they require short training times, and the results are easy to visualize and understand, but they are not suited to model non-linear relationships. Finally, SVMs are easy to implement and show good efficiency in training and generalization, but the tuning of parameters can be quite difficult.

In the last decade there have been rapid advances in a class of ML methods called deep neural networks, or deep learning (LeCun et al., 2015; Jordan & Mitchell, 2015). Given enough data, these deep networks can learn very complex, nonlinear relationships between input data and prediction targets, leading to accurate predictions on a number of challenging tasks (Bergen et al., 2019). Deep neural network architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are widely applied in image recognition, natural language processing, and robotic control tasks (Bergen et al., 2019). Their application in seismology include event detection, phase picking and identification, phase

association, hypocenter location and event characterization (for reference see Bergen et al., 2019).

2. MICROSEISM AND SEA ICE RELATIONSHIP IN ANTARCTICA

2.1 INTRODUCTION

Microseism amplitudes at temperate latitudes in both northern and southern hemispheres show strong annual periodicity with maxima during the winter seasons, when the oceans are stormier, and minima during summers (Aster et al., 2008).

However, such a pattern is different in Antarctica where during the winter, because of the sea ice, the oceanic waves cannot efficiently excite seismic energy (Aster et al., 2008; Stutzmann et al., 2009; Grob et al., 2011; Anthony et al., 2015, 2017; Turner et al., 2020). Although the link between microseism and ocean wave parameters has been quantitatively explored (e.g. Ardhuin et al., 2012; Bromirski et al., 1999, 2002), the relationship between sea ice concentration and microseism recorded in Antarctica has only been qualitatively treated.

An exception is the investigation performed by Anthony et al. (2017), who quantitatively studied such a relationship only in the Antarctic Peninsula, finding that primary microseism (PM) energy is both more sensitive to sea ice and more capable of propagating across ocean basins than secondary microseism (SM) energy. In addition, to date no technique has been found to try to infer the sea ice distribution in the Southern Ocean, based on microseism recordings.

In the following sections, we quantitatively investigate the relationship between microseism, recorded along the Antarctic coasts and filtered in three distinct period bands (primary microseism PM: 13–20 s, secondary microseism SM: 5–10 s and short-period secondary microseism SPSM: 2.5–5.0 s), and sea ice concentration by

means of statistical analysis and machine learning applications, allowing to spatially and temporally reconstruct the sea ice distribution around Antarctica based on the microseism amplitudes.

2.2 DATA

2.2.1 SEISMIC DATA

All the available seismic data, recorded by the vertical component of 20 stations from 1993 to 2017, were used. These stations, chosen because located close to the Antarctica grounding lines (maximum distance equal to ~260 km for SILY) and then to those microseism sources, mainly related to the energy transfer of ocean waves breaking/shoaling against the shoreline (PM, Hasselmann, 1963) or to local nearshore wave-wave interaction (SPSM, Bromirski et al., 2005) (Fig. 2.1), belong to the following seismic networks (Table 2.1): Antarctic Seismographic Argentinean Italian Network (BELA, ESPZ, ORCD, SMAI), POLENET (BEAR, CLRK, DNTW, HOWD, MECK, MPAT, SILY, THUR), Global Seismograph Network (CASY, PMSA, SBA, VNDA), GEOSCOPE (DRV), Geoscience Australia (MAW), GEOFON (SNAA), Pacific21 (SYO).

The data were acquired at a sampling rate of 20 or 40 Hz (Table 2.1) by seismic stations equipped with different broadband seismometers (Guralp CMG-3T, Streckeisen STS2, Streckeisen STS1, Nanometrics Trillium 240, Geotech KS-54000). The data were downloaded from the website of Incorporated Research Institutions for Seismology (IRIS; https://ds.iris.edu/SeismiQuery/station.htm). Although the considered time interval ranges from 1993 to 2017, the temporal coverage of the acquired data significantly varies from station to station (Fig. 2.2).

Station Name	Latitude	Longitude	Altitude (m a.s.l.)	Network	Sensor	Sampling Rate (Hz)
BELA	-77.875	-34.6269	262	ASAIN	Guralp CMG-3T	20
ESPZ	-63.3981	-56.9964	31	ASAIN	Guralp CMG-3T	20
SMAI	-68.1302	-67.1059	9	ASAIN	Guralp CMG-3T	20
ORCD	-60.7381	-44.7361	20	ASAIN	Guralp CMG-3T	20
MAW	-67.604	62.871	12	Geoscience Australia	Streckeisen STS-2	40
DRV	-66.664908	140.002069	40	GEOSCOPE	Streckeisen STS-1	20
SNAA	-71.6707	-2.8379	846	GEOFON	Streckeisen STS-2	20
VNDA	-77.517275	161.852758	51 (100 m depth)	Global Seismograph Network	Geotech KS-54000	40
CASY	-66.2792	110.53554	5	Global Seismograph Network	Streckeisen STS-1	20
PMSA	67.7744	-64.0489	40	Global Seismograph Network	Streckeisen STS-1	40
SBA	-77.8492	166.7572	48	Global Seismograph Network	Guralp CMG-3T	40
SYO	-69.0067	39.585	20	Pacific21	Streckeisen STS-1	20
BEAR	-74.5476	-111.8511	384	POLENET	Guralp CMG-3T	40
CLRK	-77.3231	-141.8485	1042	POLENET	Nanometrics Trillium 240	40
DNTW	-76.4512	-107.7769	1031	POLENET	Nanometrics Trillium 240	40
HOWD	-77.5286	-86.7694	1495	POLENET	Nanometrics Trillium 240	40
MECK	-75.2808	-72.185	1085	POLENET	Nanometrics Trillium 240	40
MPAT	-78.0297	-155.0221	539	POLENET	Nanometrics Trillium 240	40
SILY	-77.1332	-125.966	2092	POLENET	Nanometrics Trillium 240	40
THUR	-72.5301	-97.5606	238	POLENET	Guralp CMG-3T	40

Table 2.1: List of seismic stations used in this work. The data are coming from IRIS website.

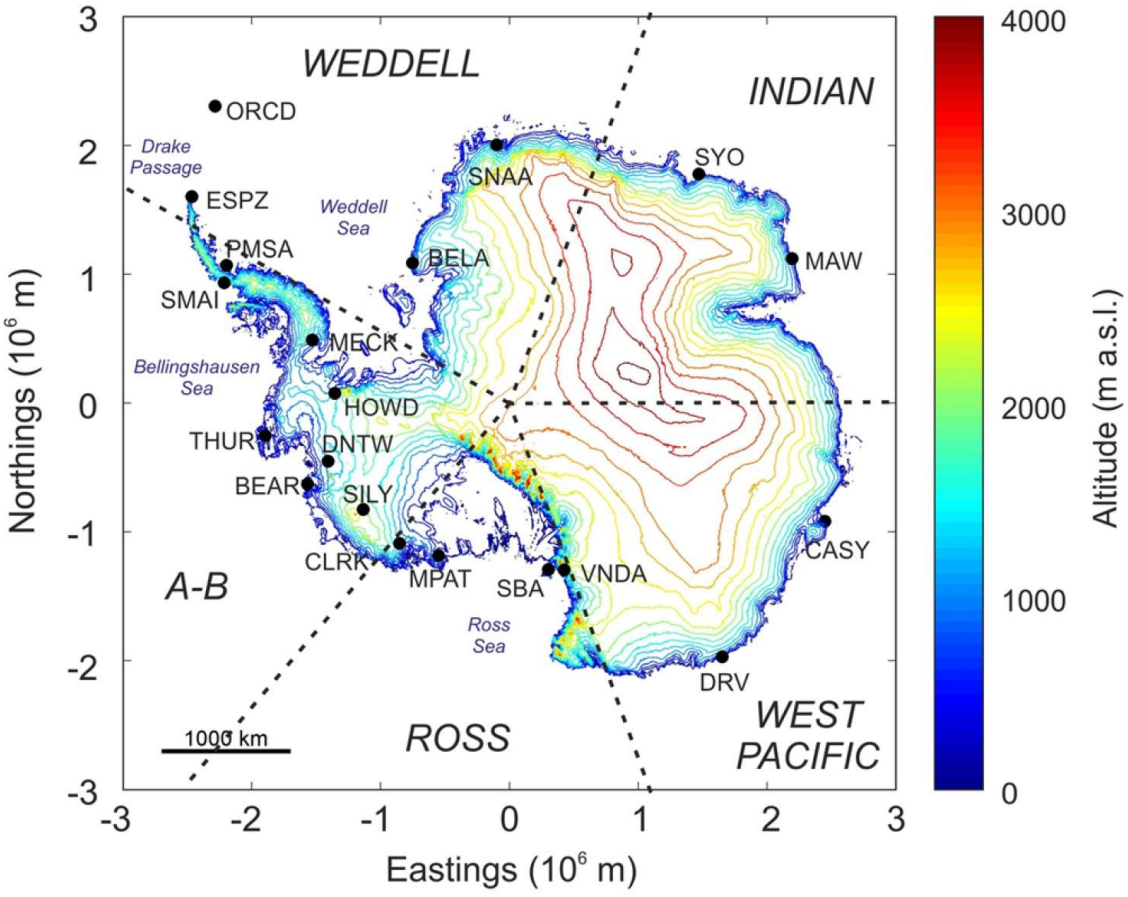

Figure 2.1: Antarctica map. Digital elevation model of Antarctica (data derived from CryoSat-2 altimetry; Slater et al., 2018); plotted by the Antarctic Mapping Tools (Greene et al., 2017), showing the locations of the seismic stations used in this study. The dashed black lines divide Antarctica into 5 sectors named Weddell, Indian, West Pacific, Ross and A-B (acronym of Amundsen-Bellinghausen) (Holland, 2014).

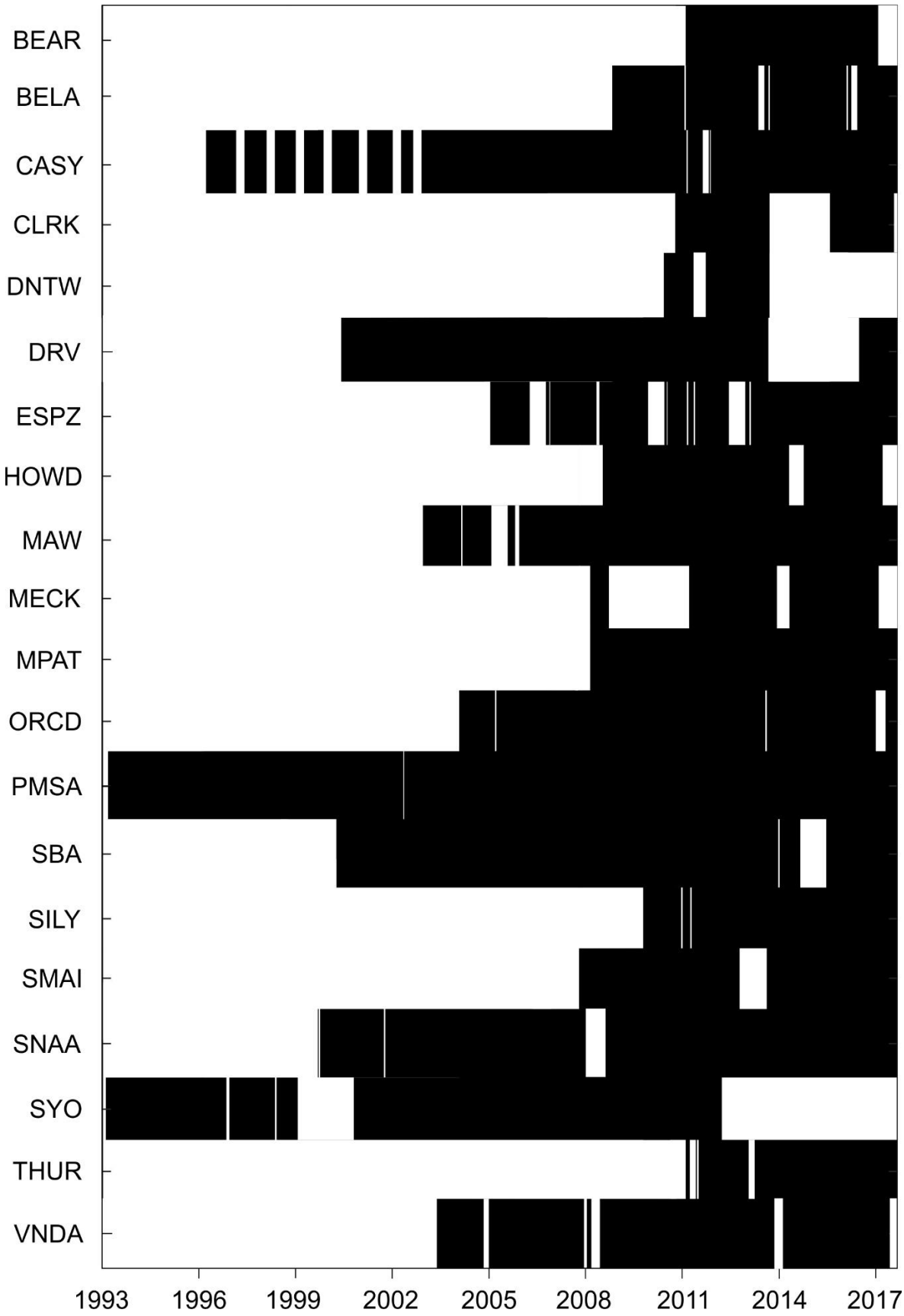

Figure 2.2: Data availability for the seismic stations used in this study. Black areas indicate periods when data were successfully recorded, white areas show periods when the stations were down.

2.2.2 SEA ICE DATA

Information about temporal and spatial variability of sea ice concentration, defined as the percentage of ice cover within each 25 × 25 km^2 cell of a grid comprising the entire Antarctic polar sea ice cover, are obtained by brightness temperature data (Fetterer et al., 2017). Such data were downloaded as GeoTIFF files, providing the daily sea ice concentration data in a georeferenced format (Fig. 2.19b-c). In particular, the version 2.1 files were used for the period 2000–2016, the version 3.0 files for the period 1993–1999 and 2017. The two versions have no difference in terms of daily sea ice concentration (Windnagel et al., 2017). The sea ice concentration is represented with a scale ranging from 0 to 1000. We divided this value by 10 to get data in percent. It has to be noted that values lower than 150 (15%) are considered statistically irrelevant because of instrumental limits. Such a parameter is affected by larger uncertainty during summer, because of the thinner sea ice (Steffen et al., 1992).

2.3 METHODS OF ANALYSIS

2.3.1 SPECTRAL AND RMS AMPLITUDE ANALYSES

Once the data were downloaded, they were corrected for the instrument response, and the Short Time Fourier Transform (STFT) of the data recorded by the vertical component of the 20 stations (shown in Fig. 2.1) was calculated as follows: spectra over 81.92-second-long sliding window were computed, and all the spectra falling in the same day were averaged by Welch's overlapped segment averaging estimator (Welch, 1967), and visualised as spectrograms (Fig. 2.3).

As for the root mean square (RMS) amplitudes, three distinct period bands were taken into account: PM (13–20 s), SM (5–10 s), and SPSM (2.5–5.0 s) (Anthony et al., 2017).

If s_n, $n = 0, 1.., N - 1$ is a segment of the signal encompassing N points, the RMS in a specified frequency band is calculated from the filtered signal s_n^F :

$$RMS = \sqrt{\frac{\sum_{n=0}^{N-1}(s_n^F)^2}{N}} \qquad (2.1)$$

It is also possible to calculate the RMS without filtering using Parseval's theorem. If we apply the discrete Fourier transform to the signal s_n, we get its spectrum S_k , $k= 0,1.., N – 1$. The RMS of the unfiltered signal is then:

$$RMS = \frac{1}{N}\sqrt{\sum_{k=0}^{N-1}|S_k|^2} \qquad (2.2)$$

If we wish to consider a specific frequency band, we only calculate the sum over the chosen part of the spectrum. For continuous signals, the RMS value is much more stable than the peak-to-peak amplitude which may be strongly influenced by a single peak in the signal or by a glitch due to a technical problem (Battaglia & Aki, 2003).

The RMS amplitude time series were obtained by gathering median daily RMS amplitudes, computed on values calculated over consecutive 81.92-sec-long windows (Fig. 2.4). Successively, the RMS amplitude time series were smoothed by a 90-day-long moving median, split in year-long windows, that were stacked and normalised by subtracting the minimum value and dividing by the maximum value (Fig. 2.5a).

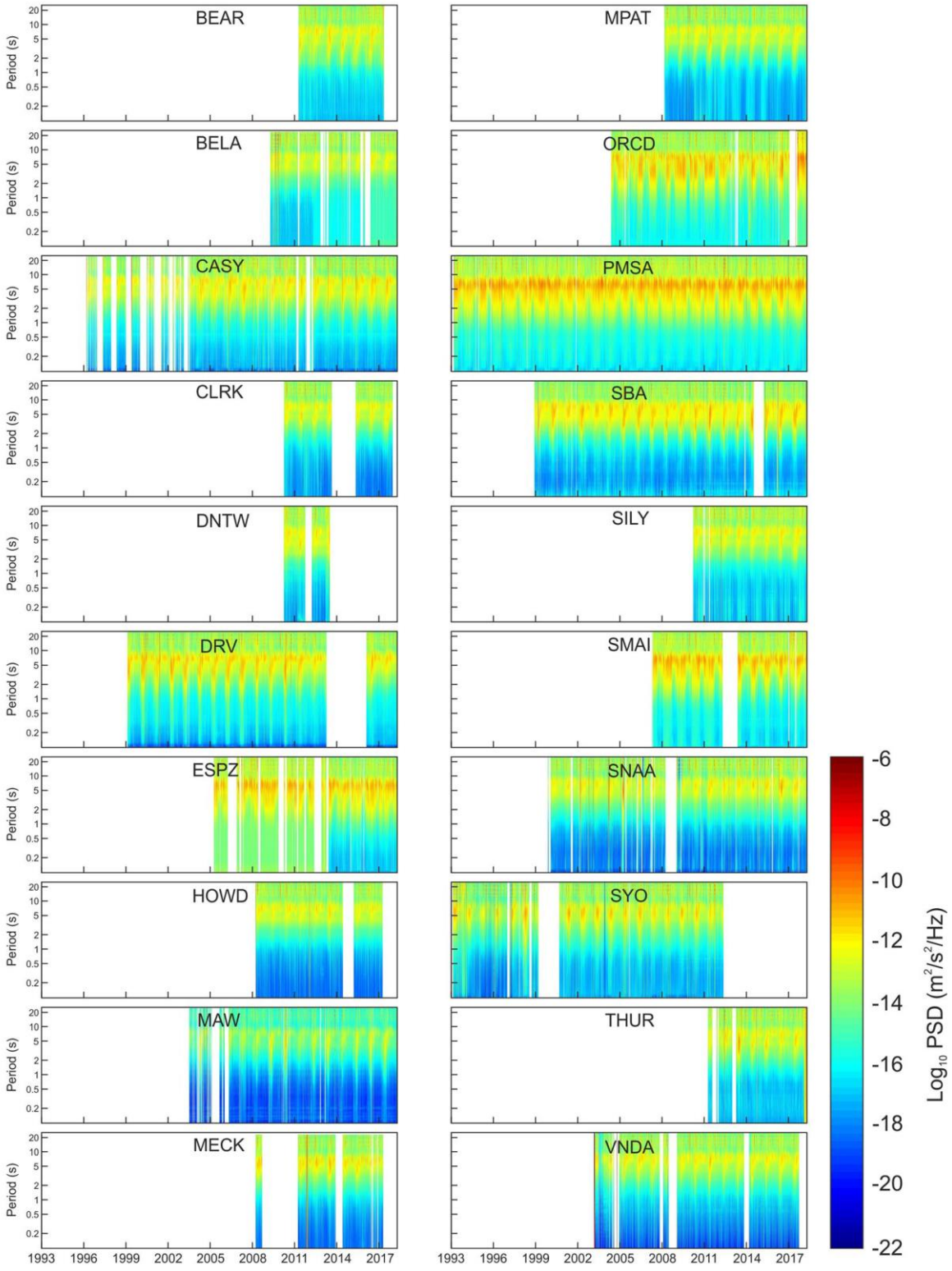

Figure 2.3: Seismic spectrograms. Spectrograms of the seismic signal recorded by the vertical component of the 20 considered stations.

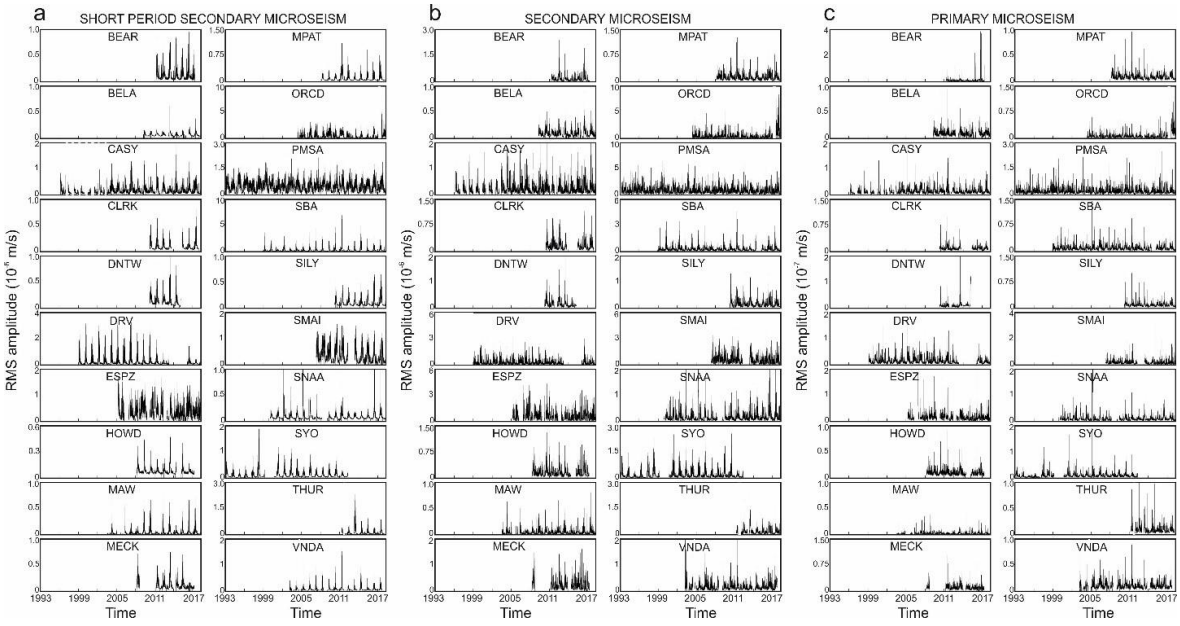

Figure 2.4: RMS amplitudes of the seismic signal recorded by the vertical component of the 20 considered stations and filtered in the band (a) 2.5-5.0 s (SPSM), (b) 5-10 s (SM) and (c) 13-20 s (PM).

Then, all these stacked normalised RMS amplitudes of all the stations were again stacked to have the overall seasonal trends in the three distinct period bands (Fig. 2.5b). Finally, the three curves were stacked, thus obtaining a single curve (Fig. 2.5c). The highest and lowest peaks in Fig. 2.5c indicate the onset time of the 90-day-long windows, characterised by the strongest (February–April) and weakest (October–December) microseism, respectively.

Furthermore, a median value of RMS amplitude was obtained per each station in the three considered period bands and, by applying a triangulation-based natural neighbour interpolation (Sibson, 1981), a map showing the spatial distribution of RMS amplitude values was plotted per each period band (Fig. 2.6a–c). Following Möllhoff & Bean (2016), to remove interpolated information from poorly constrained regions, we masked interpolated data for areas farther than 1000 km from any station (distance chosen to evidence the unreliable interpolated microseism amplitude

information in the inner part of Antarctica). The same technique was applied to generate maps focused on the periods with the maximum (February–April; Fig. 2.6d–f) and minimum (October–December; Fig. 2.6g–i) microseism amplitudes.

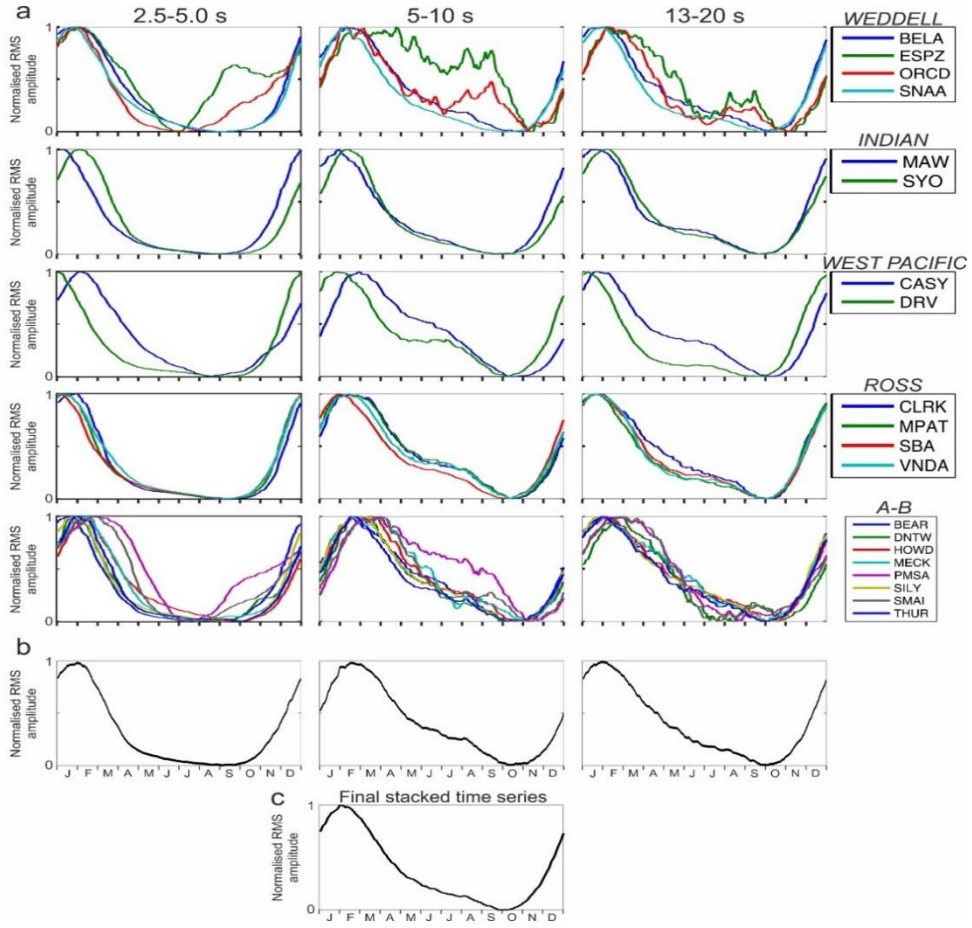

Figure 2.5: (a) RMS amplitude time series, smoothed by a 90-day-long moving median, split in year-long windows, stacked and normalised for all the considered seismic stations (see the legends on the right). In particular, left plots regard the period band 2.5-5.0 s, central plot 5-10 s and right plot 13-20 s. The time series are gathered in different plots, according to the different Antarctic sector, where the station is located (see Figure 2.1 for sector view). (b) Overall seasonal trend of microseism for the period band 2.5-5.0 s, 5-10 s and 13-20 s (left, central and right plot, respectively), obtained by stacking the RMS amplitudes of all the stations, smoothed and normalised, shown in (a). (c) Overall seasonal trend of microseism in Antarctica obtained by stacking the time series in (b). The time in x-axis in (a-c) indicates the window onset of the 90-day-long moving median.

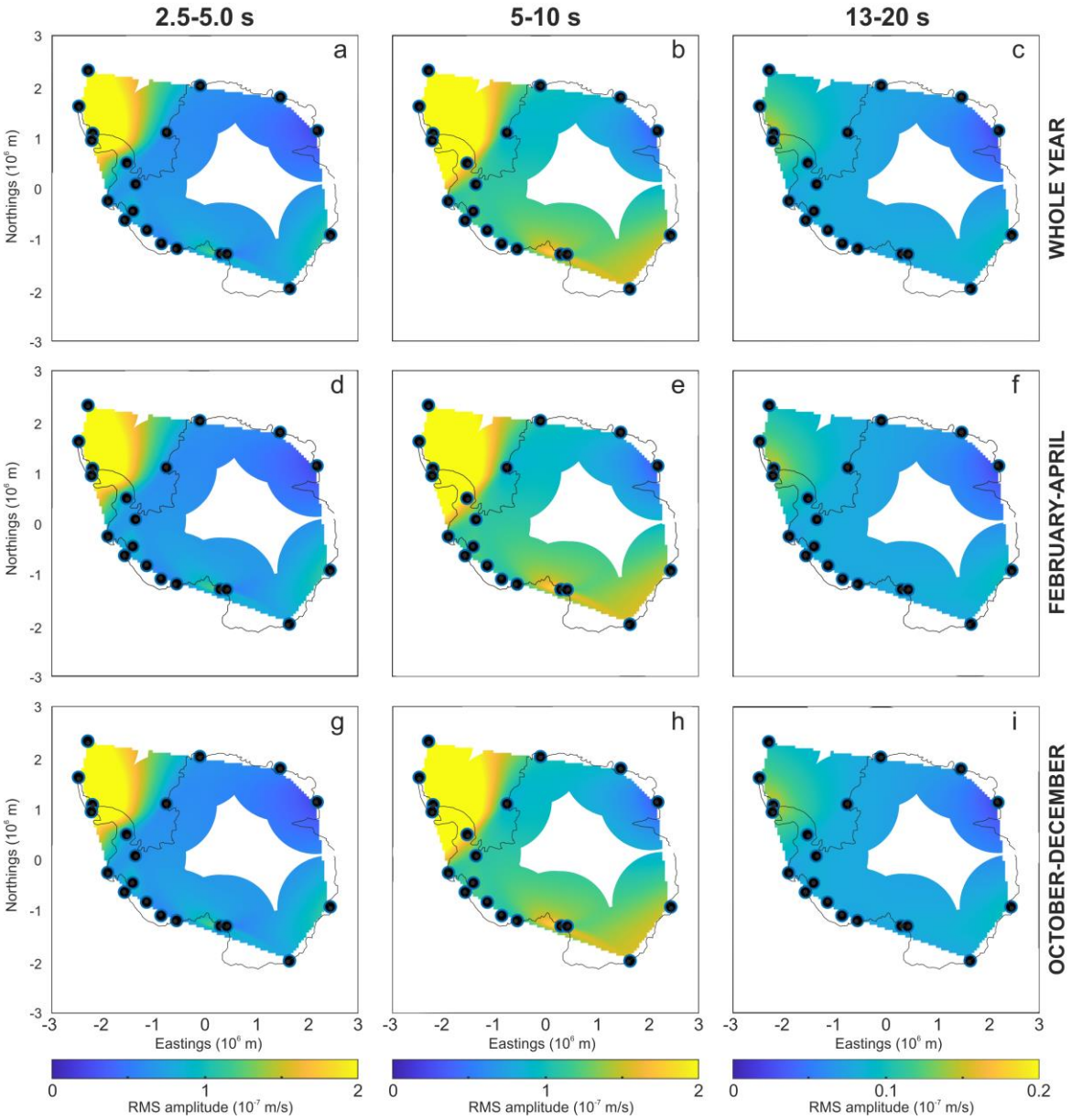

Figure 2.6: Microseism amplitude maps. Spatial distribution of the median values of RMS amplitude time series in the bands 2.5–5.0 s (SPSM; a,d,g), 5–10 s (SM; b,e,h) and 13–20 s (PM; c,f,i), obtained by applying a triangulation-based natural neighbour interpolation (Sibson, 1981). The plots (a–c) were obtained by taking into account the whole year, while the plots (d–f) and (g–i) are focused on the periods with the strongest (February–April) and weakest (October–December) microseism, respectively. The black dots indicate the locations of the considered stations. It has to be noted that the color scale used for PM 13–20 s (c,f,i) is different from the color scales used for SM 5–10 s (b,e,h) and SPSM 2.5–5 s (a,d,g).

2.3.2 CORRELATION ANALYSIS

Following the idea by Anthony et al. (2017), we quantitatively investigated the relationship between microseism and sea ice distribution for the three different period band: 2.5–5.0 s, 5–10 s and 13–20 s. Time series of daily sea ice concentration values per each grid node were obtained and compared with the RMS amplitude patterns by Spearman correlation coefficient.

Following Craig et al. (2016) and Xiao et al. (2018), in place of using the more common Pearson correlation coefficient, we made use of the Spearman correlation coefficient. This coefficient, defined as a nonparametric measure of rank correlation, was preferred with respect to the more widely used cross correlation coefficient, whose application is limited to explore linear dependence between normally distributed parameters (Myers & Well, 2003; Kanji, 2006). In addition, unlike the Pearson correlation coefficient, the Spearman correlation coefficient is less affected by outliers. The Spearman correlation coefficient can be computed as follows (Craig et al., 2016):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \qquad (2.3)$$

where d_i is the difference between the ranks of the two parameters and n is the number of observations. Then, spatial distributions of the Spearman correlation coefficient in the portion of Southern Ocean, seasonally covered by sea ice, were obtained (Fig. 2.7a,c,e).

In addition, we repeated the same correlation analysis by filtering the long-period components out in both seismic RMS amplitude and sea ice time series. To obtain

these filtered time series, we tested two methods: (i) we computed smoothed time series of RMS amplitude and sea ice by using a 15-day-long moving median, and subtracted from each value of the original RMS amplitude and sea ice series, the value of the smoothed series shifted back by one year; (ii) we divided the RMS amplitude and sea ice time series into 1-year-long windows, stacked them and smoothed them by using a 15-day-long moving median, and then we subtracted from each sample of the original time series the corresponding sample of the yearly stacked series. By using both methods, we obtained very similar time series. Also in this case, we obtained maps of spatial distribution of Spearman correlation coefficient (Fig. 2.7b,d,f).

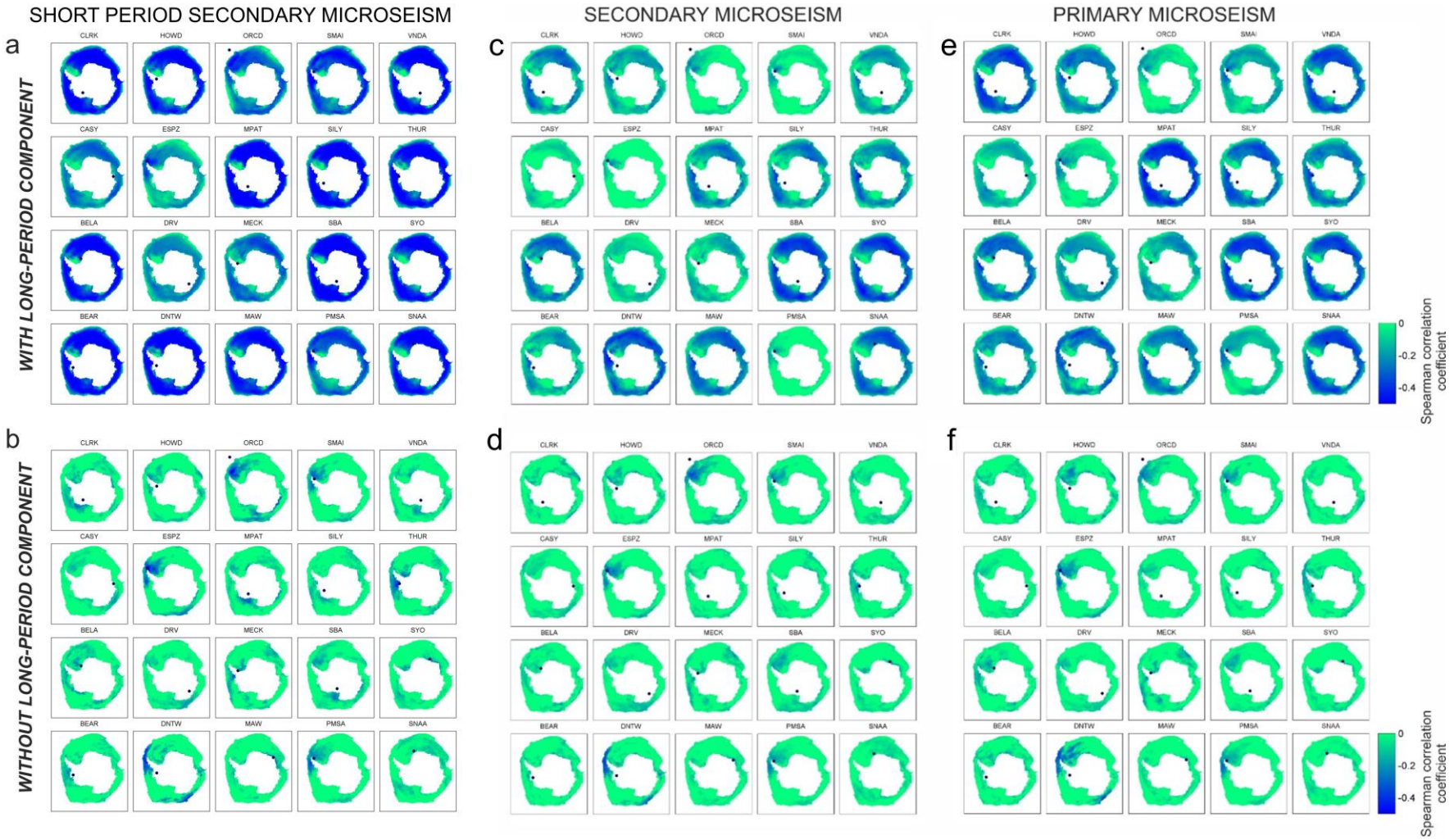

Figure 2.7: Maps of Antarctica showing the space distribution of the Spearman correlation coefficient computed between sea ice concentration time series and RMS amplitudes, calculated per per each considered station in the bands (a,b) 2.5-5.0 s, (c,d) 5-10 s and (e,f) 13-20 s. In (a,c,e) the long period components of the time series were present, in (b,d,f) the long-period components were filtered out before performing the correlation analysis. The black dots indicate the locations of the considered stations.

As both seismic RMS amplitudes and sea ice concentrations are characterized by a sampling period of 24 h (much longer than any physically possible delays between the two time series), a simple zero-lag correlation was computed.

To test whether the observed values of Spearman correlation coefficient are significantly different from zero or not (null hypothesis), the t-test, taking into account the different number of samples in the distinct RMS amplitude time series, was performed and the space distribution of p-value (probability value) was calculated; in particular, p-values lower than the significance level of 0.05 were considered sufficient to reject the null hypothesis (Fig. 2.8). Such a threshold means that the probability, that the result of the statistical test is due to chance alone, is less than 5%, so it would occur once out of 20 times the study is repeated. The value of 0.05 is a commonly accepted significance level used for this statistic test (e.g. Anthony et al., 2017).

Focusing on the filtered RMS amplitude and sea ice time series, we collected all the Spearman correlation values and the corresponding distances for all the stations and the three period bands. A cumulative 3D density plot was obtained per each period band, showing the distance in the x-axis, the Spearman correlation value in the y-axis, and the normalised number of Spearman correlation estimations with the color scale (Figs. 2.9a, 2.10a, 2.11a). Furthermore, 2D histograms, gathering the Spearman correlation values within given ranges of distance (from 0 to 6000 km, with step of 125, 250, 500, Figs. 2.12a-c, 2.13a-c, 2.14a-c and 1000 km), were obtained. Figures 2.9b-g, 2.10b-g and 2.11b-g show the results for the 1000-km step.

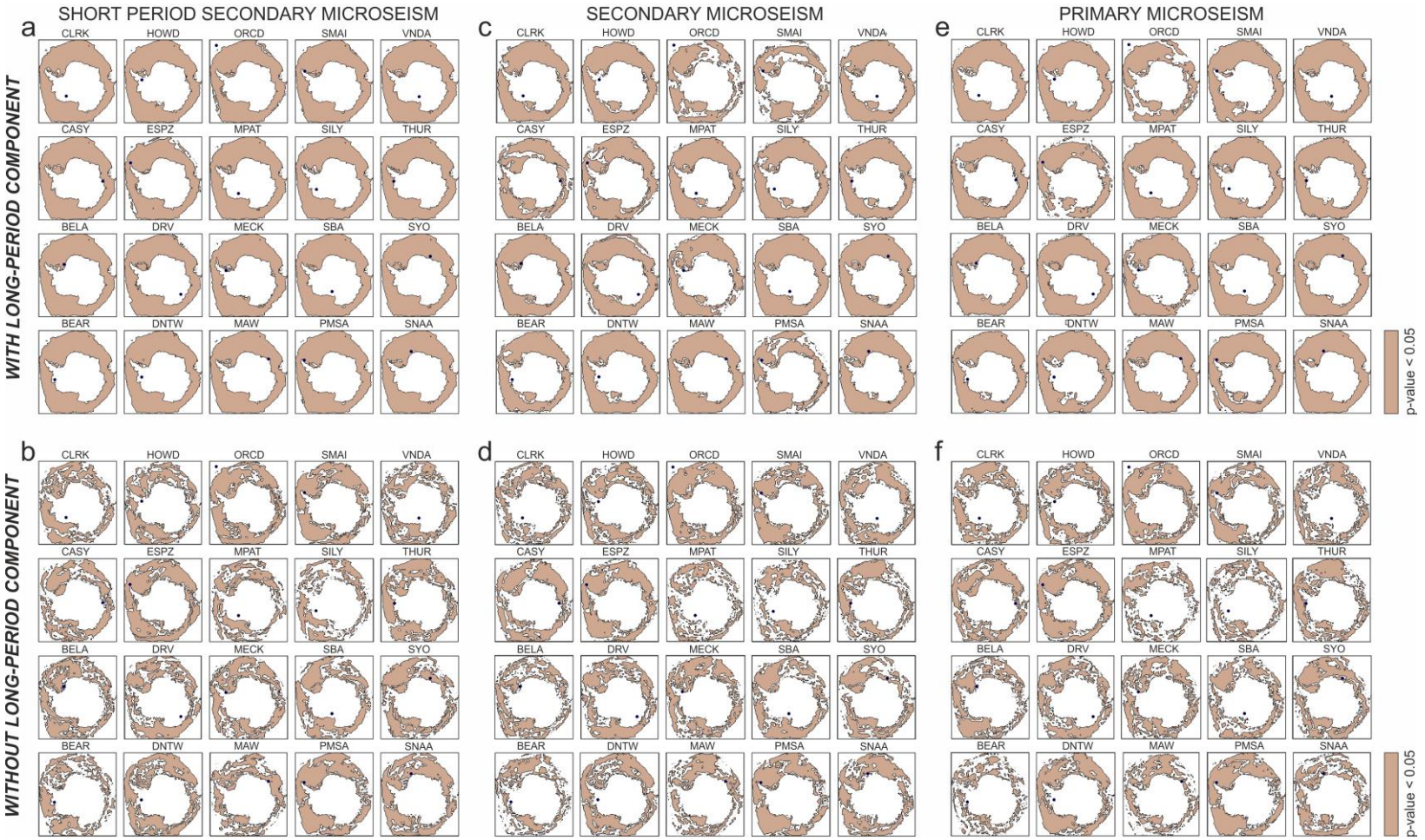

Figure 2.8: Maps showing the space distribution of p-values, obtained by the Spearman correlation analysis, calculated per each considered station in the bands (a,b) 2.5-5.0 s, (b) 5-10 s and (c) 13-20 s. The pink color indicates the areas where p-value is lower than 0.05. In (a,c,e) the long-period components of the time series were present, in (b,d,f) the long-period components were filtered out before performing the correlation analysis. The black dots indicate the locations of the considered stations.

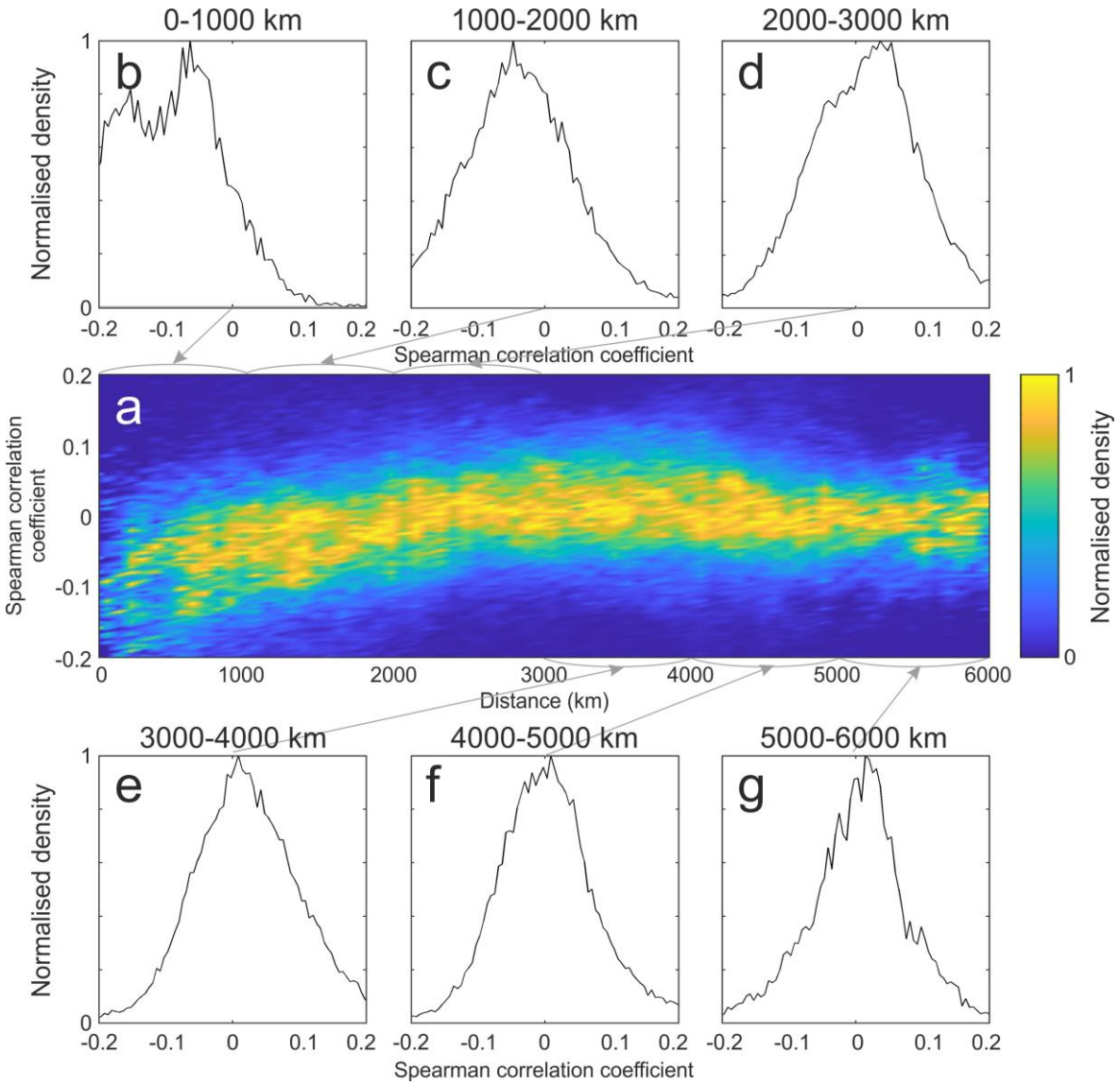

Figure 2.9: Density plots of Spearman correlation coefficient for short period secondary microseism. (a) 3D normalised density plot obtained for the band 2.5–5.0 s, showing the distance sea ice-seismic station in the x-axis, the Spearman correlation value in the y-axis, and the number of correlation estimations with the color scale. (b–g) 2D normalised histograms showing the Spearman correlation value in the x-axis and the number of correlation estimations in the y-axis for different distance ranges.

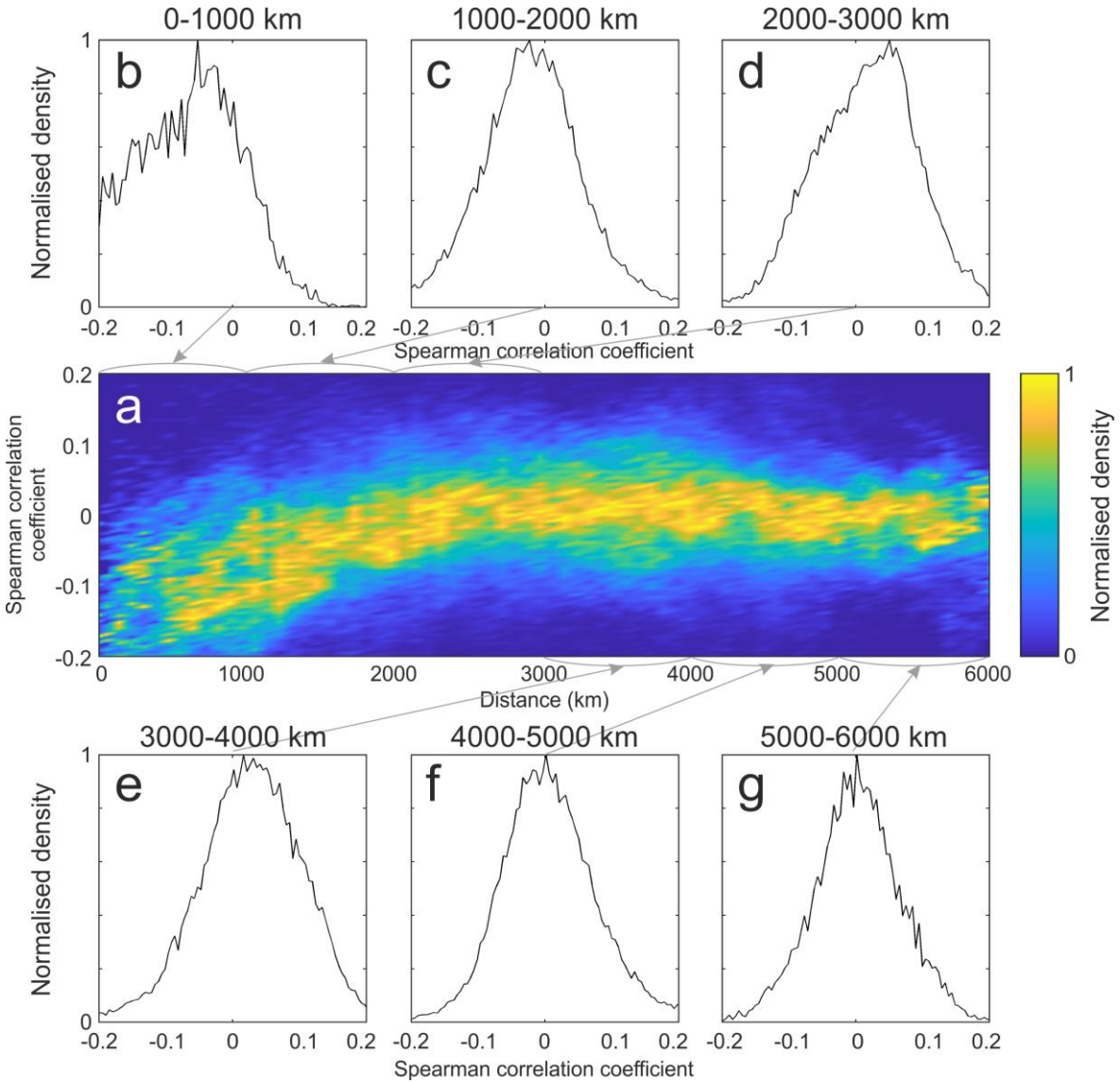

Figure 2.10: Density plots of Spearman correlation coefficient for secondary microseism. (a) 3D normalised density plot obtained for the band 5-10 s, showing the distance sea ice-seismic station in the x-axis, the Spearman correlation value in the y-axis, and the number of correlation estimations with the color scale. (b-g) 2D normalised histograms showing the Spearman correlation value in the x-axis and the number of correlation estimations in the y-axis for different distance ranges.

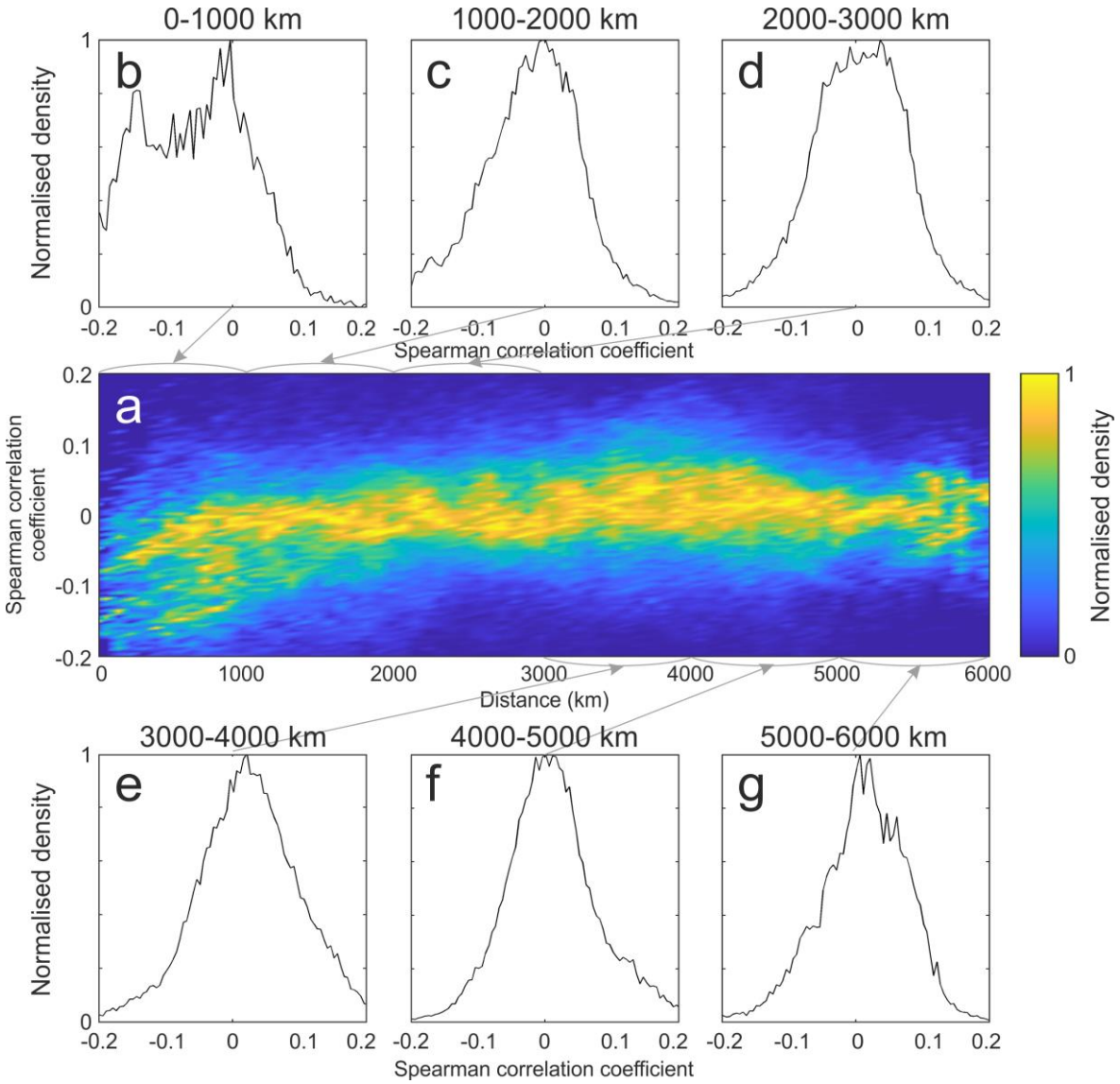

Figure 2.11: Density plots of Spearman correlation coefficient for primary microseism. (a) 3D normalised density plot obtained for the band 13-20 s, showing the distance sea ice-seismic station in the x-axis, the Spearman correlation value in the y-axis, and the number of correlation estimations with the color scale. (b-g) 2D normalised histograms showing the Spearman correlation value in the x-axis and the number of correlation estimations in the y-axis for different distance ranges.

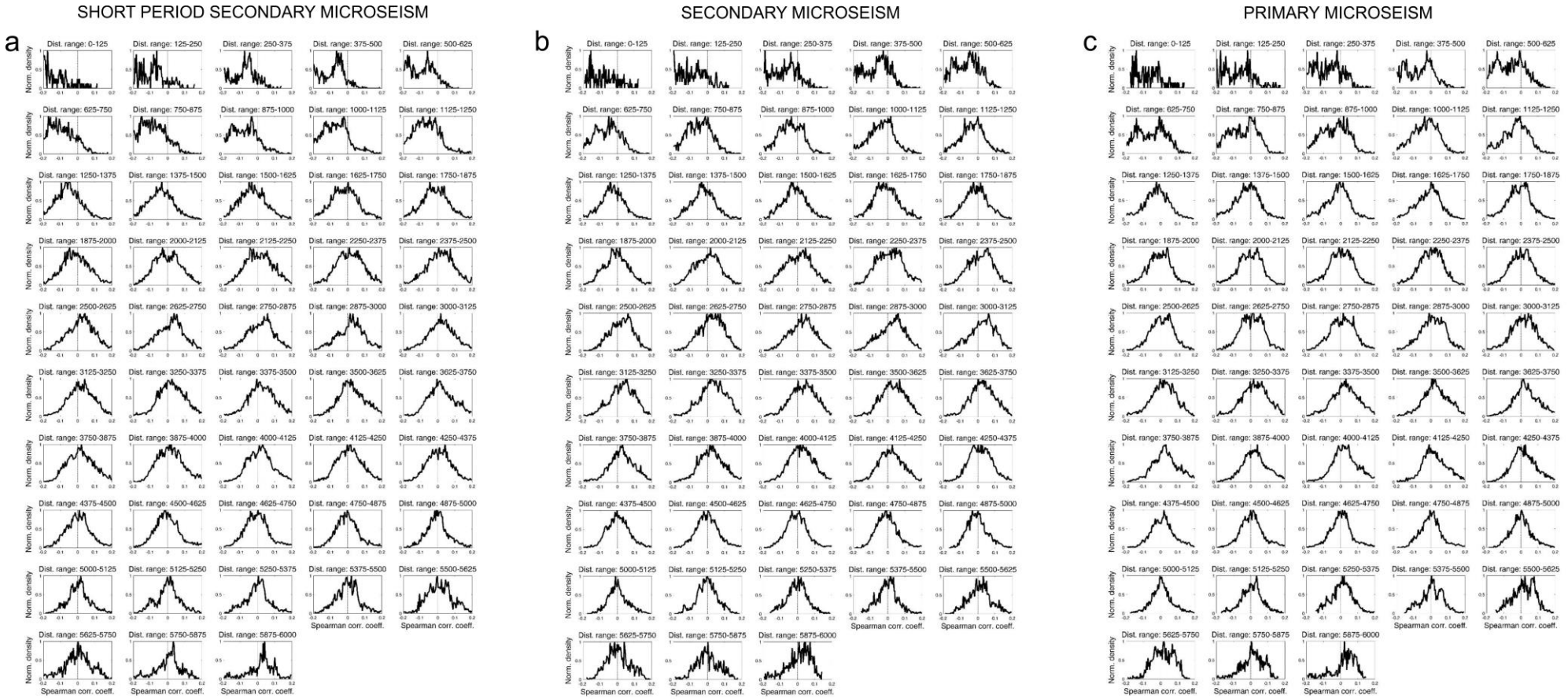

Figure 2.12: Normalised histograms showing the Spearman correlation values in the x-axis and the normalized number of correlation estimations in the y-axis for distances ranging from 0 to 6000 km, with interval of 125 km, obtained for the band (a) 2.5-5.0 s, (b) 5-10 s and (c) 13-20 s.

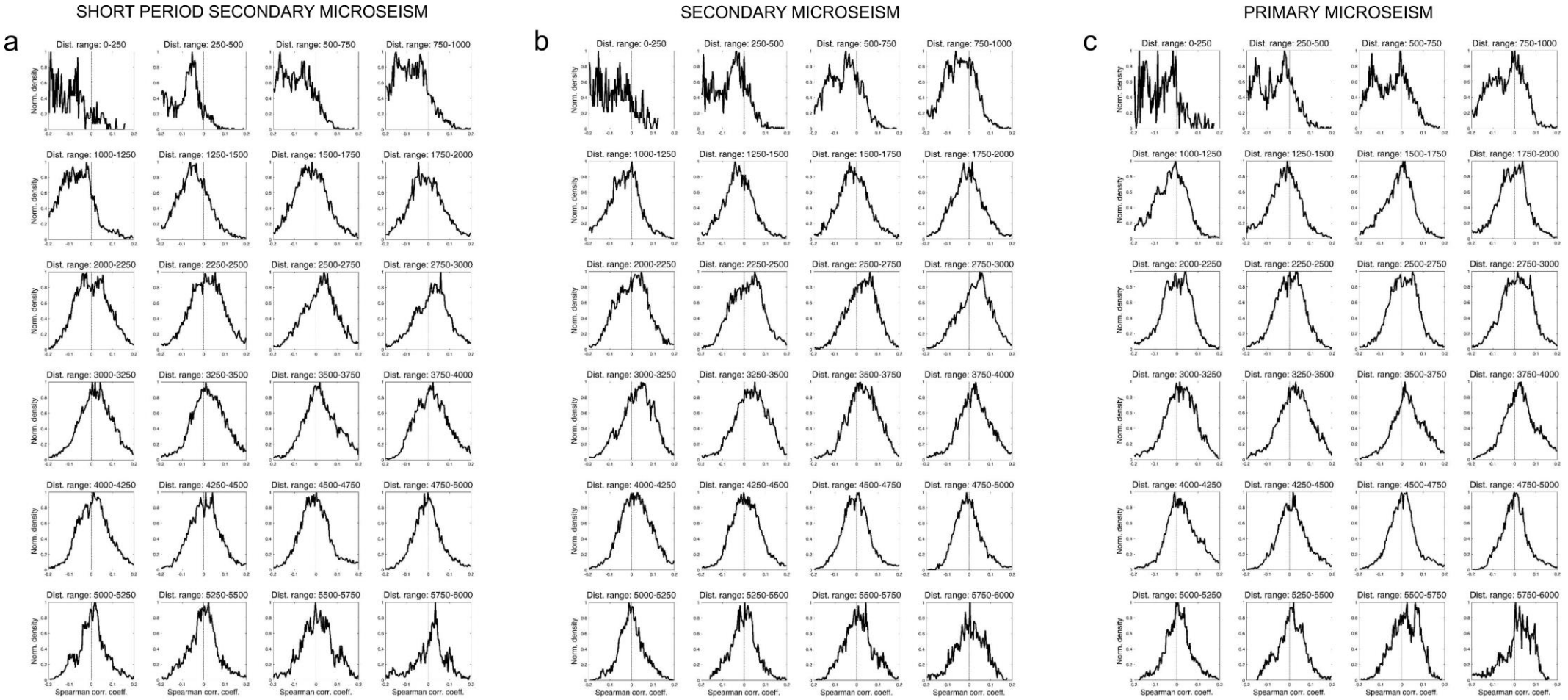

Figure 2.13: Normalised histograms showing the Spearman correlation values in the x-axis and the normalized number of correlation estimations in the y-axis for distances ranging from 0 to 6000 km, with interval of 250 km, obtained for the band (a) 2.5-5.0 s, (b) 5-10 s and (c) 13-20 s.

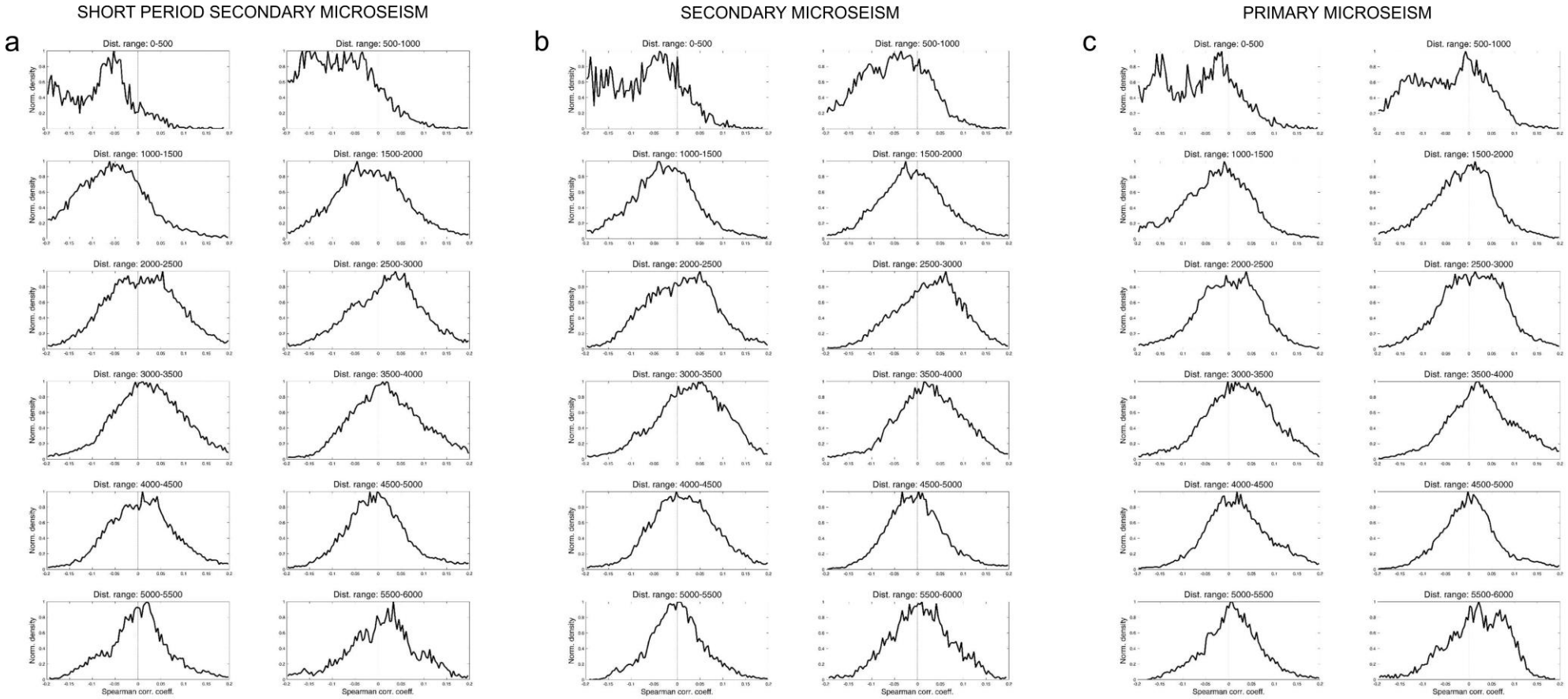

Figure 2.14: Normalised histograms showing the Spearman correlation values in the x-axis and the normalized number of correlation estimations in the y-axis for distances ranging from 0 to 6000 km, with interval of 500 km, obtained for the band (a) 2.5-5.0 s, (b) 5-10 s and (c) 13-20 s.

2.3.3 CALCULATION OF SEA ICE CONCENTRATION BY USING MICROSEISM AND MACHINE LEARNING

In this section we apply a new methodology able to predict the spatial daily sea ice concentration starting from microseism signal. The method is composed of three steps (Fig. 2.15): (i) data preparation; (ii) training; (iii) cross-validation.

As for the step (i), the time intervals characterized by an almost complete coverage of seismic data from 18 out of 20 stations were selected (2011–2012 and 2016–2017), and the corresponding RMS amplitude data were extracted. SYO and DNTW stations were not used because of the total lack of data during the most recent time interval (2016–2017). We applied the following transformations on all the RMS amplitude time series:

- Probability Integral Transformation (PIT, Dodge, 2006): to maximize the distribution regularity of the data in a range. The technique attempts to convert a random variable with any given continuous distribution into a random variable having a standard uniform distribution (Angus, 1994).
- The Linear Discriminant Analysis (LDA, McLachlan, 2004): to maximize the class-separability of clustered data. The goal of LDA is to project the dataset onto a lower-dimensional space with better class-separability. To define the classes of sea ice space distribution, we firstly found the best number of clusters that can separate the sea ice space distribution images by using the Calinski-Harabasz (Caliński & Harabasz, 1974) and the Silhouette (Rousseuw & Silhouettes, 1987) indexes in the range between 2 and 365. We found the optimum number of 32 different classes for the sea ice space distribution dataset, that we clusterized by k-means algorithm (MacQueen, 1967). Then,

we applied the LDA to the seismic features corresponding to these classes. With an explained variance greater than 95%, we found the transformed and reduced seismic features that better separate the sea ice space distribution classes.

- Time Smoothing (SMT): to reduce the time variability of the seismic data, we applied a sliding window smoothing the signal. The mean autocorrelation function of the RMS amplitude time series shows a fast decay of more than 60% after only 2 days, thus we considered a causal-consistent smoothing window of 3 days.

As potential input features (IF) for machine learning modeling, we tested all the possible ordered selections without repetition of any subset (included the empty one) of the described transformations (PIT, LDA, SMT) on the RMS amplitude data. To be sure that the model predicted the sea ice concentration from the information carried by microseism signals and not from the implicit seasonality, we compared the results obtained by using only a variable linked to the period of the year as input (see rows with "–" in the Basic data column of Table 2.2) with those obtained by the whole set of microseism-related data (see rows with "microseism" in the Basic data column of Table 2.2). The time-related feature was defined as a sinusoidal oscillator between 0 and 1 with annual period and with 0 corresponding to the time with the maximum peak of the sea ice concentration (defined as the measurement of the area of ocean where there is at least 15% of sea ice concentration). All the potential sets of IFs were tested both in presence (see rows with "x" in the Time feature column of Table 2.2) and in absence of this time-related feature that indicated the day of the year (see rows without "x" in the Time feature column of Table 2.2).

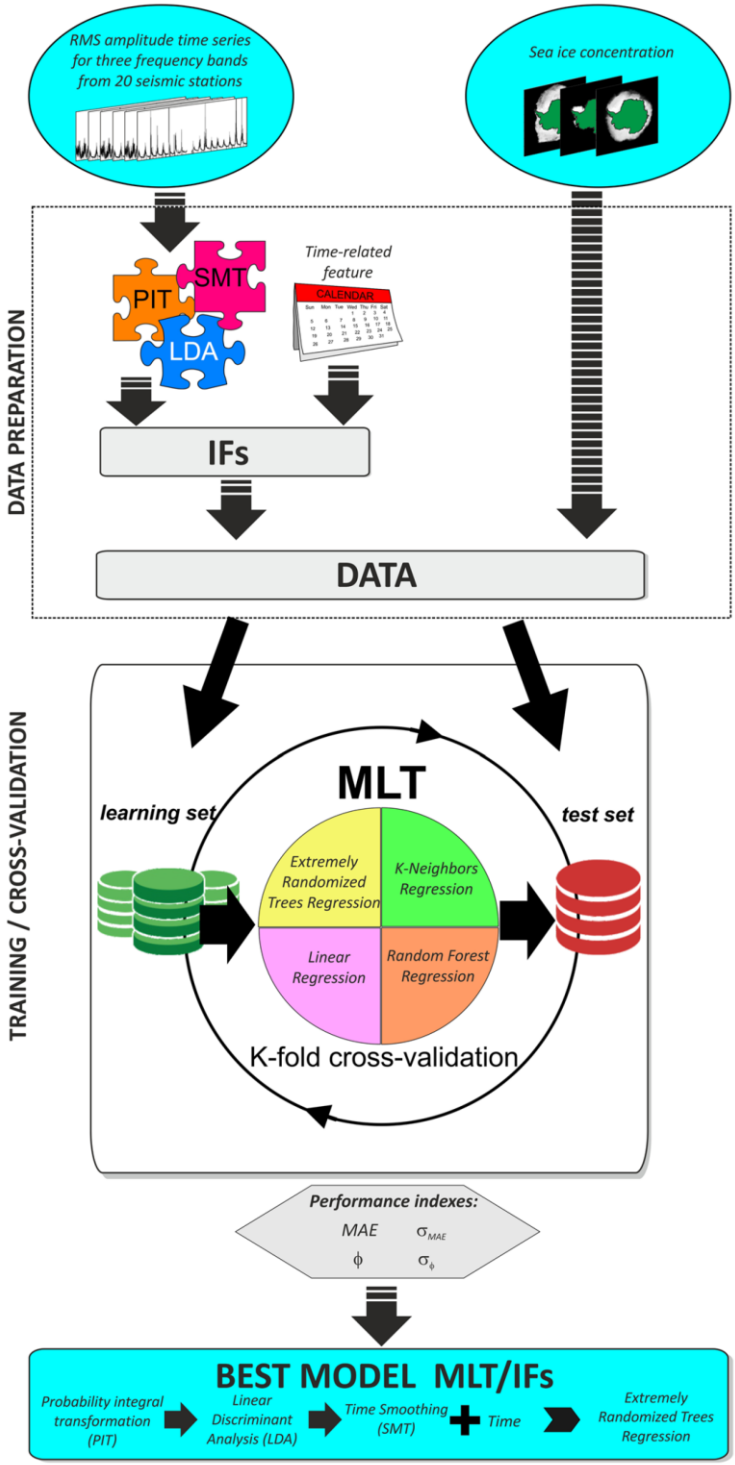

Figure 2.15: Machine learning scheme. Scheme of the modelling analysis to get the spatial distribution of sea ice concentration by using the microseism (see text for details). “IFs” stands for input features, “MLT” for machine learning technique, “PIT” for probability integral transformation, “LDA” for linear discriminant analysis, “SMT” for time smoothing, “MAE” for mean absolute error. The maps were created by Matplotlib package for Python (Hunter, 2007).

Concerning the training step (ii), the supervised machine learning techniques (MLT), taken into account to predict the sea ice concentration, were:

- Linear Regression (Kutner et al., 2004): the relationships are modeled using linear predictor functions, whose unknown model parameters are estimated from the input/output data.
- Random Forest Regression (Liaw & Wiener, 2002): it operates by constructing a multitude of decision trees at training time, outputting the mean prediction of the individual trees.
- K-Neighbors Regression (Altman, 1992): average of the values of its k nearest neighbors.
- Extremely Randomized Trees Regression (Geurts et al., 2006): based on random forest, it applies a fully random selection to split data in the test nodes.

The step (iii), cross-validation, consisted of evaluating the unbiased generalization capacity of each pair MLT/IF by calculating the prediction performance through K fold cross-validation (Kuhn & Johnson, 2013). We considered K = 10 with partially overlapped subsets of 365 timely-consecutive daily data. The choice to consider 365-consecutive daily data for testing was due to the slow dynamics of sea ice formation and melting. In a completely random selection of the test set, results could be optimistically biased for the presence of close-in-time (and hence similar) patterns in the learning set. Considering an annual (i.e. 365 daily data) test set, we avoided biases in results due to such seasonal correlations. The performance indexes used to compare the models were:

- The mean absolute error (MAE) between the observed sea ice concentration and the predicted one, calculated in a mask where the ice concentration is not null for the whole period.
- The percentage (φ) of the cells in the sea ice grid, showing an absolute error, defined as the absolute value of the difference between predicted and true sea ice concentration, lower than 50% (threshold chosen to discriminate gross errors).

Cross-validation allowed us to estimate the unbiased mean and standard deviation (σ) of the performance indexes. Table 2.2 shows the cross-validated performances of the different pairs MLT/IF sorted by ascending MAE + σ_{MAE}.

2.4 RESULTS

2.4.1 RESULTS OF THE SPECTRAL AND RMS AMPLITUDE ANALYSES

Microseism exhibits seasonal variability with maxima during February–April (austral fall) and minima during October–December (austral spring-summer) as shown in the yearly smoothed and stacked RMS amplitude time series (Fig. 2.5). There are slight differences in the seasonal pattern, regarding both the times of maxima and minima (indeed, there are clear lags between the different time series) and the shape of the patterns (Fig. 2.5a). Such a variability to a first approximation depends on the sector where the stations are located (see Fig. 2.1 for sector view). For instance, stations falling in the Weddell sector show RMS amplitude maxima for the band 2.5–5.0 s during the period end of January – end of April, while stations in Ross sector exhibit maxima slightly before, mid-January – mid-April.

In addition, stations located in the West Pacific sector (and in particular CASY) show a shorter period of minimum RMS amplitudes, with respect to the stations located in the Ross sector. Finally, a certain variability can also be observed among stations belonging to the same sector. The annual patterns of ESPZ and ORCD are evidently different from the patterns shown by the other stations of the Weddell sector.

The spectrograms show how most microseism energy is comprised in the bands 2.5–5.0 and 5–10 s (SPSM and SM, respectively) at all the considered stations (Fig. 2.3). The different energy content in the distinct period bands is also evident in the maps, showing the spatial distribution of the median value of RMS amplitudes (Fig. 2.6a–c). Moreover, these maps highlight that the area with maximum microseism amplitude is the West Antarctica, and in particular the Antarctic Peninsula. This feature has also been noted by Anthony et al. (2015), who analysed a shorter time period between 2007 and 2012, and interpreted this feature as due to the circumpolar westerlies, making Drake Passage and Bellingshausen Sea very stormy. Strong microseism sources located in this area have also been reported in other papers (Lee et al., 2011; Prat et al., 2017). Moreover, the maps, displaying the spatial distribution of the median value of RMS amplitudes in the three investigated period bands (2.5–5.0, 5–10 and 13–20 s) during February–April (Fig. 2.6d–f) and October–December (Fig. 2.6g–i), confirm the strong seasonal modulation of the microseism: higher amplitude during austral fall, and lower amplitude during austral spring-summer.

2.4.2 RESULTS OF THE CORRELATION ANALYSIS

The results of the correlation analysis highlight clear anti-correlation patterns for all the bands (in Fig. 2.7a the Spearman correlation coefficient map for period band 2.5-5.0 s is shown). The values of Spearman correlation strongly depend on the considered station and period band. In particular, the period band showing the strongest anti-correlation is 2.5–5.0 s (SPSM; Fig. 2.7a). Indeed, the median value calculated on the Spearman correlation minima obtained for all the stations is equal to −0.63 for 2.5–5.0 s band, while it is equal to −0.43 and −0.40 for 5–10 s and 13–20 s (SM and PM), respectively. Furthermore, the estimated space distribution of the p-values highlights how the anti-correlation obtained for most of the stations is significantly different from zero in wide areas, in some cases coinciding with almost the whole Southern Ocean (Fig. 2.8a,c,e). This is because both the parameters show strong seasonal periodicity.

In addition, to seek a prospective spatial dependence between sea ice concentration and the location of the station recording microseism, the long-period components (among which the seasonal modulation) were filtered out and the correlation analysis was performed again between the filtered time series (Figs 2.7b,d,f and 2.8b,d,f).

The correlation maps show less strong anti-correlations: the median value calculated on the Spearman correlation minima obtained for all the stations is equal to −0.34, −0.28 and −0.29 for 2.5–5.0, 5–10 and 13–20 s, respectively. More importantly, in most cases the Spearman correlation coefficient displays the lowest values in the areas close to the stations where the microseism is recorded. This is evident in many regions, such as the Drake Passage, Bellingshausen Sea and Weddell Sea (i.e. stations ESPZ, HOWD, ORCD, PMSA, SMAI), and the Ross Sea (i.e. stations MPAT, SBA,

VNDA) (Figs 2.1 and 2.6b). According to Anthony et al. (2017), in regions with strong anti-correlation we expect that sea ice interferes with the microseism generation.

In a few cases, it is also possible to note low values of Spearman correlation coefficient in areas far away from the stations recording the microseism. This is especially evident for stations located on or nearby the Antarctica Peninsula, such as ORCD, ESPZ and DNTW (see Figs 2.1 and 2.7b), whose Spearman correlation maps show anti-correlation both close to the stations (Weddell Sea and/or Bellinghausen Sea) and far away (mostly Western Ross Sea). However, such an apparent link between microseism and very distant sea ice is due to the fact that time series of sea ice concentration in areas far away from each other can have similar patterns.

To verify the dependence of the anti-correlation from the distance between the sea ice and the seismic station recording microseism, a cumulative 3D density plot of all the correlation maps was obtained per each period band, showing the distance in the x-axis, the Spearman correlation value in the y-axis, and the number of Spearman correlation estimations, performed on the filtered deseasonalized time series, with the color scale (Figs. 2.9a, 2.10a, 2.11a). These 3D density plots have the aim to highlight common patterns among the 20 maps, and reduce the contribution of features regarding single stations or a small number of stations, such as the afore mentioned apparent link between microseism and very distant sea ice. Furthermore, 2D histograms, gathering the Spearman correlation values within given ranges of distance (from 0 to 6000 km, with step of 125, 250, 500 and 1000 km), were obtained (Figs. 2.9b-g, 2.10b-g, 2.11b-g, 2.12a-c, 2.13a-c, 2.14a-c). Both the 3D density plots and the 2D histograms show a fairly symmetric shape with maxima in

correspondence with zero correlation values, that is what we expect in case of unrelated random signals. However, if we focus on the short distances (<1000 km; Figs. 2.9b, 2.10b, 2.11b, 2.12a-c, 2.13a-c, 2.14a-c), the distributions show a clear asymmetry with higher number of negative correlation values with respect to the positive correlation values. This feature, evident for all the investigated period bands, suggests that microseism is mostly affected by the sea ice concentration within 1000 km from the station recording the seismic signal. Such a decrease in sea ice sensitivity of microseism, due to the increasing distance from the seismic station, has never been observed in Antarctica. In the Arctic, similar observations led (Tsai & McNamara, 2011) to build an equation linking sea ice and microseism amplitudes in Bering Sea.

2.4.3 RESULTS OF THE MICROSEIM-SEA ICE LINK BY MEANS OF MACHINE LEARNING

The observed decrease in sea ice sensitivity of microseism, due to the increasing distance from the seismic station, paves the way to implement an algorithm to spatially and temporally reconstruct the sea ice distribution around Antarctica on the basis of the microseism amplitudes. However, to do that, an analytical approach, based on microseism wave propagation, seems to be impracticable for the few and sparse data available in a highly heterogeneous and complex environment that would conduct to a strongly underdetermined ill-posed inversion problem. For this reason, we exploited the capabilities of the newest regression algorithms in machine learning to reconstruct the sea ice field starting from the knowledge of the microseism features or their transformations. In particular, the method we used is composed of three main

steps (summarized in Fig. 2.15 and explained in detail in the section 2.3.3): (i) data preparation; (ii) training; (iii) cross-validation.

By listing the cross-validated performances of the different pairs MLT/IF sorted by ascending MAE plus its standard deviation (σ_{MAE}) (Table 2.2), we found that the best performance is obtained by the Extremely Randomized Trees Regression applied on RMS amplitude data sequentially post-processed by PIT, LDA and SMT, with the addition of the time-related feature. The analysis of the results also shows that the time-related feature by itself (see rows with "–" in column Basic data in Table 2.2) is not able to give performance comparable with those obtained with microseism features.

Figure 2.16a–f shows two cases of actual and predicted sea ice concentration, together with the corresponding error, obtained by using the Extremely Randomized Trees model trained on the above-mentioned optimal IFs, acquired during 2011–2012 and from the beginning of 2016 to half year before the considered day.

For the identified optimal pair MLT/IF, we also estimated and mapped the unbiased spatial MAE through K-fold cross-validation (Fig. 2.17a). It is evident how the area characterized by the lowest prediction error is the southern part of Weddell Sea, that is shielded from the westerlies by the Antarctica Peninsula (Anthony et al., 2017) and for this reason it is characterized by almost permanent sea ice (see an example of sea ice concentration map during a period with very low sea ice extent in Fig. 2.16d).

In addition, we computed the seasonal trend of the prediction error through K-fold cross-validation. The error for each day was computed as the average of the absolute values of the difference between the true sea ice concentration and the predicted one. Then, the median of all the errors concerning the same day during the year was

computed to get the seasonal trend of prediction error (Fig. 2.17b). It is evident how the error is higher during the time periods characterised by high sea ice concentration.

#	Regression Model	Basic data	Sequence of transformations	Time feature	MAE (%)	σ_{MAE}(%)	φ(%)	σ_{φ} (%)	MAE+σ_{MAE}(%)
1	ExtraTrees	microseism	PIT SMT LDA	x	10.33	3.29	96.00	2.04	13.62
2	ExtraTrees	microseism	PIT LDA SMT	x	10.28	3.34	95.95	2.11	13.62
3	ExtraTrees	microseism	SMT PIT LDA	x	10.39	3.41	95.89	2.22	13.79
4	RandomForest	microseism	PIT LDA SMT	x	10.38	3.46	95.44	2.26	13.84
5	RandomForest	microseism	PIT SMT LDA	x	10.49	3.49	95.35	2.34	13.99
6	RandomForest	microseism	SMT PIT LDA	x	10.47	3.52	95.39	2.48	13.99
7	ExtraTrees	microseism	PIT LDA	x	10.52	3.54	95.65	2.37	14.06
8	ExtraTrees	microseism	PIT SMT	x	10.68	3.43	96.08	2.08	14.11
9	ExtraTrees	microseism	LDA PIT SMT	x	10.64	3.52	95.75	2.05	14.16
10	ExtraTrees	microseism	SMT PIT	x	10.71	3.47	96.10	2.05	14.18
11	ExtraTrees	microseism	LDA SMT PIT	x	10.66	3.55	95.76	2.13	14.21
12	ExtraTrees	microseism	PIT	x	10.79	3.49	95.89	2.23	14.28
13	ExtraTrees	microseism	PIT LDA SMT		10.79	3.51	96.12	2.49	14.30
14	ExtraTrees	microseism	PIT SMT LDA		10.86	3.50	96.06	2.53	14.36
15	ExtraTrees	microseism	SMT LDA PIT	x	10.80	3.59	95.77	2.16	14.39
16	RandomForest	microseism	PIT LDA	x	10.67	3.76	95.27	2.71	14.43
17	RandomForest	microseism	PIT SMT	x	10.81	3.70	95.35	2.55	14.51
18	ExtraTrees	microseism	SMT PIT LDA		10.96	3.56	95.99	2.68	14.52
19	RandomForest	microseism	SMT PIT	x	10.84	3.72	95.46	2.58	14.56
20	ExtraTrees	microseism	SMT	x	10.96	3.62	95.77	2.39	14.58
21	RandomForest	microseism	LDA SMT PIT	x	10.83	3.80	95.24	2.60	14.63
22	RandomForest	microseism	PIT LDA SMT		10.93	3.70	95.45	2.90	14.63
23	RandomForest	microseism	LDA PIT SMT	x	10.84	3.80	95.22	2.64	14.64
24	RandomForest	microseism	PIT	x	10.86	3.80	95.40	2.82	14.66
25	RandomForest	microseism	PIT SMT LDA		11.01	3.69	95.45	2.89	14.70
26	ExtraTrees	microseism	LDA PIT	x	10.96	3.79	95.41	2.58	14.75
27	RandomForest	microseism	SMT LDA PIT	x	10.91	3.86	95.31	2.63	14.77
28	ExtraTrees	microseism		x	11.13	3.65	95.59	2.50	14.78
29	RandomForest	microseism	SMT PIT LDA		11.10	3.78	95.44	3.09	14.87
30	Knn	microseism	PIT LDA SMT	x	10.88	4.00	94.47	3.02	14.88
31	Knn	microseism	PIT SMT LDA	x	10.94	4.07	94.53	3.17	15.00
32	Knn	microseism	SMT PIT LDA	x	10.97	4.10	94.45	3.26	15.07
33	ExtraTrees	microseism	PIT LDA		11.28	3.85	95.79	3.01	15.13
34	Knn	microseism	PIT LDA	x	11.07	4.08	94.82	3.14	15.15
35	RandomForest	microseism	LDA PIT	x	11.17	4.08	95.03	3.13	15.25
36	Knn	microseism	PIT LDA SMT		11.30	4.08	94.50	3.48	15.38
37	ExtraTrees	microseism	LDA PIT SMT		11.43	4.06	95.55	2.84	15.49
38	ExtraTrees	microseism	SMT LDA PIT		11.48	4.04	95.60	2.89	15.52
39	ExtraTrees	microseism	LDA SMT PIT		11.47	4.05	95.56	2.87	15.53
40	RandomForest	microseism	PIT LDA		11.48	4.12	95.27	3.41	15.60
41	Knn	microseism	PIT SMT LDA		11.61	4.20	94.32	3.78	15.81
42	RandomForest	microseism	LDA SMT PIT		11.62	4.19	95.11	3.16	15.81
43	RandomForest	microseism	LDA PIT SMT		11.59	4.23	95.12	3.26	15.82
44	Knn	microseism	SMT PIT LDA		11.62	4.27	94.19	3.87	15.89
45	RandomForest	microseism	SMT LDA PIT		11.60	4.32	94.99	3.47	15.92
46	ExtraTrees	microseism	SMT PIT		11.86	4.15	95.99	3.07	16.01
47	ExtraTrees	microseism	PIT SMT		11.91	4.22	95.93	3.14	16.13
48	ExtraTrees	microseism	SMT		12.13	4.03	96.16	2.84	16.16
49	Knn	microseism	PIT SMT	x	11.91	4.41	93.96	3.95	16.32
50	Knn	microseism	PIT LDA		11.95	4.40	94.48	3.88	16.35
51	Knn	microseism	LDA PIT SMT	x	12.09	4.38	93.87	4.21	16.47
52	ExtraTrees	microseism	LDA PIT		12.06	4.45	95.31	3.48	16.51
53	RandomForest	microseism	LDA PIT		12.10	4.53	95.00	3.72	16.63
54	ExtraTrees	microseism	PIT		12.35	4.38	95.82	3.50	16.73
55	Knn	microseism	SMT PIT	x	12.16	4.58	93.65	4.26	16.74
56	RandomForest	microseism	SMT PIT		12.17	4.69	95.00	4.38	16.85
57	RandomForest	microseism	PIT SMT		12.19	4.68	95.03	4.26	16.87
58	Knn	microseism	PIT	x	12.40	4.52	94.13	4.07	16.92
59	Knn	microseism	LDA SMT		12.19	4.84	93.54	4.53	17.03
60	ExtraTrees	microseism			12.69	4.35	95.91	3.47	17.03
61	RandomForest	microseism	PIT		12.51	4.72	95.30	4.31	17.24
62	Knn	microseism	SMT LDA		12.39	4.98	93.68	4.71	17.37
63	Linear	microseism	LDA PIT SMT	x	14.23	3.16	96.58	2.19	17.39
64	Linear	microseism	LDA SMT PIT	x	14.35	3.12	96.58	2.27	17.47
65	Knn	microseism	PIT SMT		12.56	4.98	93.44	4.85	17.54
66	Knn	microseism	LDA SMT PIT	x	13.03	4.55	93.16	4.92	17.58
67	Linear	microseism	PIT SMT LDA	x	14.27	3.32	96.74	2.36	17.58
68	Linear	microseism	PIT LDA SMT	x	14.30	3.33	96.72	2.45	17.64
69	Linear	microseism	SMT LDA PIT	x	14.47	3.21	96.35	2.50	17.68
70	Linear	microseism	SMT PIT LDA	x	14.35	3.36	96.75	2.40	17.71
71	Linear	microseism	LDA PIT SMT		14.54	3.22	96.62	2.33	17.76
72	Linear	microseism	PIT LDA	x	14.40	3.39	96.69	2.47	17.79
73	Knn	microseism	LDA PIT	x	13.19	4.64	93.75	4.95	17.82
74	Linear	microseism	LDA SMT PIT		14.68	3.19	96.61	2.32	17.86
75	Knn	microseism	LDA PIT SMT		13.03	4.86	93.27	5.05	17.88
76	Linear	microseism	PIT SMT LDA		14.60	3.30	96.68	2.37	17.90
77	Linear	microseism	PIT LDA SMT		14.63	3.31	96.67	2.46	17.94
78	Linear	microseism	SMT LDA PIT		14.72	3.23	96.32	2.54	17.94
79	Linear	microseism	SMT PIT LDA		14.66	3.30	96.71	2.40	17.96
80	Knn	microseism	SMT PIT		12.79	5.18	93.09	5.24	17.97
81	Knn	microseism	LDA		12.85	5.13	93.77	4.84	17.98
82	Linear	microseism	LDA PIT	x	14.56	3.48	96.43	2.61	18.04
83	Knn	microseism	SMT LDA PIT	x	13.23	4.89	93.12	5.50	18.12
84	Knn	microseism	PIT		13.11	5.17	93.63	5.15	18.27
85	Linear	microseism	PIT LDA		14.84	3.50	96.58	2.70	18.35
86	Linear	microseism	PIT	x	15.09	3.54	96.25	2.89	18.63
87	Linear	microseism	SMT LDA	x	15.10	3.55	95.69	3.24	18.64
88	RandomForest	microseism		x	13.36	5.36	93.39	5.88	18.73
89	Linear	microseism	LDA PIT		15.03	3.70	96.34	2.95	18.73
90	Linear	microseism	LDA SMT	x	15.14	3.61	95.74	3.26	18.75
91	RandomForest	microseism	SMT	x	13.25	5.52	93.05	6.02	18.77
92	Linear	microseism	PIT SMT	x	15.37	3.57	96.01	3.12	18.94
93	Linear	microseism	SMT PIT	x	15.47	3.69	95.95	3.29	19.16
94	Linear	microseism	PIT		15.54	3.66	96.07	3.11	19.20
95	Linear	microseism	PIT SMT		15.73	3.55	95.89	3.14	19.28
96	Linear	microseism	LDA	x	15.44	3.85	95.65	3.35	19.29
97	Linear	microseism	SMT LDA		15.86	3.65	95.72	3.52	19.51
98	Linear	microseism	SMT PIT		15.82	3.71	95.82	3.36	19.53
99	Knn	microseism	SMT LDA PIT		14.38	5.46	92.30	6.65	19.83
100	Linear	microseism	LDA SMT		16.06	3.78	95.68	3.77	19.84
101	Knn	microseism	LDA SMT PIT		14.40	5.61	92.06	6.70	20.02
102	Knn	microseism	LDA PIT		14.85	5.43	92.60	6.40	20.28
103	Linear	microseism	LDA		16.61	4.19	95.42	4.34	20.79
104	Knn	microseism	SMT		15.16	6.46	91.37	7.59	21.62
105	Linear	microseism		x	16.57	5.29	94.66	5.19	21.86
106	ExtraTrees	microseism	LDA	x	15.03	6.86	88.37	7.85	21.90
107	ExtraTrees	microseism		x	15.43	7.55	87.78	8.46	22.98
108	ExtraTrees	-		x	15.43	7.55	87.78	8.46	22.98
109	ExtraTrees	microseism	SMT LDA	x	15.44	7.55	87.77	8.47	22.99
110	Knn	microseism			16.16	6.84	90.95	8.38	22.99
111	ExtraTrees	microseism	LDA SMT	x	15.44	7.56	87.77	8.48	22.99
112	Knn	microseism	LDA SMT	x	16.88	6.63	91.06	8.20	23.51
113	Knn	microseism		x	16.88	6.63	91.06	8.19	23.51
114	Knn	microseism		x	16.88	6.63	91.06	8.20	23.51
115	Knn	-		x	16.88	6.63	91.06	8.20	23.51
116	Knn	microseism	SMT	x	16.88	6.63	91.06	8.20	23.51
117	Knn	microseism	LDA	x	16.88	6.63	91.06	8.20	23.51
118	Knn	microseism	SMT LDA	x	16.88	6.63	91.06	8.20	23.51
119	Linear	microseism	SMT	x	17.46	6.12	93.77	6.26	23.58
120	Linear	microseism		x	17.73	6.06	92.15	7.01	23.79
121	Linear	-		x	17.73	6.06	92.15	7.01	23.79
122	Linear	microseism			17.75	6.27	94.06	6.66	24.02
123	RandomForest	microseism	LDA SMT	x	17.31	6.84	89.11	8.35	24.15
124	RandomForest	microseism	SMT LDA	x	17.30	6.86	89.10	8.39	24.17
125	RandomForest	microseism	LDA	x	17.33	6.85	89.10	8.38	24.18
126	RandomForest	microseism		x	17.32	6.87	89.10	8.37	24.19
127	RandomForest	-		x	17.32	6.87	89.10	8.37	24.19
128	Linear	microseism	SMT		18.37	6.91	93.24	7.44	25.29
129	RandomForest	microseism			20.64	5.93	91.45	6.88	26.57
130	RandomForest	microseism	SMT		22.12	5.43	90.60	6.67	27.55
131	ExtraTrees	microseism	LDA		24.23	4.36	89.05	6.79	28.58
132	ExtraTrees	microseism	LDA SMT		24.72	4.41	87.79	7.54	29.13
133	ExtraTrees	microseism	SMT LDA		24.72	4.41	87.79	7.54	29.13
134	RandomForest	microseism	LDA SMT		24.72	4.41	87.80	7.54	29.14
135	RandomForest	microseism	LDA		24.73	4.41	87.79	7.54	29.14
136	RandomForest	microseism	SMT LDA		24.72	4.42	87.79	7.55	29.14

Table 2.2: List of the cross-validated performances of the different pairs MLT/IF sorted by ascending MAE plus its standard deviation (σ_{MAE}). "ExtraTrees" stands for Extremely Randomized Trees Regression, "RandomForest" for Random Forest Regression, "Knn" for K-Neighbors Regression, and "Linear" for Linear Regression. The "x" in the Time feature column means that the time-related feature was included in the training and cross-validation steps. The "-" in the Basic data means that only the time-related feature was included in the training and cross-validation steps (the microseism data were not used).

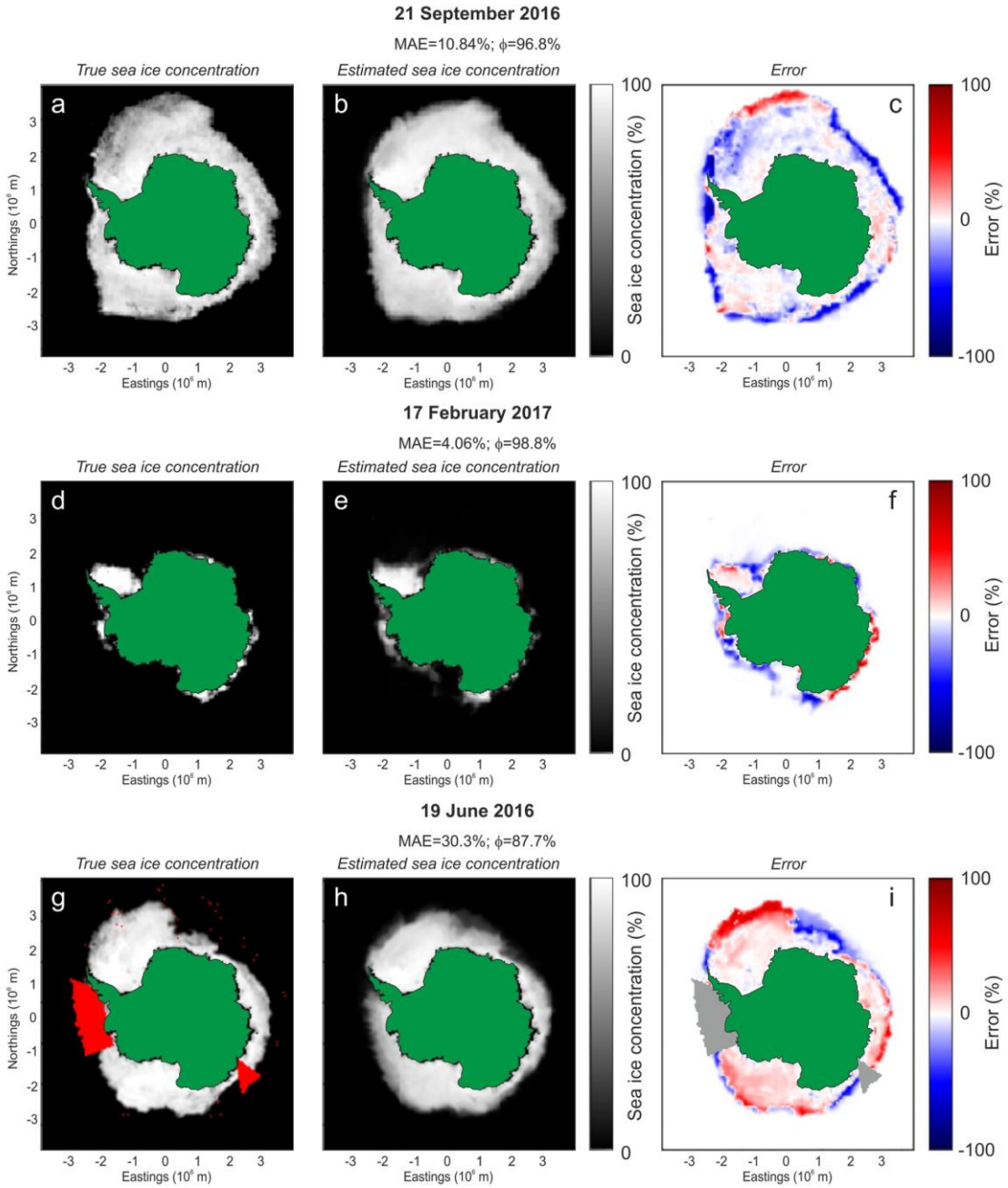

Figure 2.16: Sea ice concentration prediction. Examples of prediction based on microseism features for different patterns of sea ice concentration. In (c,f,i) the spatial distribution of the prediction error (computed as the difference between the true sea ice concentration and the predicted one) is plotted. The used model is obtained by Extremely Randomized Trees techniques applied on RMS amplitude data sequentially post-processed by PIT, LDA and SMT, with the addition of the time-related feature. The red areas in (g) and the grey areas in (i) represent two extended coastal regions, where the sea ice satellite data were missing on 19 June 2016. The maps were created by Matplotlib package for Python (Hunter, 2007).

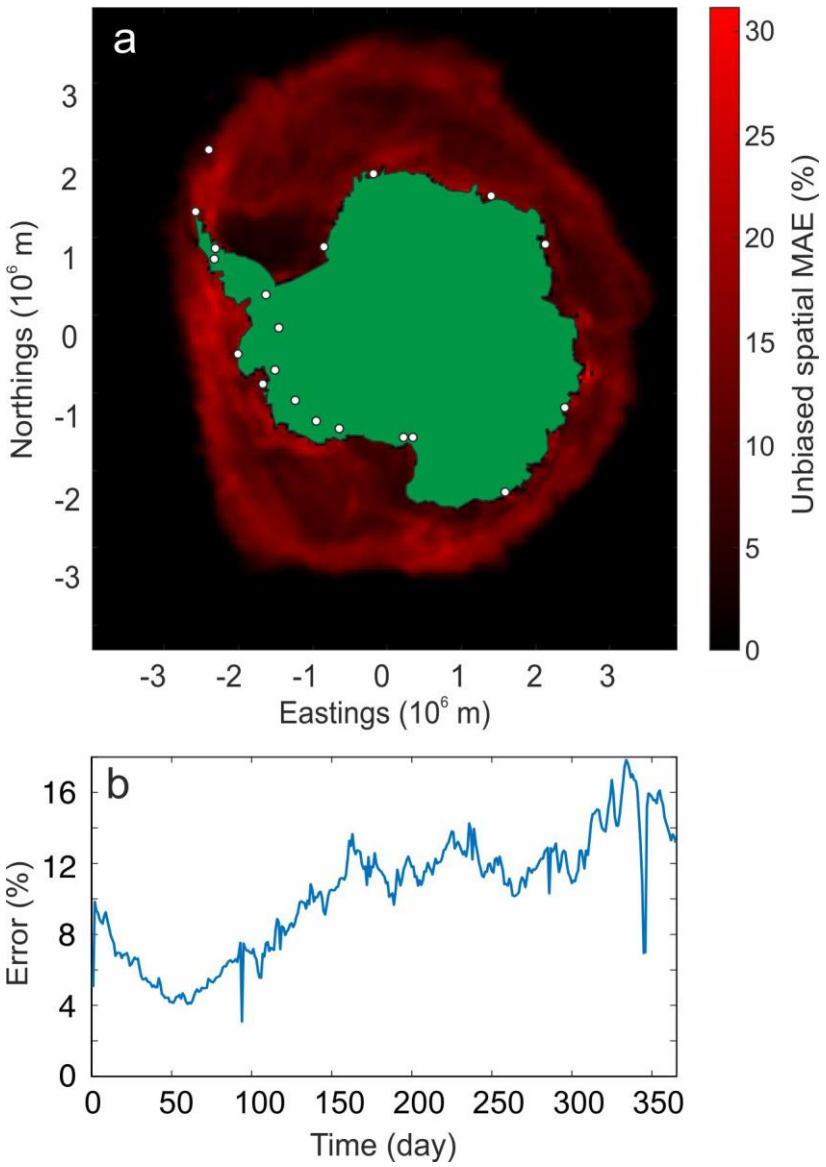

Figure 2.17: Map showing the unbiased spatial mean absolute error (MAE) obtained through K-fold cross-validation. The white dots indicate the locations of the seismic stations. (b) Seasonal trend of the prediction error computed through K-fold cross-validation from the 1st January to the 31st December (see text for explanation about error computation). The map in (a) was created by Matplotlib package for Python (Hunter, 2007).

The ability of the best identified model to predict the sea ice concentration could be particularly useful when the satellite data are partial and present large uncovered areas. As an example, we considered the microseism recorded on 19 June 2016, when satellite data of two extended coastal areas were missing (red areas in Fig. 2.16g). We

trained the model with microseism data collected during 2011–2012 and predicted the sea ice concentration of that day. Figure 2.16h shows the predicted ice field also in the areas without satellite data coverage.

The Extremely Randomized Trees approach has the advantage to easily supply an index of input importance (Fig. 2.18). Even if the time-related feature shows a fairly high importance score (~0.28), due to the seasonality of sea ice concentration, the sum of the importance of the other microseism-related features is much higher (~0.72).

This demonstrates that microseism data carry much more information about sea ice concentration than the simple seasonality. Concerning microseism, the PM data and the SPSM data roughly share the same importance for reconstructing the sea ice field, while the SM shows a slightly smaller contribution to the same purpose (Fig. 2.18a,b). This observation strongly supports the mostly near-coastal origin of the microseism in primary and short-period secondary bands, as evidenced by other authors (Cessaro, 1994; Chen et al., 2011; Grob et al., 2011). Conversely, SM shows the weakest link with sea ice, testifying that its source is likely also influenced by wave–wave interaction in deep ocean, as supposed by previous authors (Kedar et al., 2008; Obrebski et al., 2012). Moreover, the importance of each of the 18 stations, used in this analysis, was evaluated (Fig. 2.18c).

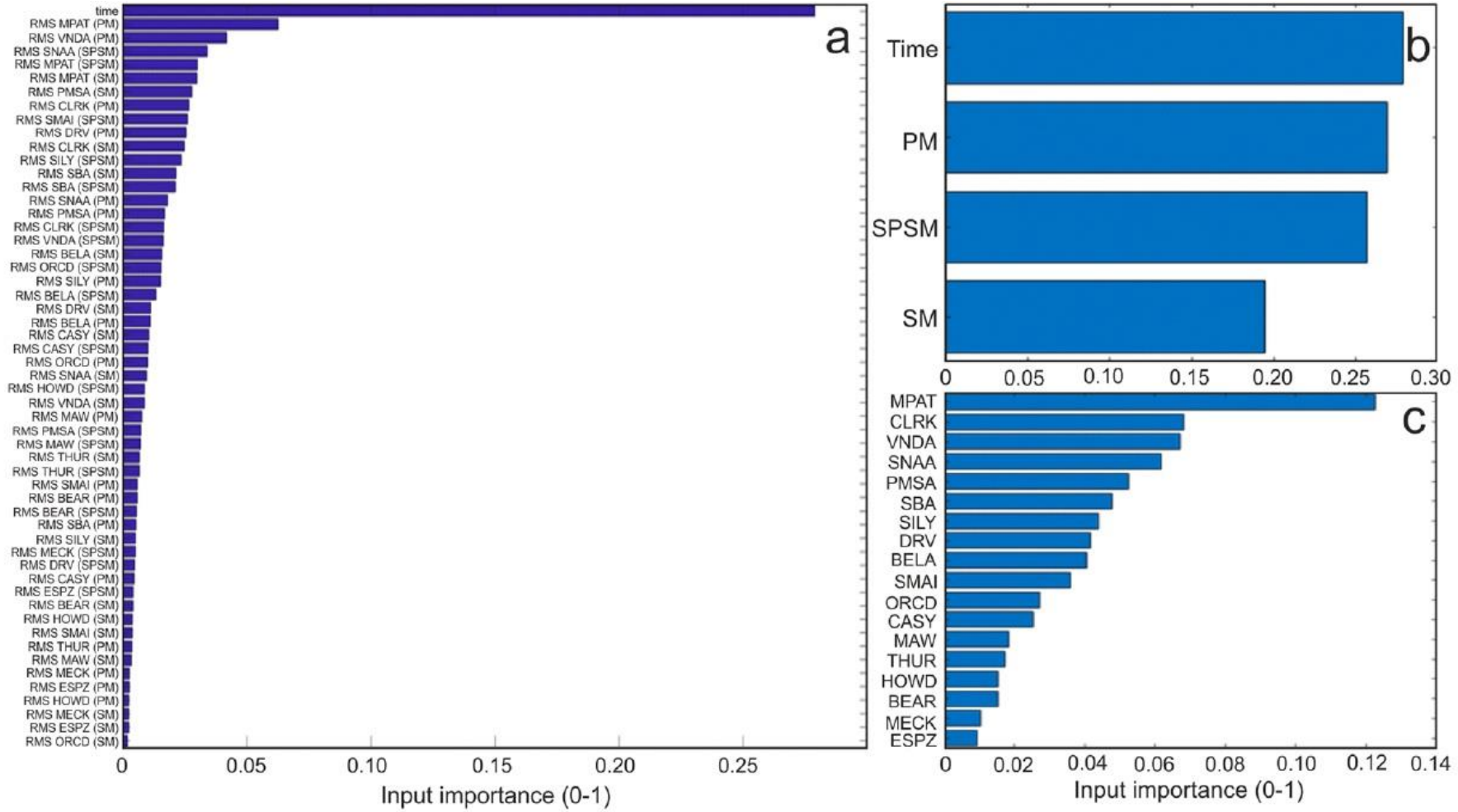

Figure 2.18: Input importance histograms. (a) Index of importance for all the input taken into account. Aggregation through summation of the input importance allowing to rank the microseism bands (b) and the seismic station (c) for sea ice concentration reconstruction purposes. “PM”, “SM”, “SPSM” and “time” in the y-label of (a,b) indicate Primary Microseism, Secondary Microseism, Short Period Secondary Microseism and the time-related feature, respectively.

One of the factors that seem to mostly affect the station importance is the temporal variability of sea ice extent in the coast nearby the station. In particular, we estimated the linear extent of the sea ice, with direction approximately normal to the coastline, closest to the station during two time periods, September 2014 and March 2017, characterized by the maximum and minimum sea ice extent during the investigated time intervals, respectively (Fig. 2.19a–c). As shown in Fig. 2.19d, there is a positive correlation between the station importance and the difference of the two linear sea ice extents.

It has to be underlined that, since microseism can have distant sources unrelated to Antarctic nearshore dynamics (hence theoretically not affected by the sea ice presence, Koch et al., 2011; Gualtieri et al., 2019), bias in the microseism-retrieved space-time distribution of sea ice can occur in case of strong distant microseism sources linked for instance to distant storms. A further issue, highlighted also by Tsai & McNamara (2011), could be related to the fact that sea ice concentration is assumed to be directly linked to the sea ice strength and then to the decrease in the efficiency of energy transfer from ocean to solid earth. Such an assumption could be not entirely verified. In spite of these issues, the prediction capability of the proposed model is satisfying as testified by the MAE value equal to 10.3%, obtained by the optimal pair MLT/IF.

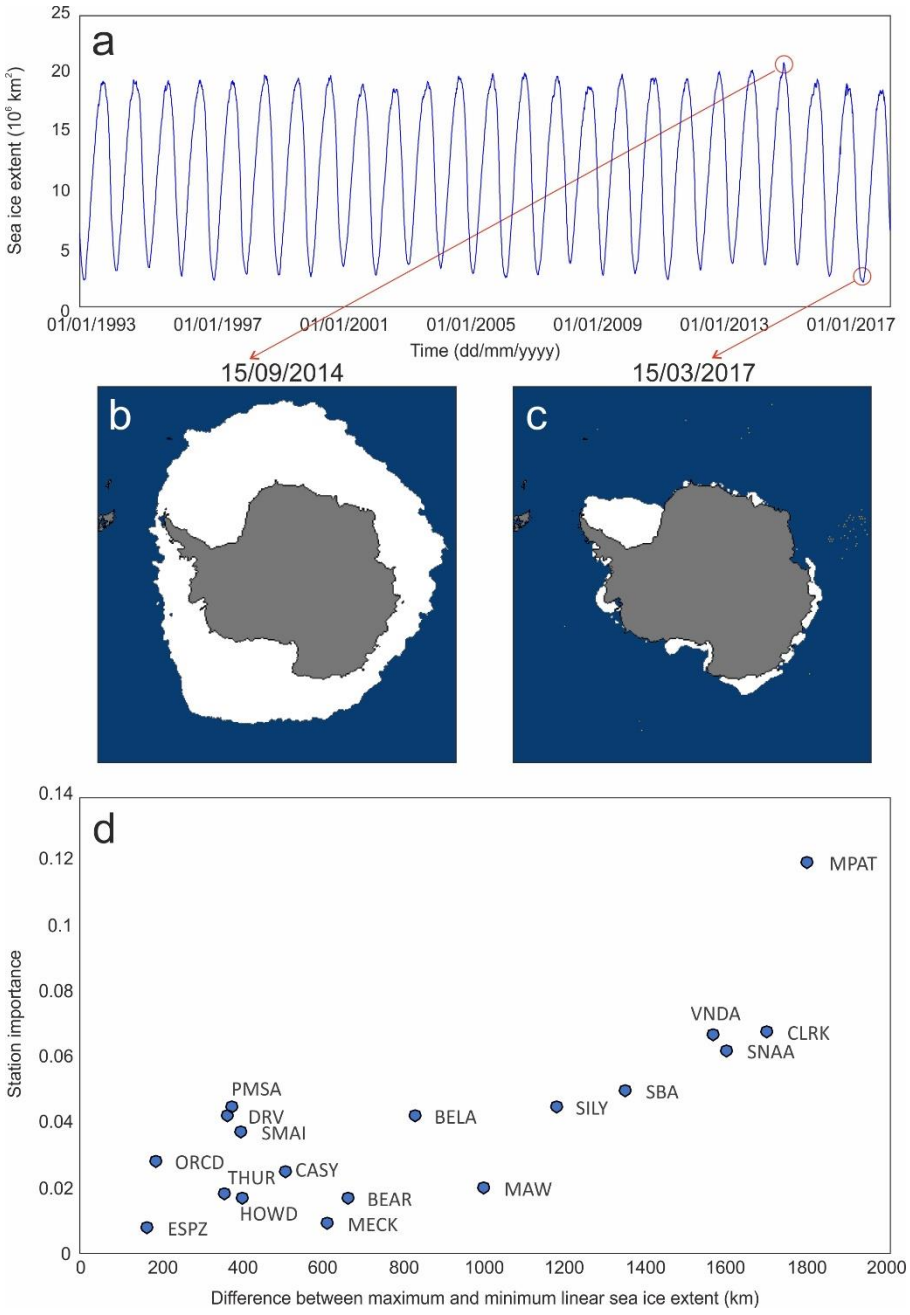

Figure 2.19: Temporal variation of sea ice extent during 1993-2017 (a), computed by counting the daily number of grid cells with sea ice concentration higher than 15% and multiplying it by the cell area. Map of sea ice extent during 15 September 2014 and 15 March 2017, respectively (b,c). Station importance, derived by the MLT, plotted versus the difference between maximum and minimum linear extent of the sea ice measured approximately normal to the coastline closest to the station during September 2014 and March 2017, respectively (d).

2.5 DISCUSSION AND CONCLUSIONS

In summary, we quantitatively investigated the relationship between microseism recorded along the coasts of Antarctica and the sea ice concentration in the whole Southern Ocean.

Clear anti-correlation patterns between microseism and sea ice concentration were found at all the investigated microseism bands (Fig. 2.7). Such a relationship depends on the fact that microseism amplitudes are influenced by ocean wave heights (Ardhuin et al., 2012; Bromirski et al., 1999, 2002). Indeed, if we assume that sea ice concentration is a proxy of sea ice strength (Tsai & McNamara, 2011), the increase in sea ice concentration: (i) prevents swell from reaching the coast, decreasing the efficiency of PM generation and (ii) inhibits the swell reflection along the coast, reducing the SM energy (Stutzmann et al., 2009; Grob et al., 2011). In particular, as for the SPSM, such a band is likely to be generated by local nearshore wave-wave interaction (Bromirski et al., 2005), heavily modulated by the presence of sea ice.

The microseism bands, showing the strongest link with sea ice, are PM and SPSM (Fig. 2.18b), corroborating their mostly near-coastal origin. On the other hand, the weakest link identified between sea ice concentration and SM (Fig. 2.18b) is indicative of influences by wave–wave interaction in deep ocean on the SM source.

In addition, we clearly show a decrease in sea ice sensitivity of microseism, due to the increasing distance from the station recording the seismic signal. The influence seems to disappear for distances above 1,000 km (Figs. 2.9 - 2.14). Following the reasoning of Tsai & McNamara (2011), such a distance could be related to the attenuation length. Indeed, considering the entire period band of microseism (from 20 s to 2.5 s), attenuation coefficient of ~10^{-3} km^{-1} can be found in literature

(Mitchell, 1995; Prieto et al., 2009). Hence, the corresponding attenuation lengths turn out to have the same order of magnitude as the estimated maximum distance of influence of sea ice on microseism. Moreover, this distance could also be linked to the temporal variability of sea ice extent in the coast nearby the station. Indeed, the average value of the difference between maximum and minimum linear extent of the sea ice, measured approximately normal to the coastline closest to the station during September 2014 and March 2017 (characterized by the maximum and minimum sea ice extent during the investigated time intervals, respectively; Fig. 2.19b,c), is equal to 850 km (average of the x-axis values of Fig. 2.19d), and then very similar to the estimated maximum distance of influence of sea ice on microseism. Furthermore, this 1,000 km-distance does not seem to be related to the bathymetry. Indeed, focusing on PM at period of 20 s and following the Airy linear wave theory approximation, most PM generation should occur at water depths less than ~150 m (Bromirski et al., 2002). As for the SPSM (the other microseism type showing the strongest link with sea ice), its sources are mostly located at even shallower depths (e.g. Chen et al., 2011; Juretzek & Hadziioannou, 2016a). If we consider the 150 m depth, this limit corresponds to distances from the Antarctic coastline much shorter than 1,000 km (Arndt et al., 2013).

It is worth noting that such a 1,000 km-threshold, that has to be considered as an average value among the different stations, does not signify that microseism recorded in Antarctica cannot have also distant sources, as highlighted by several authors (e.g. Anthony et al., 2017; Pratt et al., 2017; Koch et al, 2011; Gualtieri et al., 2019). This study does not constrain the locations of the microseism sources, their amplitudes and the corresponding decay with distance, but rather suggests that sea ice with maximum

distance of 1,000 km, on average, contributes to modulate the microseism amplitude. Indeed, location of microseism sources is a hard task, as microseism signals are non-impulsive, and the sources are generally diffuse and variable in time (Anthony et al., 2017). Hence, the classical location algorithms, used in earthquake seismology and based on the picking of the different seismic phases, cannot be applied to locate microseism sources. Array processing techniques, that can overcome the above-mentioned difficulties, have provided locations of microseism source areas surrounding Antarctic (Pratt et al., 2017). However, as the array data available in Antarctica are sparse, the microseism array locations have been obtained only for short time intervals (for instance a couple of months in Pratt et al., 2017). Preliminary information about the direction of the microseism sources, with less accuracy than by using array data, can also come from polarization analysis of single 3-component station (Lee et al., 2011).

The microseism sensitivity in sea ice is reflected in the slightly different annual patterns of RMS amplitude time series observed among the stations (Fig. 2.5). Such differences can be partly interpreted as due to the different sector, where the stations are located. Indeed, as stated by Zwally et al. (2002) there are regional changes in the annual cycle of sea ice extent in the five Antarctic sectors. For instance, the least sea ice cover observed in the West Pacific sector, compared to the other sectors, justifies the shorter duration of the time interval characterised by minimum RMS amplitudes (particularly evident in CASY). In addition, it is also possible to observe peculiar patterns of specific stations, such as ESPZ and ORCD. For these two stations, the duration of the time interval characterised by microseism RMS amplitude minima is

shorter compared to the others, likely reflecting the shorter-lived effect of sea ice modulation on microseism at the relatively lower latitudes of ESPZ and ORCD.

Finally, we propose an innovative method, based on up-to-date machine learning techniques, able to reconstruct the spatial-temporal distribution of sea ice concentration by using microseism recorded in different period bands by distinct seismic stations. The importance of each station in the prediction of sea ice concentration was evaluated (Fig. 2.18c) and turned out to be mostly affected by the temporal variability of sea ice extent in the coasts nearby the station (Fig. 2.19). Hence, the wider the area, close to the station and characterised by intense sea ice time variability, the stronger the modulation effect on the microseism amplitude recorded by the station, and then the higher the station importance for sea ice concentration prediction.

2.6 SUGGESTED FUTURE WORK

The quality of the modelling results, obtained by the machine learning techniques for the relatively small dimension of measured data, indicates that microseism signal carries significant information about the surrounding sea ice concentration. This technique will allow reconstructing the sea ice concentration in both Arctic and Antarctica in periods when the satellite images, routinely used for sea ice monitoring (Fetterer et al., 2017), are not available, using the longest-lived seismic stations, with wide applications in many fields, first of all the climate studies.

A future development of this study will be the inclusion of time series of horizontal seismic component amplitudes as input in the machine learning modelling. While the vertical component mainly brings information about Rayleigh waves composing

microseism, the horizontal components allow taking into account both Rayleigh and Love waves. Indeed, it has been highlighted by recent papers that the contribution of Love waves in microseism can be significant (Juretzek & Hadziioannou, 2016a, 2016b).

In the light of this, also the wavefield features (quantified by the polarization parameters, i.e. incidence angle, azimuth angle, rectilinearity, planarity; Jurkevics, 1988) could add additional information to the machine learning modelling, if included as further inputs.

Moreover, as pointed out by Ferretti et al. (2018), the teleseisms (i.e., earthquakes localized at a distance greater than 1000 km from the recording stations) with frequency content that is similar to that of microseism recordings could affect the results of the predicted sea ice concentration. Therefore, a future development of this study will be the removal of the time windows that include teleseism signals.

The inclusion of these new inputs will likely improve the capability to reconstruct the spatial-temporal distribution of sea ice concentration around Antarctica by the microseism.

3. MICROSEISM AND SEA WAVE HEIGHT RELATIONSHIP IN EASTERN SICILY USING BUOY DATA

3.1 INTRODUCTION

The link between microseism amplitudes and the ocean wave height has been empirically explored by several authors (e.g. Bromirski et al., 1999; Bromirski & Duennebier, 2002; Ardhuin et al., 2012; Ferretti et al., 2013, 2018). For instance, Bromirski et al. (1999) determined site-specific seismic-to-wave transfer functions in the San Francisco Bay area (California). Ferretti et al. (2013, 2018) found empirical relations to predict the significant wave height along the Ligurian coast (Italy). In addition, other authors have derived physics-based models of the generation of the different kinds of microseism from the sea state (e.g. Gualtieri et al., 2013; Ardhuin et al., 2015; Gualtieri et al., 2019).

In spite of the availability of seismic and buoy data in the Ionian and Tyrrhenian Seas and coastlines, the link between sea waves and microseism has never been explored in such areas. Furthermore, although the spectral features of the microseism recorded in this area have been studied (e.g. De Caro et al., 2014), the locations of its sources have never been constrained.

In the following sections we will study the microseism recorded along the coastline of Eastern Sicily in terms of spectral content, amplitude seasonal pattern, and source location. In addition, we will present a novel algorithm, based on up-to-date MLTs, able to reconstruct significant wave height time series in points located in both the Ionian and the Tyrrhenian Seas from the microseism recordings.

3.2 DATA

3.2.1 SEISMIC DATA

In order to investigate microseism, seismic signals recorded from 2010 to 2014 by the vertical component of six stations, belonging to the seismic permanent network run by Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo – Sezione di Catania (INGV-OE), were used (Figure 3.1a, Table 3.1). These stations are equipped with broadband three-component Trillium 40-s seismometers (Nanometrics^{TM}) recording at a sampling rate of 100 Hz.

Moreover, to carry out array analysis, seismic signals recorded in January 2010–February 2012 by the vertical component of the seven stations (equipped with the same sensors as above), composing the summit ring of the Mt. Etna permanent seismic network, were used (Figure 3.1b, Table 3.2). These stations were chosen because of: (i) the availability of continuously recorded data during the time interval 2010–beginning of 2012 (in February– March 2012, EBEL and ETFI stations were destroyed by lava flows); (ii) the ring-shaped geometry; and (iii) the distance from the coastline (and then from the prospective closest microseism sources associated with the nearshore wave–coast or wave– wave interaction).

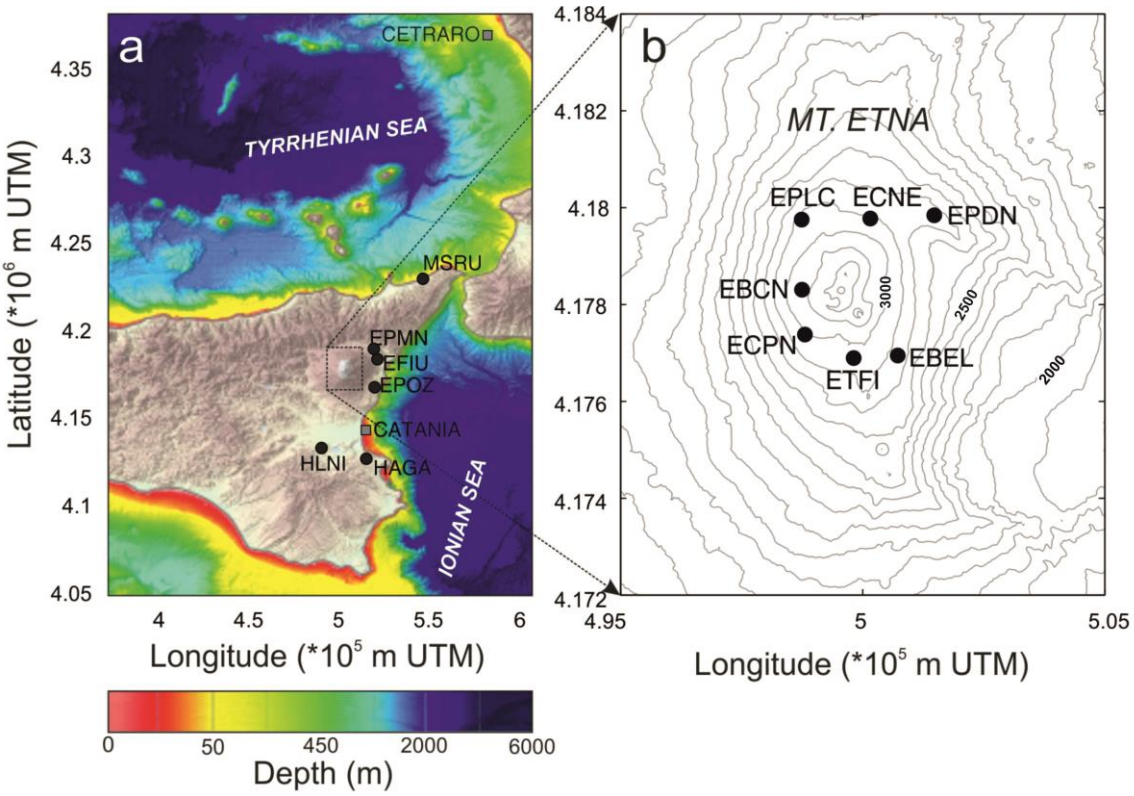

Figure 3.1: (a) Bathymetric and topographic map (EMODnet Bathymetry Consortium, 2018), with the locations of the seismic stations (black dots), used to perform spectral and amplitude analysis of the microseism and to investigate its relationship with significant wave height, recorded by Catania and Cetraro buoy stations (gray squares). (b) Digital elevation model of Mt. Etna, with the locations of the seismic stations (black dots), used to perform array analysis.

Station Name	Latitude	Longitude	Altitude (m a.s.l.)
MSRU	38.2638	15.5083	401
EFIU	37.7897	15.2101	98
EPMN	37.820	15.177	541
EPOZ	37.6718	15.1885	119
HAGA	37.2853	15.1552	126
HLNI	37.3485	13.872	147

Table 3.1: Coordinates of the seismic stations used for microseism and machine learning analyses.

Station Name	Latitude	Longitude	Altitude (m a.s.l.)
EBCN	37.752	13.9858	3085
EBEL	37.7408	15.0091	2808
ECPN	37.7428	13.9885	2985
ECNE	37.7653	15.0018	2901
EPDN	37.7659	15.0168	2823
EPLC	37.7651	13.9857	2917
ETFI	37.7393	13.9979	2948

Table 3.2: Coordinates of the seismic stations used for array analysis.

3.2.2 BUOY DATA

Significant wave height data, recorded from 2010 to 2014 with a 30-min sampling step by two stations (Catania and Cetraro; see Figure 3.1a) belonging to the Italian Data Buoy Network, managed by Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), were used (Bencivenga et al., 2012; Figure 3.2).

The significant wave height is defined as:

$$H_s = 4\sqrt{M_0} \qquad (3.1)$$

where M_0 is the 0-moment of the auto-spectral correlation of the Fourier transformations of the buoy displacements in the frequency/time domain (Steele and Mettlach, 1993):

$$M_0 = \sum_{f_l}^{f_u} \big(S(f)d(f)\big) \qquad (3.2)$$

where the sum of the spectral density $S(f)$ is over all frequency bands, from the lowest frequency f_l to the highest frequency f_u of the non-directional wave spectrum (calculated only for the elevation of the sea surface), and $d(f)$ is the bandwidth of each band.

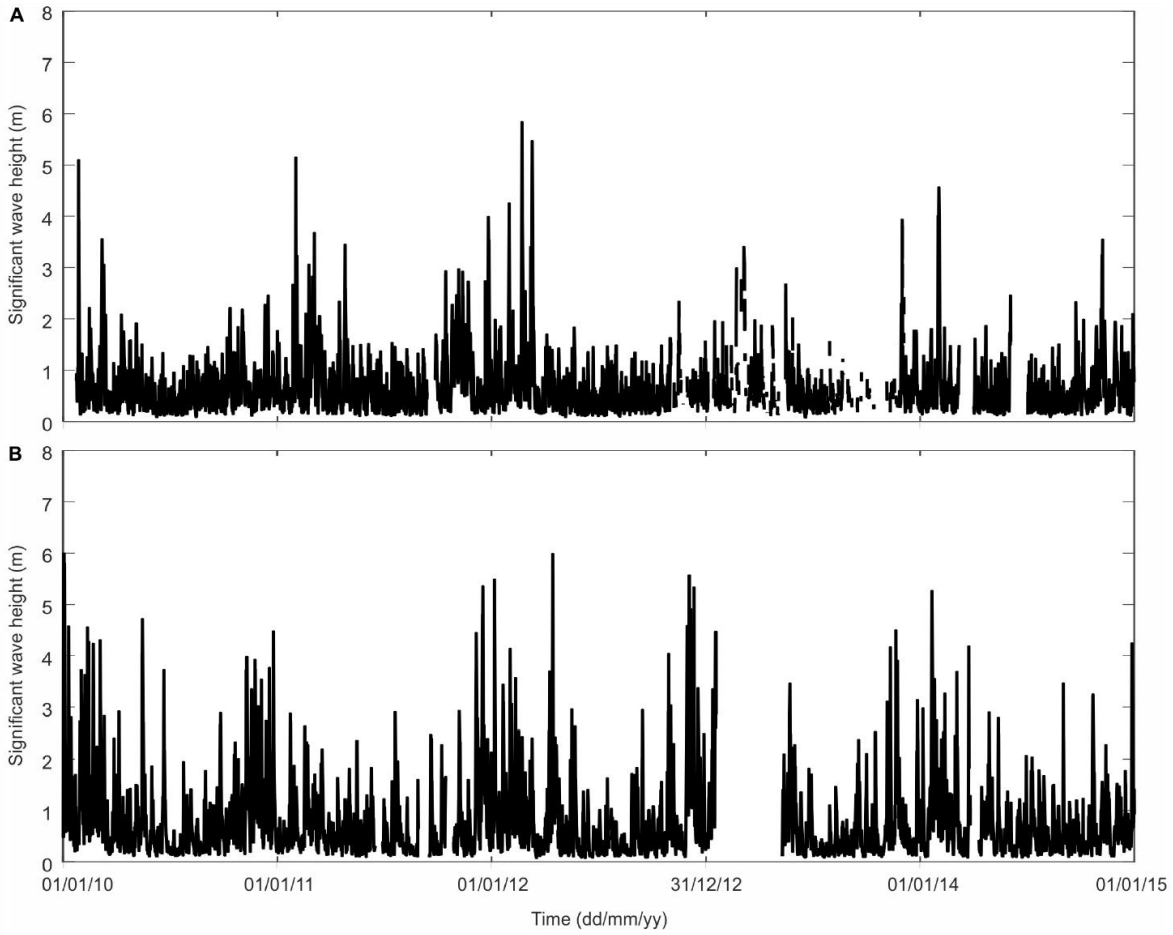

Figure 3.2: Significant wave height time series recorded by the Catania (A) and Cetraro (B) buoys.

3.3 METHODS OF ANALYSIS

3.3.1 SPECTRAL AND AMPLITUDE ANALYSES

The spectral content of the seismic data recorded by the vertical component of the six seismic stations shown in Figure 3.1a, was analysed following the same workflow adopted in paragraph 2.3.1. All the daily spectra were collected and visualized as spectrograms, which are 3D plots with time on the x-axis, frequency on the y-axis, and power spectral density (PSD) indicated by a color scale (Fig. 3.3a). In addition, to obtain information on the spectral features of the seismic signals recorded by the

different stations during the whole investigated period, all the daily spectra composing the spectrograms were averaged (Fig. 3.3b).

The time variability of the microseism amplitude was investigated by calculating the root-mean-square (RMS) amplitude of the seismic signal filtered in three period bands (PM, 13–20 s; SM, 5–10 s; and SPSM, 2.5–5.0 s) using the equation 2.1, computed with both daily and hourly rates. The daily RMS amplitude time series (Figure 3.4) were smoothed by a 90-day-long moving median, split in year-long windows, stacked, and rescaled between 0 and 1 (Figure 3.5).

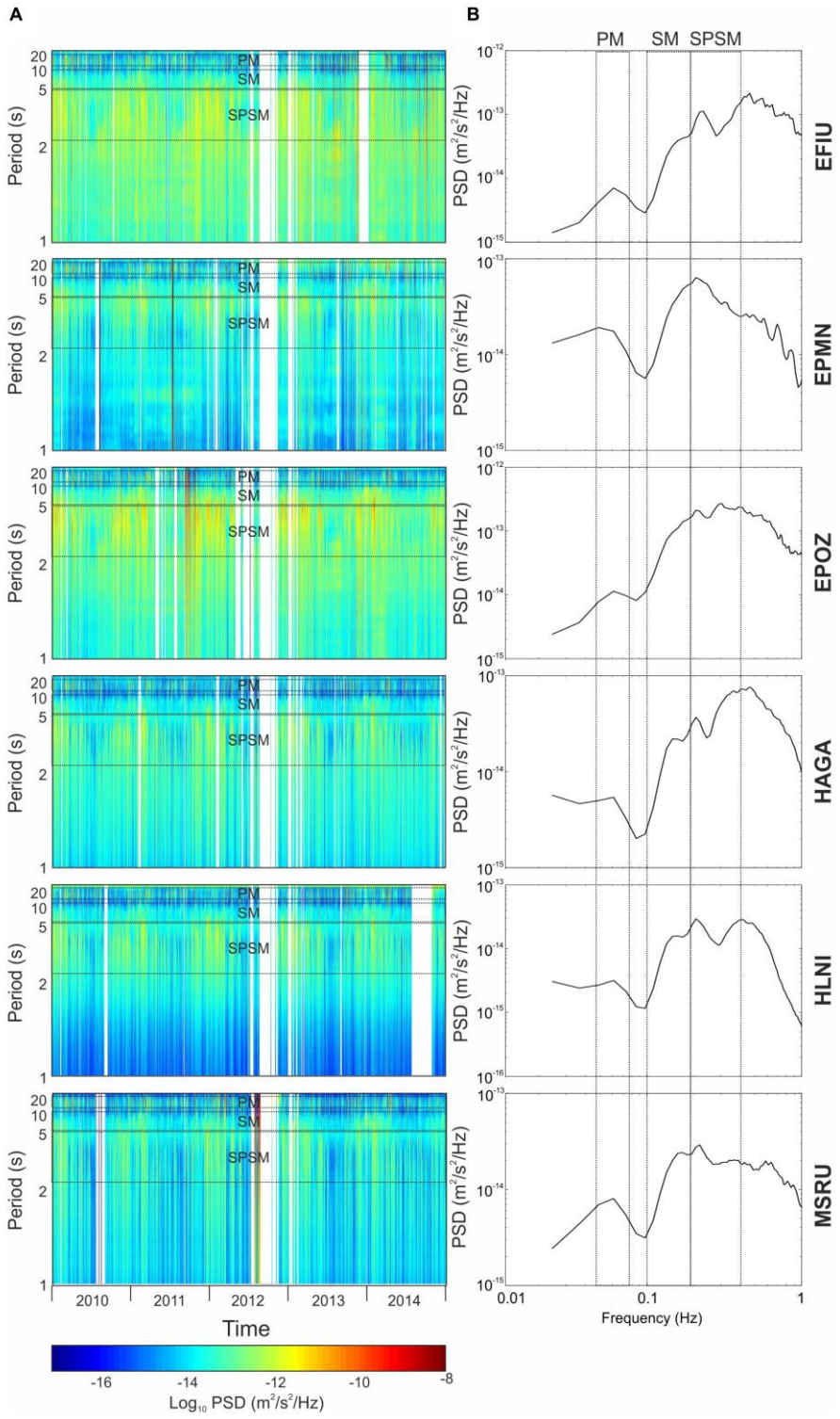

Figure 3.3: (A) Spectrograms of the seismic signal recorded by the vertical component of the six considered stations. (B) Median spectra of the seismic signal recorded by the vertical component of the six considered stations. The acronyms PM, SM, and SPSM indicate primary microseism, secondary microseism, and short-period secondary microseism, respectively.

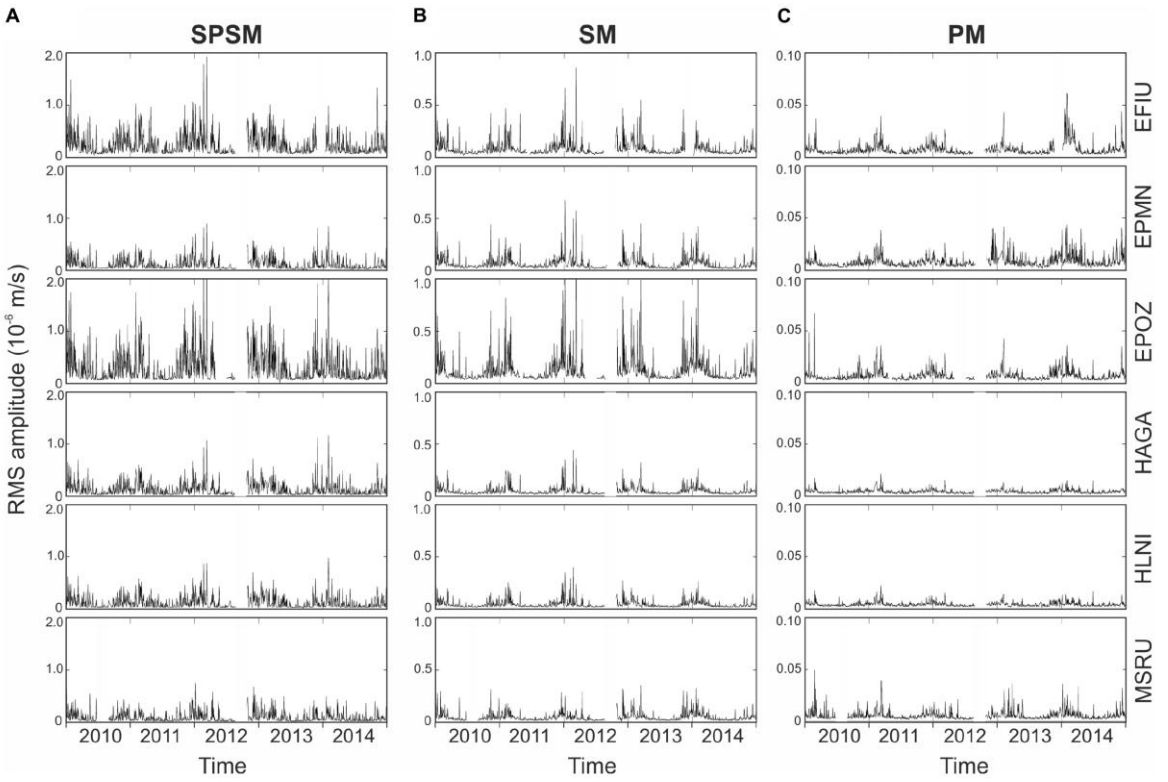

Figure 3.4: RMS amplitude time series of the seismic signal recorded by the vertical component of the six considered stations and filtered in the bands (A) 2.5–5.0 s (SPSM, short-period secondary microseism); (B) 5–10 s (SM, secondary microseism); and (C) 13 - 20 s (PM, primary microseism).

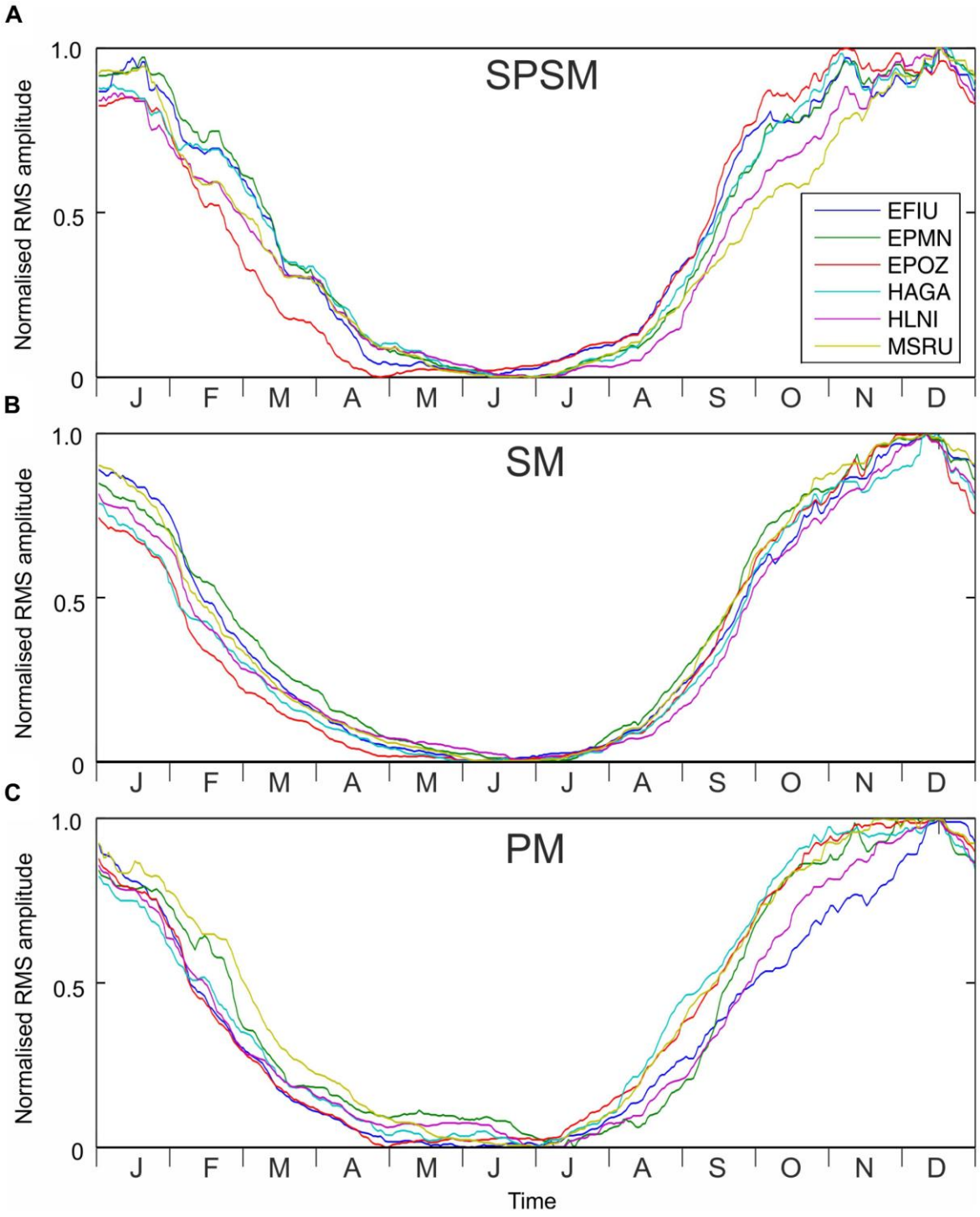

Figure 3.5: RMS amplitude time series smoothed by a 90-day-long moving median, split into 1-year-long windows, stacked, and normalized for all the considered seismic stations (see the legends on the bottom right corner of (A)). In particular, regarding the period bands (A) 2.5–5 s (SPSM), (B) 5–10 s (SM), and (C) 13–20 s (PM). The time on the x-axis of (A–C) indicates the window onset of the 90-day-long moving median.

3.3.2 ARRAY ANALYSIS

To get an idea on the locations of the main microseism sources surrounding the Eastern Sicilian coastlines, the seven stations composing the summit ring of the Mt. Etna seismic permanent network were used as a roughly circular array (Figure 3.1b). The array response functions (ARFs) were computed for the PM, SM, and SPSM for a plane wave arriving with a slowness of 0 s deg^{-1} (Figure 3.6). Such ARFs exhibit that only the SPSM case shows a fairly good resolution. This is due to the very long wavelength of PM and SM compared to the array aperture (~ 5 km). Indeed, taking into account a velocity of the S-waves (Vs) in the first kilometers of the crust equal to ~ 2 km/s (e.g. Hirn et al., 1991; Patanè et al., 1994), the wavelengths of PM, SM, and SPSM are ~ 26, 10, and 5 km, respectively. When the wavelength is much greater than the array aperture (as in the case of PM and SM), the array behaves like a single station (e.g. Schweitzer et al., 2012).

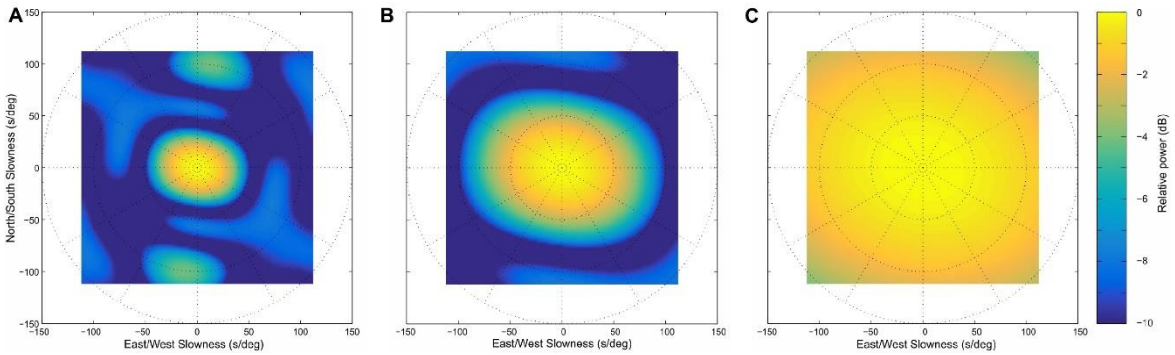

Figure 3.6: Array response functions of the seven stations composing the summit ring of the Mt. Etna seismic permanent network (see Figure 3.1b) for a unit amplitude incident wave with slowness of 0 s deg^{-1} at periods of 2.5 s (A), 5 s (B), and 13 s (C).

The portions of the Ionian and Tyrrhenian coastlines, where the microseism sources closest to the array could supposedly be located, are characterized by a minimum

distance of ~ 20 and ~ 45 km, respectively, from the array center. Such distances are greater than two to three times the array aperture, and hence, on the basis of the synthetic tests performed by Almendros et al. (2002), the Etna summit ring array should be able to locate the microseism sources with a planar wavefront assumption. Then, to apply array analysis, the following processing steps were carried out on the seismic signals: demeaning and detrending, correction for the instrument response, filtering within a 0.2 - to 0.4 Hz band by a second-order Butterworth filter, and subdivision in 60-s-long windows, tapered with a Tukey window. The filter is also used to exclude volcanic tremor, whose energy at Mt. Etna is mainly radiated in the band 0.5–5.0 Hz (Cannata et al., 2010). Successively, the STA/LTA technique (acronym for short time average over long time average; e.g. Trnkoczy, 2012) was applied to detect prospective amplitude transients that could be related to volcano activity (i.e., long period events and very long period events). Windows containing amplitude transients were excluded from the array analysis. Finally, the frequency–wavenumber (*f*–*k*) analysis was carried out, allowing to calculate the power distributed among different slownesses and back-azimuths (e.g. Capon, 1973; Rost & Thomas, 2002).

The array analysis was performed in January 2010–February 2012 on specific time intervals characterized by one of the following two conditions: (i) intense wave activity in the Ionian Sea, as shown by the Catania buoy data and/or by the high RMS amplitude values at EPOZ station; or (ii) intense wave activity in the Tyrrhenian Sea, as suggested by the Cetraro buoy data and/or by the high amplitude RMS amplitude values at MSRU station. Examples of the results for the days 26/04/2011 and 18/12/2011, exhibiting conditions (i) and (ii), respectively, are shown in Figures 3.7.

To evaluate the error associated with the back-azimuth estimation, the jackknife technique (Efron, 1982) was employed as follows. Firstly, the signal window was analyzed by the *f–k* technique by using all the seven stations composing the array. Successively, the analysis was repeated seven times, leaving one station out at a time, so providing further seven back-azimuth values. An arithmetic mean of these estimates was assessed by the following equation:

$$\bar{P} = \frac{1}{n}\sum_{i=1}^{n} P_i \qquad (3.3)$$

where P_i is the back-azimuth value computed by omitting the *i*-th station and *n* is the number of stations composing the array. Then, it is possible to estimate the *i*-th so-called pseudovalue as:

$$J_i = n\hat{P} - (n-1)P_i \qquad (3.4)$$

where $\hat{P}$ is the back-azimuth value computed by considering all the seven array stations. The jackknife estimator of parameter P is given by:

$$J(\hat{P}) = \frac{1}{n}\sum_{i=1}^{n} J_i = n\hat{P} - (n-1)\bar{P} \qquad (3.5)$$

The standard error of the jackknife estimates is given by:

$$\delta_{J(\hat{P})} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(J_i - J(\hat{P}))^2} \qquad (3.6)$$

Finally, median error estimations were calculated separately for the back-azimuths oriented toward the Ionian Sea and the Tyrrhenian Sea.

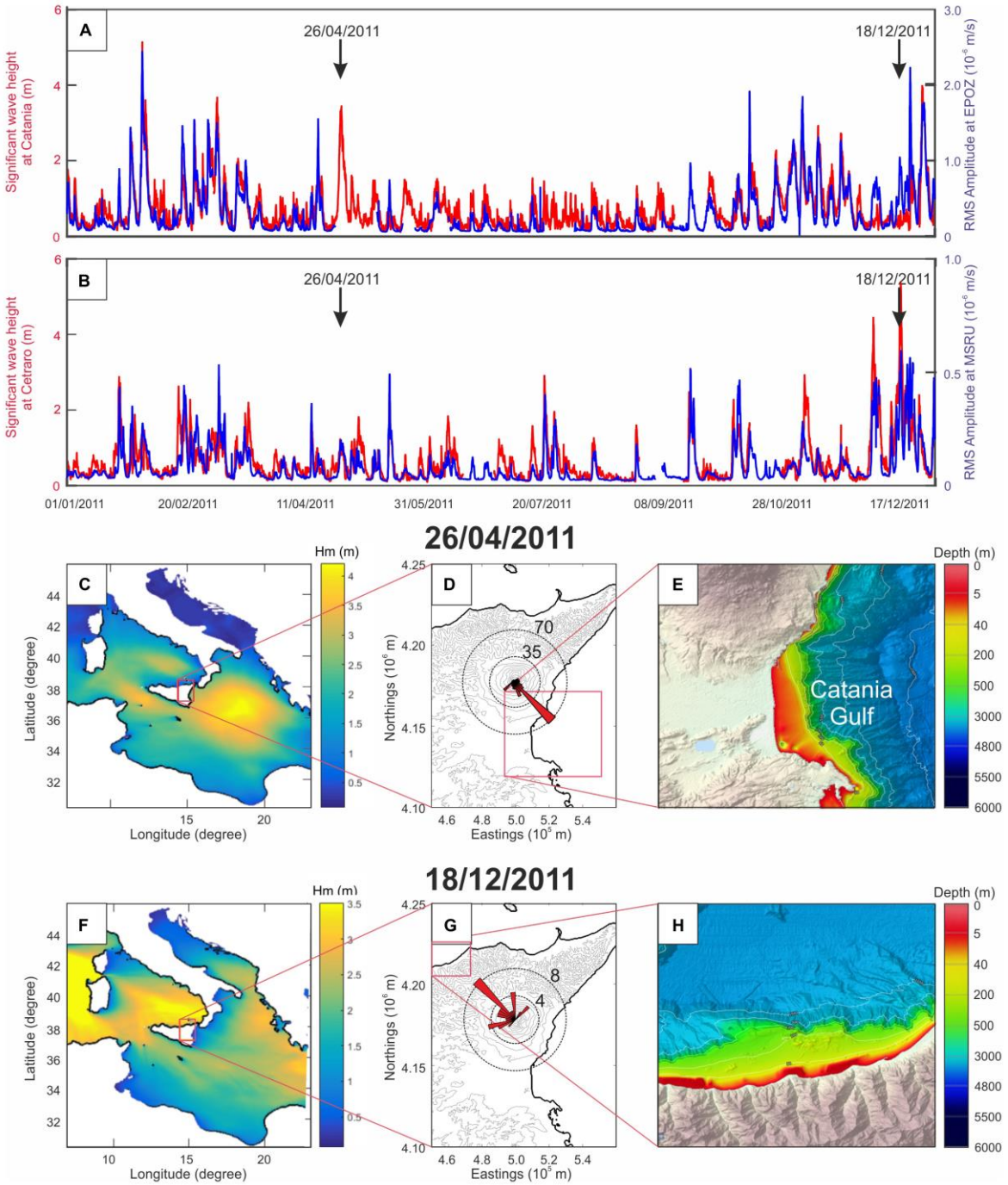

Figure 3.7: (A,B) Time series of significant sea wave height recorded by the Catania and Cetraro buoys (red lines) and the RMS amplitude computed in the period band 2.5–5.0 s (SPSM) by EPOZ and MSRU stations (blue lines) in 2011. (C,F) Maps of a portion of the Mediterranean Sea showing the spatial distribution of the significant wave height on 26/04/2011 at 12:00 and on 18/12/2011 at 12:00, respectively (MEDSEA_HINDCAST_WAV_006_012 from http://marine.copernicus.eu/services-portfolio/access-to-products/). (D,G) Digital elevation models of Eastern Sicily with rose

diagrams, located at the center of the seismic summit ring of Mt. Etna (see Figure 3.1b), showing the distribution of the back-azimuth values on 26/04/2011 and 18/12/2011, computed by *f–k* analysis. (E,H) Maps showing the bathymetry of portions of Sicily coastlines (EMODnet Bathymetry Consortium, 2018).

3.3.3 MACHINE LEARNING APPLICATIONS

Modern machine learning techniques (MLTs) have been tested to build reliable predictive models able to calculate the time series of significant wave height from microseism data. The method, similar to the one proposed in chapter 3 to spatially and temporally reconstruct the sea ice distribution around Antarctica based on the microseism amplitudes, is composed of four main steps (summarized in Figure 3.8): (a) data preparation; (b) training; (c) cross-validation; and (d) testing.

Step (a) consisted of centering and scaling the predictor variables (Kuhn & Johnson, 2013), that is, the 18 time series of the microseism hourly RMS amplitudes from January 2010 to August 2014 (six stations by three frequency bands). The remaining data (September–December 2014) is used for the testing step (d). To center the microseism predictor, the average is subtracted from all the values. Successively, to scale the data, each value of the microseism predictor is divided by its standard deviation. Hence, all the time series of the microseism RMS amplitudes share a common scale.

As for step (b), we made use of the following four MLTs to build predictive models: (i) random forest (RF) regression; (ii) K-nearest neighbors (KNN) regression; (iii) linear regression; and (iv) support vector machine (SVM) regression. For all the above-mentioned MLTs, the 18 time series of the centered and scaled seismic RMS

amplitudes from January 2010 to August 2014 were used as the input, while the two time series of significant wave heights, recorded by the Catania and Cetraro buoys, were resampled by a sampling step of 1 h (the same rate as the seismic RMS amplitude time series) and considered as the output to build the regression models.

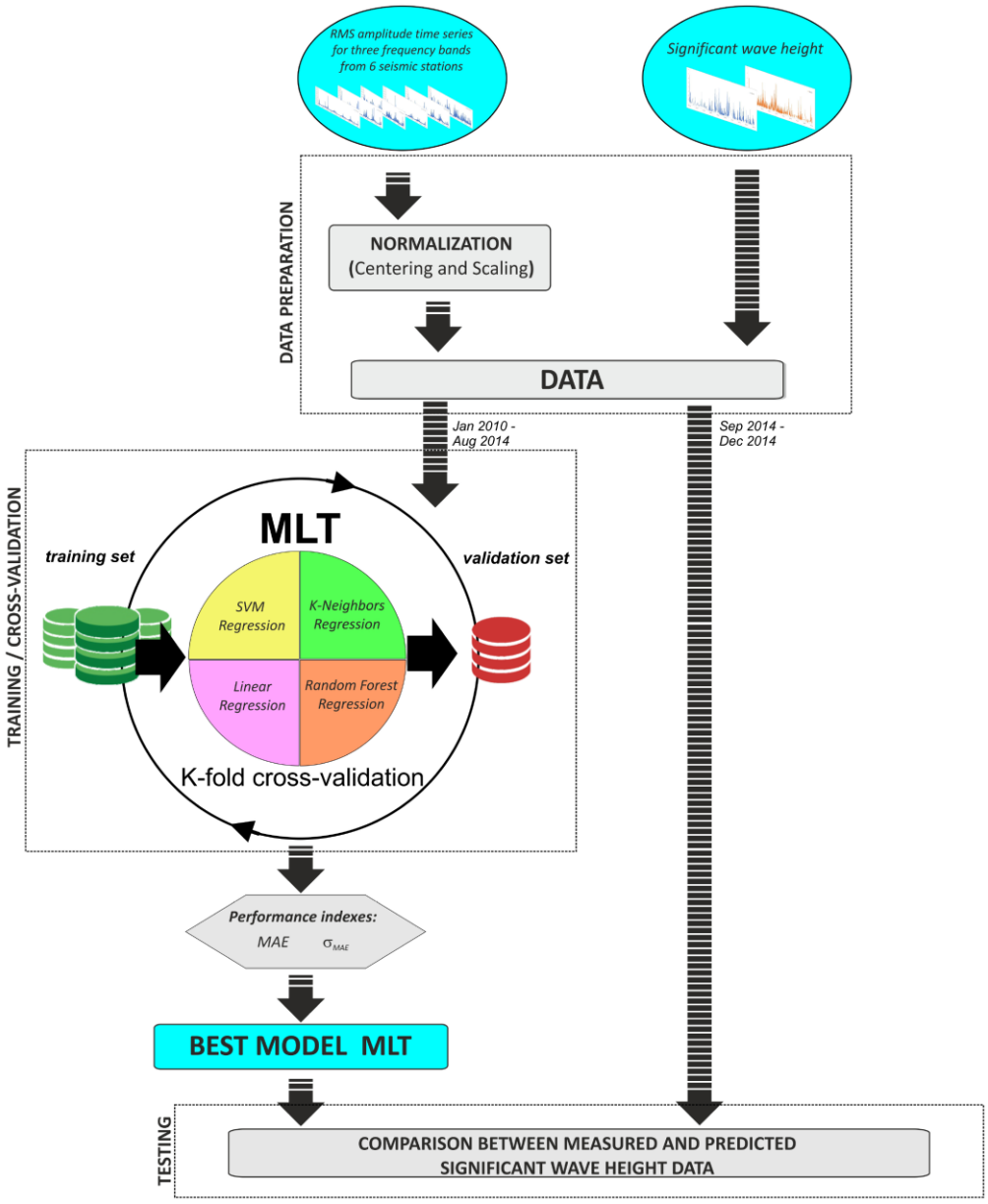

Figure 3.8: Scheme of the modeling analysis to obtain the time series of significant wave height in the Catania and Cetraro buoy locations by using the microseism. MLT, machine learning technique; MAE, mean absolute error; σ_{MAE}, standard deviation computed on the mean absolute error; SVM, support vector machine.

Step (c) consisted of evaluating the best MLT by carrying out the k-fold cross-validation, as explained in detail in the paragraph 2.3.3. The parameters we used to estimate the model performance are: mean absolute error (MAE) between the observed significant wave height and the predicted one and the corresponding standard deviation (σ_{MAE}). The results are shown in Figures 3.9a,b.

The final model was trained with the whole dataset from January 2010–August 2014 and tested on the test set from September–December 2014 (testing step (d)). The comparisons between the predicted and measured significant wave heights for the testing period are reported in Figures 3.10 – 3.13.

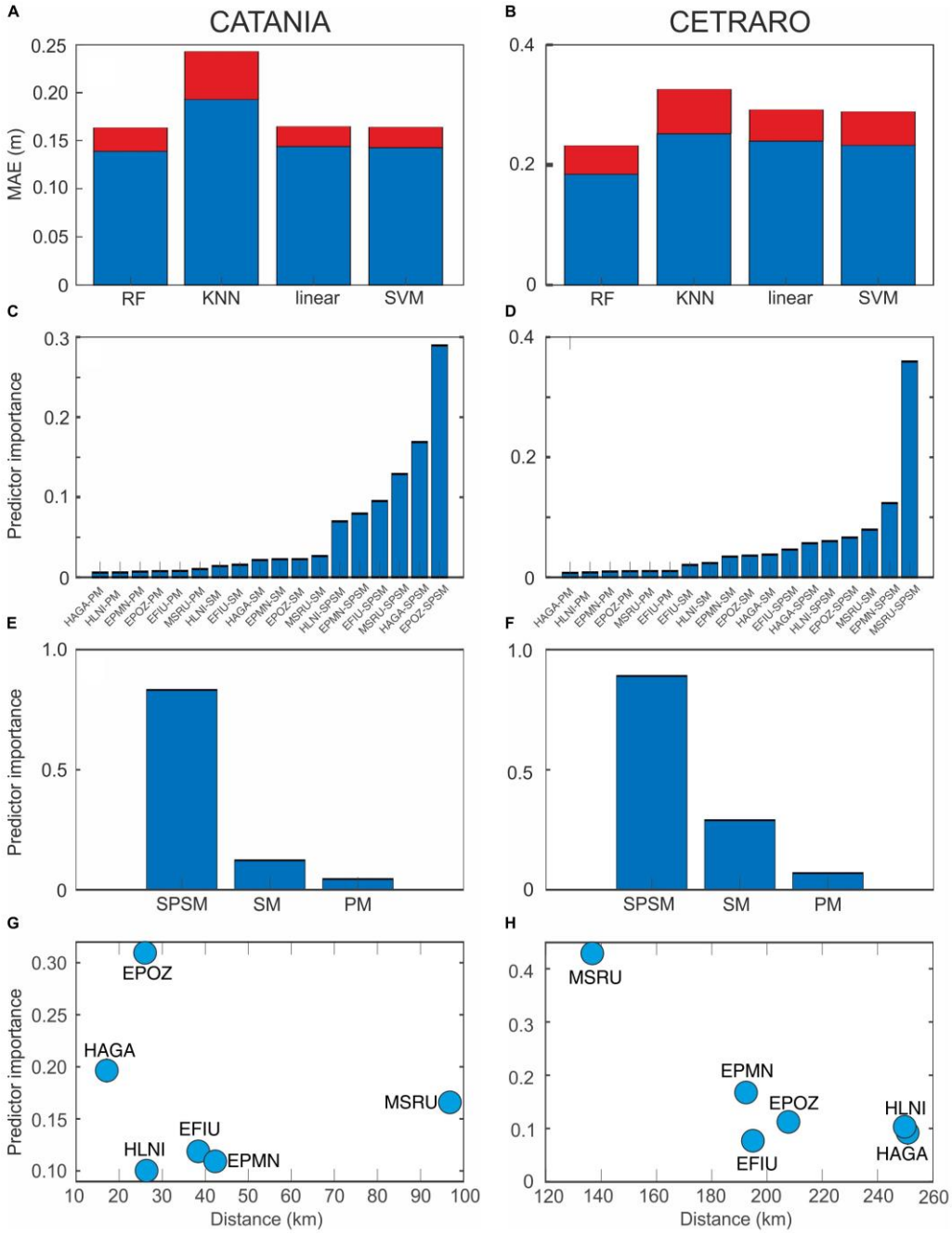

Figure 3.9: Results of the machine learning analysis. (A,B) Average (blue bars) and standard deviation (red bars) of the mean absolute error (MAE), estimated by k-fold cross-validation, for the Catania and Cetraro buoy data, respectively. (C,D) Index of importance for all the input taken into account to model the Catania and Cetraro buoy data, respectively. (E,F) Aggregation through a summation of the input importance allowing to rank the microseism bands for the Catania and Cetraro buoy data prediction, respectively. (G,H) Aggregation through a summation of the station importance for the Catania and Cetraro buoy data prediction plotted versus the distance from the Catania and Cetraro buoys, respectively.

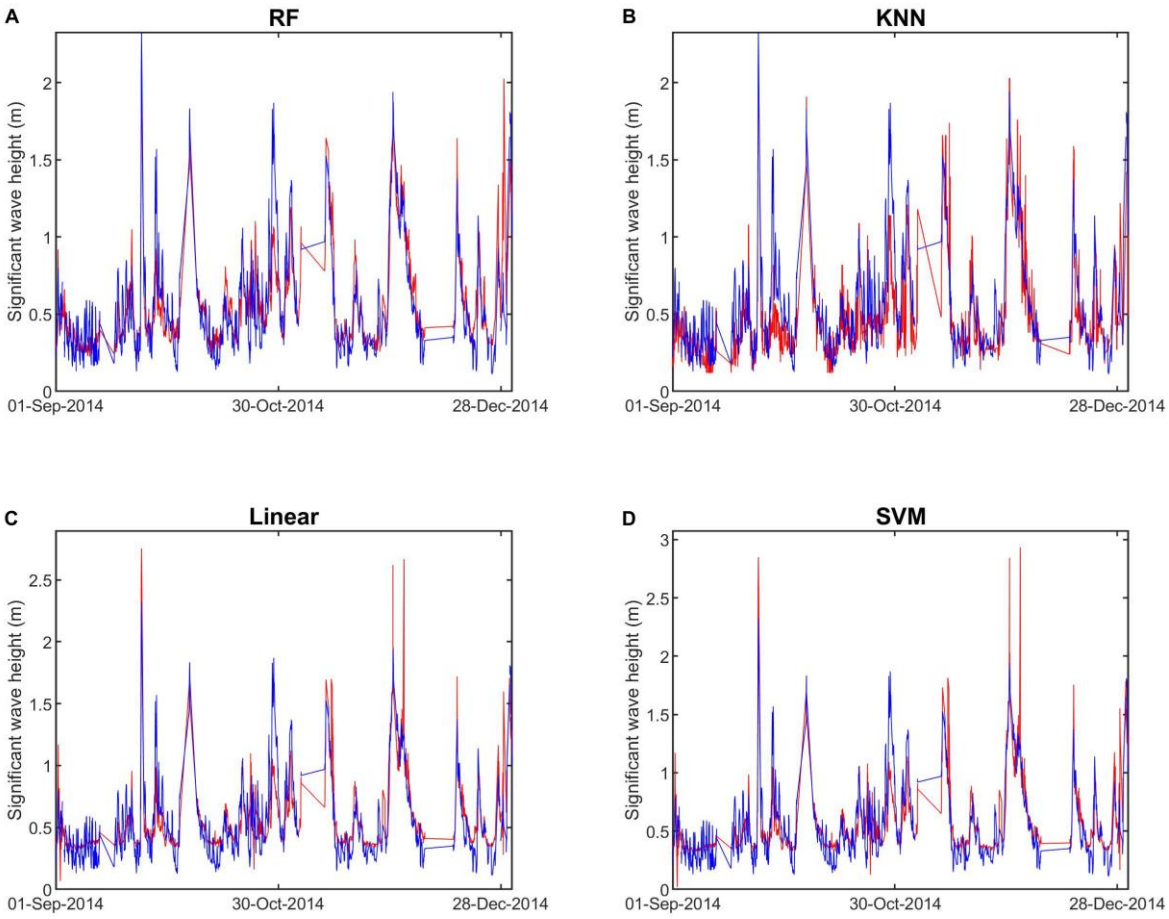

Figure 3.10: Measured (blue line) and predicted (red line) significant wave height time series of the Catania buoy from 1 September to 31 December 2014. The prediction was carried out by RF regression (A), KNN regression (B), linear regression (C), and SVM regression (D).

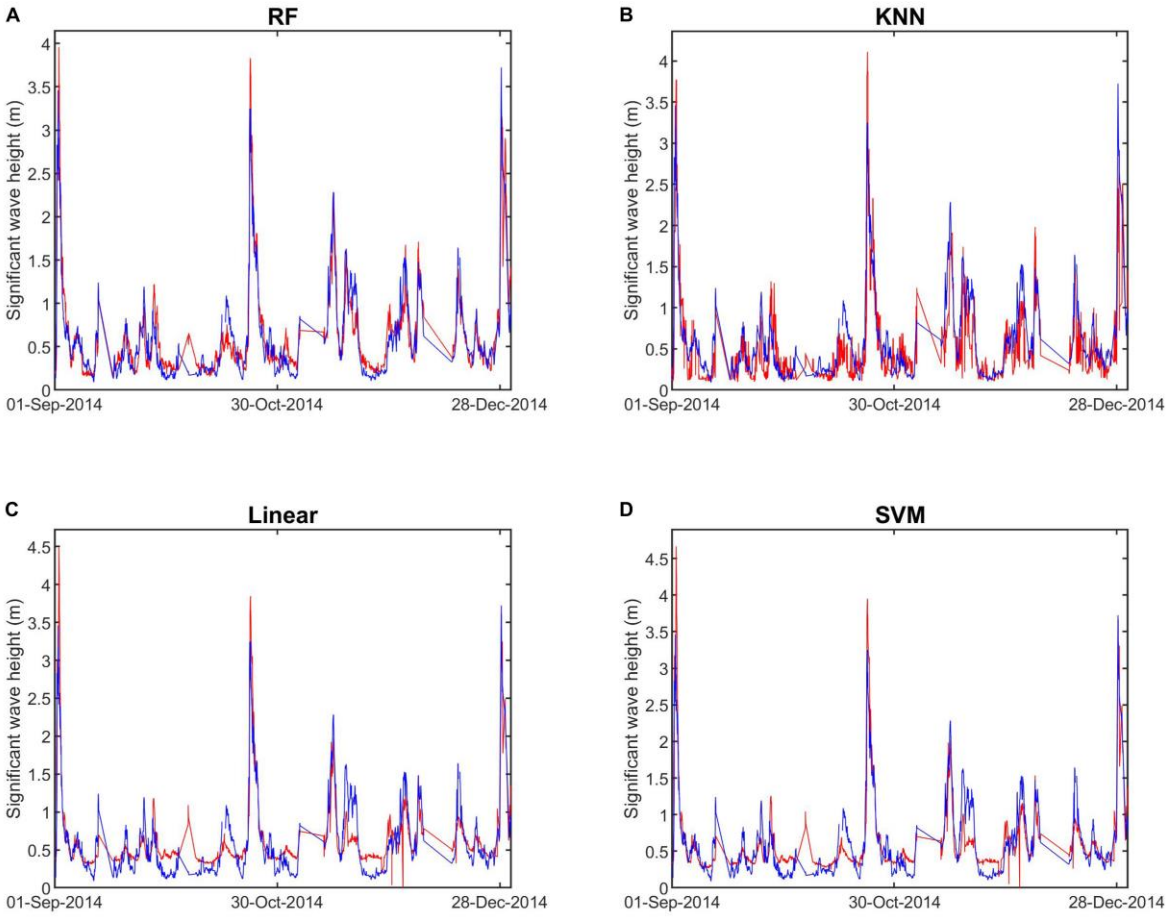

Figure 3.11: Measured (blue line) and predicted (red line) significant wave height time series of the Cetraro buoy from 1 September to 31 December 2014. The prediction was carried out by RF regression (A), KNN regression (B), linear regression (C), and SVM regression (D).

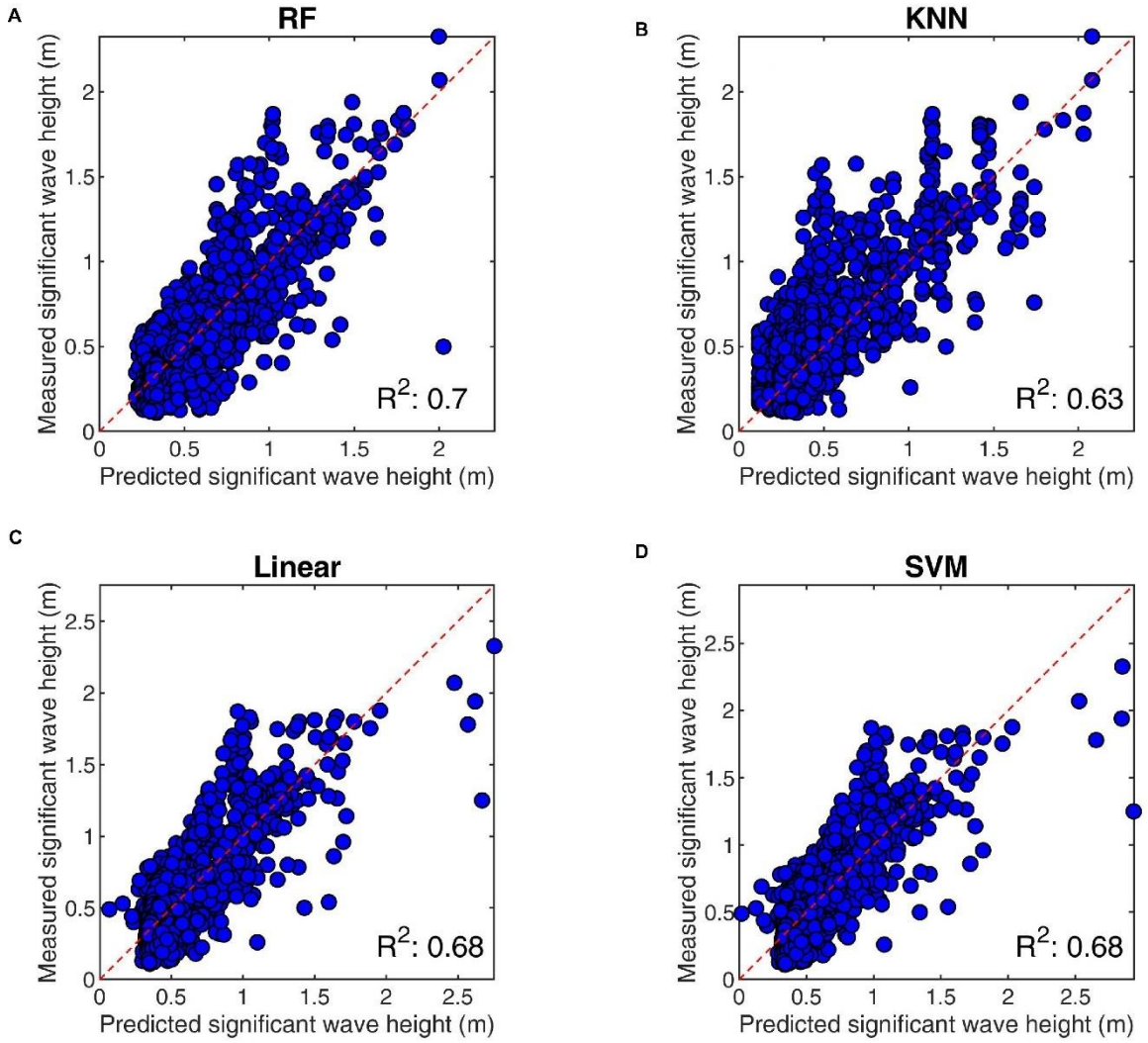

Figure 3.12: Scatter plots showing the measured versus the predicted significant wave heights of the Catania buoy from 1 September to 31 December 2014. The prediction was carried out by RF regression (A), KNN regression (B), linear regression (C), and SVM regression (D). The red dashed line in (A–D) is the y = x line. The value of the determination coefficient (R^2) is also reported in the bottom right corner of the plots.

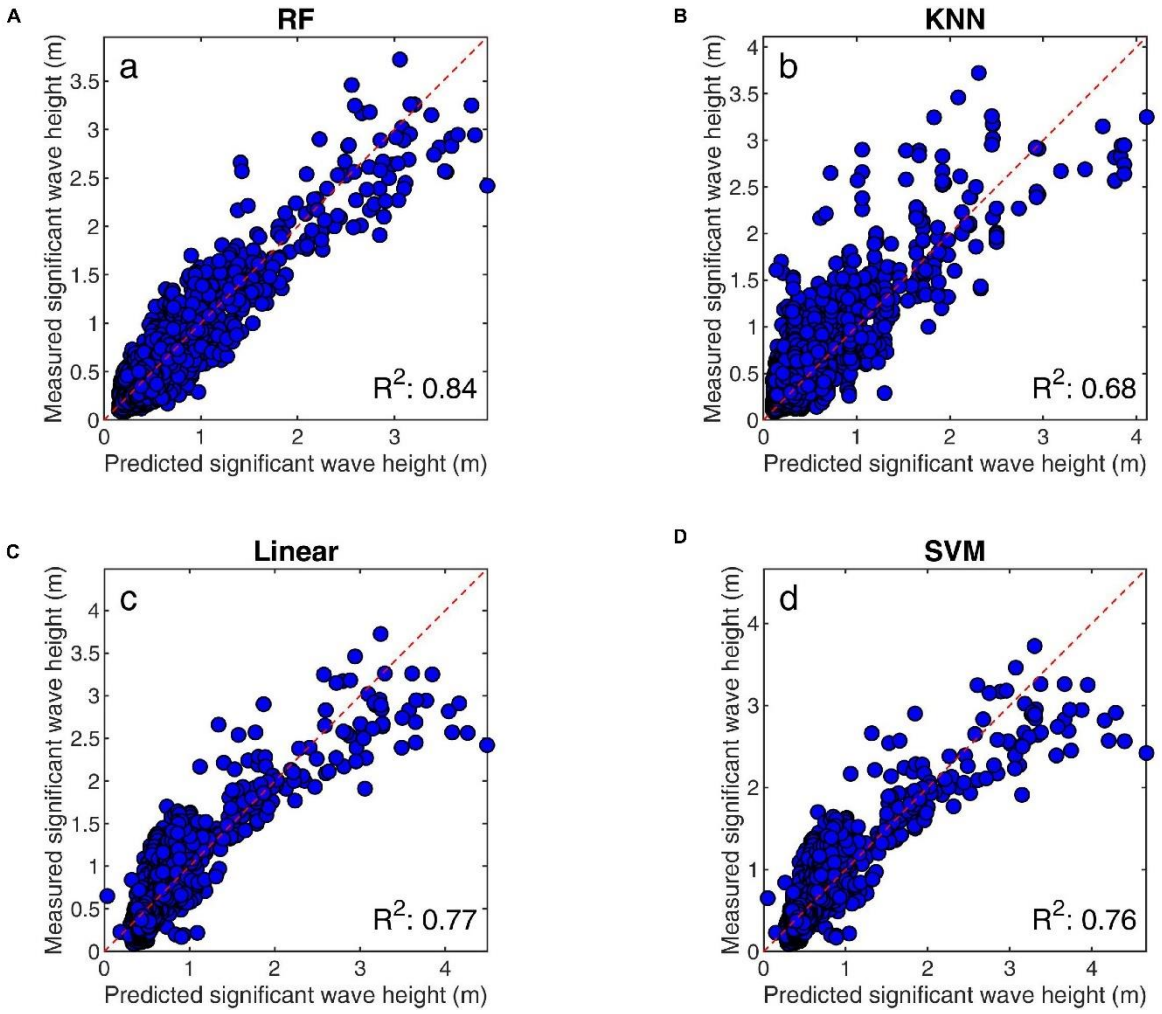

Figure 3.13: Scatter plots showing the measured versus the predicted significant wave heights of the Cetraro buoy from 1 September to 31 December 2014. The prediction was carried out by RF regression (A), KNN regression (B), linear regression (C), and SVM regression (D). The red dashed line in (A–D) is the y = x line. The value of the determination coefficient (R^2) is also reported in the bottom right corner of the plots.

3.4 RESULTS

3.4.1 RESULTS OF THE SPECTRAL AND RMS AMPLITUDE ANALYSES

Concerning the spectral analysis, it is worth noting that in most of the considered stations (EPOZ is the only exception), it is possible to clearly distinguish three peaks in the spectra, corresponding to primary, secondary, and short-period secondary microseisms (PM, SM, and SPSM, respectively, in Figure 3.3b). Furthermore, as observed in the seismic recordings around the world (e.g. Aster et al., 2010), the spectral analysis of the microseism acquired in Eastern Sicily shows that most of the energy is recorded in the secondary and short-period secondary microseism bands, while the primary microseism exhibits a much weaker spectral amplitude (Figure 3.3). In particular, the short-period secondary microseism peak is the strongest among the three in all the considered stations (Figure 3.3b). High values of spectral amplitude can also be observed at frequencies above the short-period secondary microseism band (> 0.4 Hz). In addition, in both spectrograms and RMS amplitude time series, it is possible to observe a seasonal modulation with maxima during the winters (December – February) and minima during the summers (June – August) (Figures 3.4, 3.5).

3.4.2 RESULTS OF THE ARRAY ANALYSIS

As for the array analysis, the summit ring of Mt. Etna seismic permanent network turned out to be effective in locating the microseism sources in the SPSM band (Figure 3.4a). During Ionian stormy days, the back-azimuth values indicate the Catania Gulf (Figure 3.7d), while during Tyrrhenian stormy days the back-azimuth

rotates, pointing north–westward (Figure 3.7g). In both cases, the SPSM sources appear to be located in areas of extended shallow water depths (Figure 3.7e,h). Concerning the median error in the back-azimuth estimations obtained by the jackknife technique, it was equal to 21° and 12° for back azimuths oriented toward the Tyrrhenian and Ionian Seas, respectively. As for the apparent seismic velocity estimations, the histograms in Figure 3.14 show values of ~1.5–2.0 km/s.

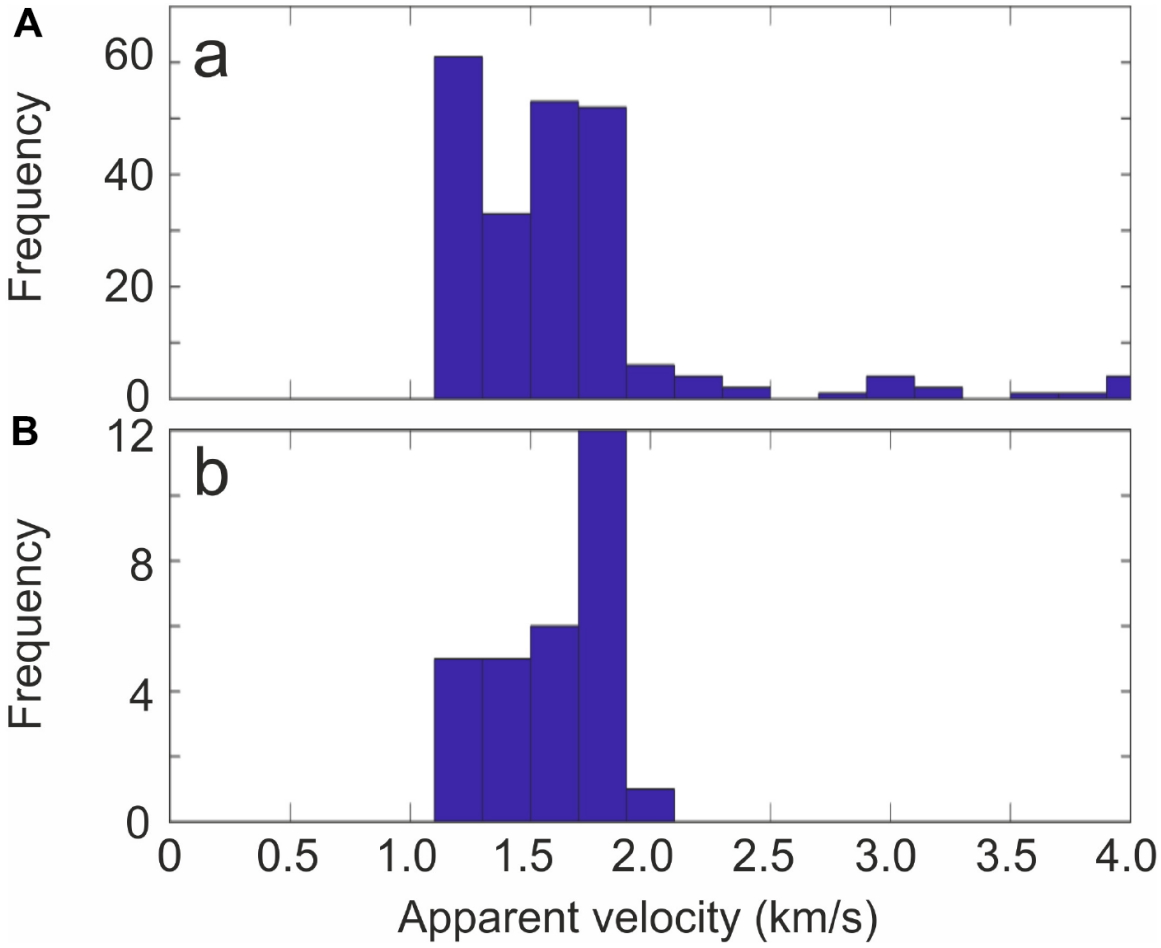

Figure 3.14: Histograms showing the apparent velocity estimated by *f–k* analysis on 26/04/2011 (a) and on 18/12/2011 (b).

3.4.3 RESULTS OF MACHINE LEARNING APPLICATIONS

MLTs have been able to reconstruct the time series of significant sea wave height on the basis of microseism data. The technique showing the best performance was RF regression (Figures 3.9a,b), allowing to get the minimum MAEs equal to 0.14 ± 0.02 m and 0.18 ± 0.05 m for the Catania (the Ionian Sea) and Cetraro (the Tyrrhenian Sea) data, respectively.

It has to be underlined that the RF, linear, and SVM regressions show very similar MAE values, especially in the case of the Catania buoy. The RF approach has the advantage of easily supplying an index of predictor importance (Figures 3.9c,d), calculated by exploiting the random permutation of out-of-bag samples (Breiman, 2001). To get information on the importance of the different microseism bands in the prediction, aggregation through summation was performed (Figures 3.9e,f), showing how the SPSM band has the highest weight in reconstructing the significant wave height time series at the two buoys. In addition, aggregation through summation was performed also for the station importance and exhibited how the importance tends to decrease with increasing distance of station–buoy (Figures 3.9g,h).

Finally, the comparison between the measured and predicted significant wave height data during the testing period (September–December 2014; Figures 3.12, 3.13) showed very similar patterns for the two time series, as also confirmed by the high values of determination coefficient equal to 0.7 and 0.84 for the Catania and Cetraro buoys, respectively, in the case of RF regressions.

3.5 DISCUSSION AND CONCLUSION

Concerning the microseism characterization, as measured in the seismic signals acquired worldwide (e.g. Aster et al., 2010), most of its energy is contained in the SPSM and SM bands (Figure 3.3). This can be due to the fact that the considered stations are very close to the sea coastlines (maximum distance equal to ~19 km for HLNI) and, hence, record well the short-period secondary microseism, whose sources are linked to local sea state and wave activity, and influenced by local winds (e.g. Bromirski et al., 2005; Chen et al., 2011). According to Bromirski et al. (2005), the clear split between the spectral peaks of short-period secondary microseisms and secondary microseisms, observed in most of the computed spectra (Figure 3.3b), is a common feature for stations located nearby the shoreline.

Spectral analysis also shows relatively high amplitudes at frequencies above the short-period secondary microseism band (Figure 3.3b). According to previous studies (e.g. Gal et al., 2015; Möllhoff, & Bean, 2016) such a high-frequency microseism could be related to wind-driven ocean waves breaking on the shoreline. In any case, we cannot exclude in this band (> 0.4 Hz) small contributions from anthropogenic seismic noise, as well as from the continuous volcanic tremors recorded at the seismic stations located on or nearby the Etna volcano edifice (such as EPOZ and EFIU) and characterized by a frequency band 0.5–5.0 Hz (Cannata et al., 2010).

Seasonal modulation, identified in both spectrograms and seismic RMS amplitude time series (Figures 3.3 – 3.5), has been observed in all the areas worldwide with temperate latitudes (e.g. Aster et al., 2008; Stutzmann et al., 2009), and is due to the more efficient energy transfer from the sea to the solid earth in winters, when the seas are stormier. Areas located close to the Equator, as well as regions at very high

latitudes, do not show such a pattern: in the former, noise amplitude is mostly stable over the year (e.g. Stutzmann et al., 2009), while in the latter, the microseism seasonal modulation is different, as during the winters the oceanic waves cannot efficiently excite seismic energy because of the sea ice (e.g. Grob et al., 2011; Cannata et al., 2019).

Taking into account the array analysis, performed by the seven seismic stations in Figure 3.1b by the *f–k* array technique in the SPSM band, we were able to obtain the slowness vector direction and, therefore, to get an idea on the locations of the microseism source in the SPSM band. It was observed that the SPSM sources appear to be located in areas of extended shallow water depths: the Catania Gulf and a part of the Northern Sicily coastlines (Figure 3.7d,e,g,h).

The array analysis results are in agreement with Chen et al. (2011), who analyzed microseism data collected in Taiwan and showed how a stronger excitation in SPSM takes place in the narrow Taiwan Strait where the water depth is very shallow, while the excitations are relatively weak in the eastern offshore area, an open sea with water depth increasing rapidly off the coast. Although Juretzek & Hadziioannou (2017) focused on a different frequency band (PM), they also constrained the source locations of the microseism recorded in Europe in regions with extended shallow water areas, that is, Norwegian and Scottish coasts.

It has to be noted that the error associated with the microseism source locations is higher in the case of the Northern Sicily coastlines compared to the Catania Gulf. It derives from both the higher back-azimuth error (21° for the Tyrrhenian Sea versus 12° for the Ionian Sea) as well as from the longer distance array-Northern Sicily coastlines (~ 45 km) compared to the distance array–Catania Gulf (~ 20 km).

The apparent seismic velocity estimations of 1.5–2.0 km/s in the SPSM band (Figure 3.14) are in agreement with the Rayleigh wave velocity calculated by using beamforming analysis, applied on the ambient seismic noise in New Zealand, by Brooks et al. (2009), as well as with the results obtained from investigating the seismic noise in the northeast of the Netherlands by Kimman et al. (2012). In addition, Rivet et al. (2015) also estimated comparable velocities (of 1.5 km/s at 1 Hz and 2.0 km/s at 0.5 Hz) by using a time–frequency analysis to measure the group velocity of Rayleigh wave on noise cross-correlation.

Machine learning analysis is able to reconstruct the time series of significant wave height by using microseism recorded in different period bands by distinct seismic stations. Such a method allows to reliably compute the significant wave height in two locations, coinciding with the two buoys in the Ionian and Tyrrhenian Seas, with fairly low error (MAE equal to ~ 0.14 m for the Catania buoy and ~ 0.18m for the Cetraro buoy; Figures 3.9a,b).

In particular, the MLT which showed the best performance was the RF regression. This can be related to several factors, such as: (i) the performance of the RF regression is not much affected by parameter selection (e.g. Li et al., 2011; Kuhn & Johnson, 2013); (ii) by making use of an ensemble of decision trees, RF regression does not overfit with respect to the source data (e.g. Li et al., 2011); and (iii) RF shows robustness to outliers and noise (Breiman, 2001). Finally, compared to linear regressions, RF regression is able to deal with non-linear relationships between the input and output. Indeed, according to Essen et al. (2003) and Craig et al. (2016), the relationship linking microseism amplitude and significant wave height is likely to be non-linear.

Focusing on the comparison between the highest measured and predicted (by RF regression) significant wave height data during the testing period, it is possible to note a slight underestimation and overestimation of the predicted values compared to the measured ones in the Catania and Cetraro cases, respectively (Figures 3.12a, 3.13a). These different behaviours could be related to the different distances between the seismic stations and the buoys.

The RF regression also provides an index of importance of the distinct predictor variables, which are the seismic RMS amplitude time series. The aggregated importance of the different frequency bands exhibits how the SPSM band contains most of the information for the buoy data reconstruction (Figures 3.9e,f). According to the literature (e.g. Bromirski et al., 2005; Chen et al., 2011; Gualtieri et al., 2015), such a microseism band, characterized by high frequencies and then by quick attenuation with distance, is mostly generated by sources located in relatively shallow water close to the shelf break, close to the seismic stations. Such sources are likely related to local nearshore non-linear wave–wave interaction (e.g. Bromirski et al., 2005). This is in agreement with the location of the considered buoys, close to the coastlines, in shallow water conditions (90 and 100 m for Catania and Cetraro, respectively; Bencivenga et al., 2012). Both PM and SM turned out to have a much smaller importance for the buoy data reconstruction. Indeed, as for PM, its dominant source regions can be located thousands of kilometers away from the seismic stations (Gualtieri et al., 2019). Concerning SM, it has been shown how it can also have pelagic sources in deep ocean (e.g. Chevrot et al., 2007; Kedar et al., 2008).

In addition, the difference in the predictors with the maximum importance for the two buoys (EPOZ-SPSM for the Catania buoy and MSRU-SPSM for the Cetraro buoy;

Figures 3.9c,d) reflects the different locations of the seismic stations. Indeed, EPOZ is very close to the coastline of the Ionian Sea, where the Catania buoy is installed, while MSRU is placed nearby the Tyrrhenian Sea, where the Cetraro buoy is located (Figures 3.1a, 3.9g,h). Hence, the closer the seismic station is to the sea, the more information concerning the sea state are contained in the recorded microseism.

4. MICROSEISM AND SEA WAVE HEIGHT RELATIONSHIP IN EASTERN SICILY USING HINDCAST MAPS

4.1 INTRODUCTION

The monitoring of sea state is a fundamental task for economic activities in coastal areas, such as transportation, tourism, and design of infrastructures (e.g. Von Storch et al., 2015; Ferretti et al., 2018). In particular, the importance of monitoring sea wave height for marine risk assessment and mitigation is increasing day-by-day. This is partially due to global warming that is making sea waves stronger and, hence, the extreme wave events more intense and frequent (Reguero et al., 2019).

Among the different instruments used to monitor the sea state (e.g. Holthuijsen, 2007), wave buoys can be currently considered the most used and reliable instruments for in situ measurements of the offshore and coastal wave climate. However, they show serious drawbacks because of the high costs of both installation and maintenance and the permissions needed for their installation (Orasi et al., 2018).

Other useful tools to monitor the sea state are remote sensing instruments such as radar altimeters and synthetic aperture radar, providing data from large marine areas. Despite their good spatial coverage, they suffer from limited temporal resolution (Fu et al., 2000; Musa et al., 2015; Wiese et al., 2018). An important improvement in sea state monitoring is represented by high-frequency (HF) radar measurements, providing directional wave spectra at a high temporal resolution, with a spatial resolution that depends on both bandwidth and antenna (Wyatt & Green, 2009a, 2009b; Orasi et al., 2018).

Finally, seismic signals recorded by seismometers can also be used as a sea "wavemeter" (Zopf et al., 1976). Indeed, ocean gravity waves cause pressure fluctuations that transfer energy from the ocean to the solid earth, generating a seismic noise called microseisms (e.g. Bromirski et al., 2002).

In the following sections, by statistical and machine learning approaches, we will explore the relationship between microseism amplitude, recorded by seismic stations located in Eastern Sicily (Italy), and the sea wave height in the Ionian Sea, Tyrrhenian Sea, and Sicily Channel. Following the idea by Ferretti et al. (2018), the latter data have been provided by the hindcast maps, which show several advantages with respect to other measured data, such as high spatial resolution, widespread coverage on large areas, and temporal continuity of the information (hindcast maps are always available, as they are not affected by instrument breakage). In particular, we will try to answer the question of whether, and to what extent, it is possible to reconstruct the spatial distribution of sea wave height by using microseisms recorded at distinct stations and in different frequency bands.

4.2 DATA

4.2.1 SEISMIC DATA

In order to investigate microseism, seismic signals recorded from 2010 to 2017 by the vertical component of six stations, belonging to the seismic permanent network run by Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo – Sezione di Catania (INGV-OE), were used (Figure 4.1, Table 4.1). These stations are equipped with broadband three-component Trillium 40-s seismometers (Nanometrics^{TM}) recording at a sampling rate of 100 Hz.

A different time period has been chosen from the one analysed in chapter 4 because of the different time coverage of buoy data (until 2014) with respect to the hindcast maps (available up to 2017).

Moreover, the six seismic stations used in this chapter are different respect to the six used in chapter 3. In particular, the AIO seismic station was used instead of EMPN because of the shorter data gaps in AIO from 2015 to 2017.

Station Name	Latitude	Longitude	Altitude (m a.s.l.)
MSRU	38.2638	14.5083	401
AIO	37.9689	14.2278	715
EFIU	37.7897	14.2101	98
EPOZ	37.6718	14.1885	119
HAGA	37.2853	14.1552	126
HLNI	37.3485	14.872	147

Table 4.1: Coordinates of the seismic stations used for correlation and machine learning analyses.

4.2.2 SEA DATA

As for the sea data, the "MEDSEA_HINDCAST_WAV_006_012" product, shared by the Copernicus Marine Environment Monitoring Service (CMEMS) http://marine.copernicus.eu/, was used. Such a product is the hindcast product of the Mediterranean Sea Waves forecasting system, based on the third-generation wave model WAM Cycle 4.4.4 (Behrens, 2013). In particular, the hourly significant wave height data of a portion of the Mediterranean Sea (shown in Fig.4.1) at 1/24° horizontal resolution was extracted and used to make comparisons to the microseism.

To speed up the computation, one grid cell every two in both the east–west and north–south directions was taken into account for the analyses performed in this work.

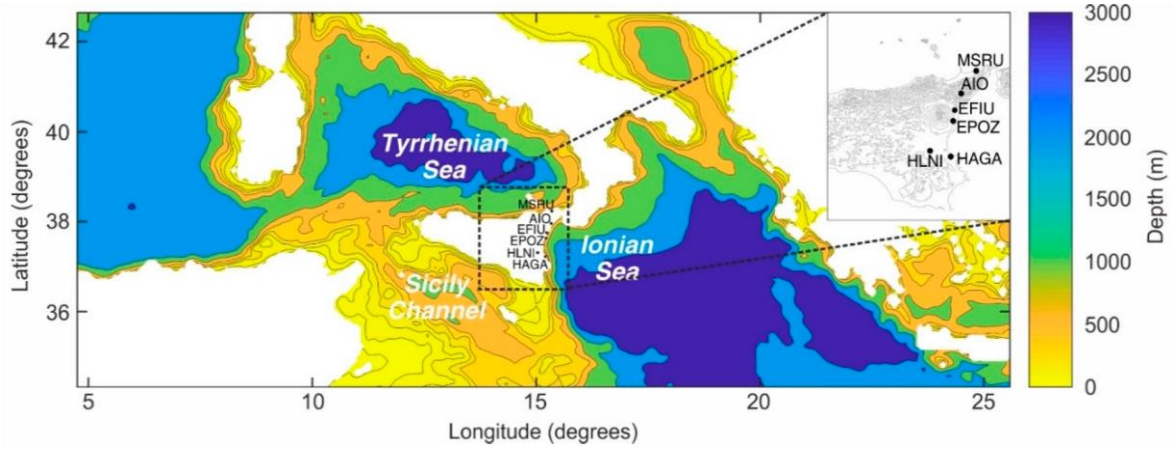

Figure 4.1: Map of a portion of the Mediterranean Sea, showing bathymetry (data from http://marine.copernicus.eu/) and the location of the six seismic stations used for correlation and machine learning analysis with hindcast significant sea wave heights (black dots and labels). The inset in the right upper corner shows the digital elevation model of Eastern Sicily, with the location of the seismic stations.

4.3 METHODS OF ANALYSIS

4.3.1 SPECTRAL AND AMPLITUDE ANALYSES

The spectral content of the seismic data recorded by the vertical component of the six seismic stations shown in Figure 4.1 was analysed following the same workflow adopted in paragraph 2.3.1. All the daily spectra were collected and visualized as spectrograms, which are 3D plots with time on the x-axis, frequency on the y-axis, and power spectral density (PSD) indicated by a color scale (Figs. 4.2a-f). In addition, to obtain information on the spectral features of the seismic signals recorded by the different stations during the whole investigated period, all the daily spectra composing the spectrograms were averaged (Figs. 4.2g-l).

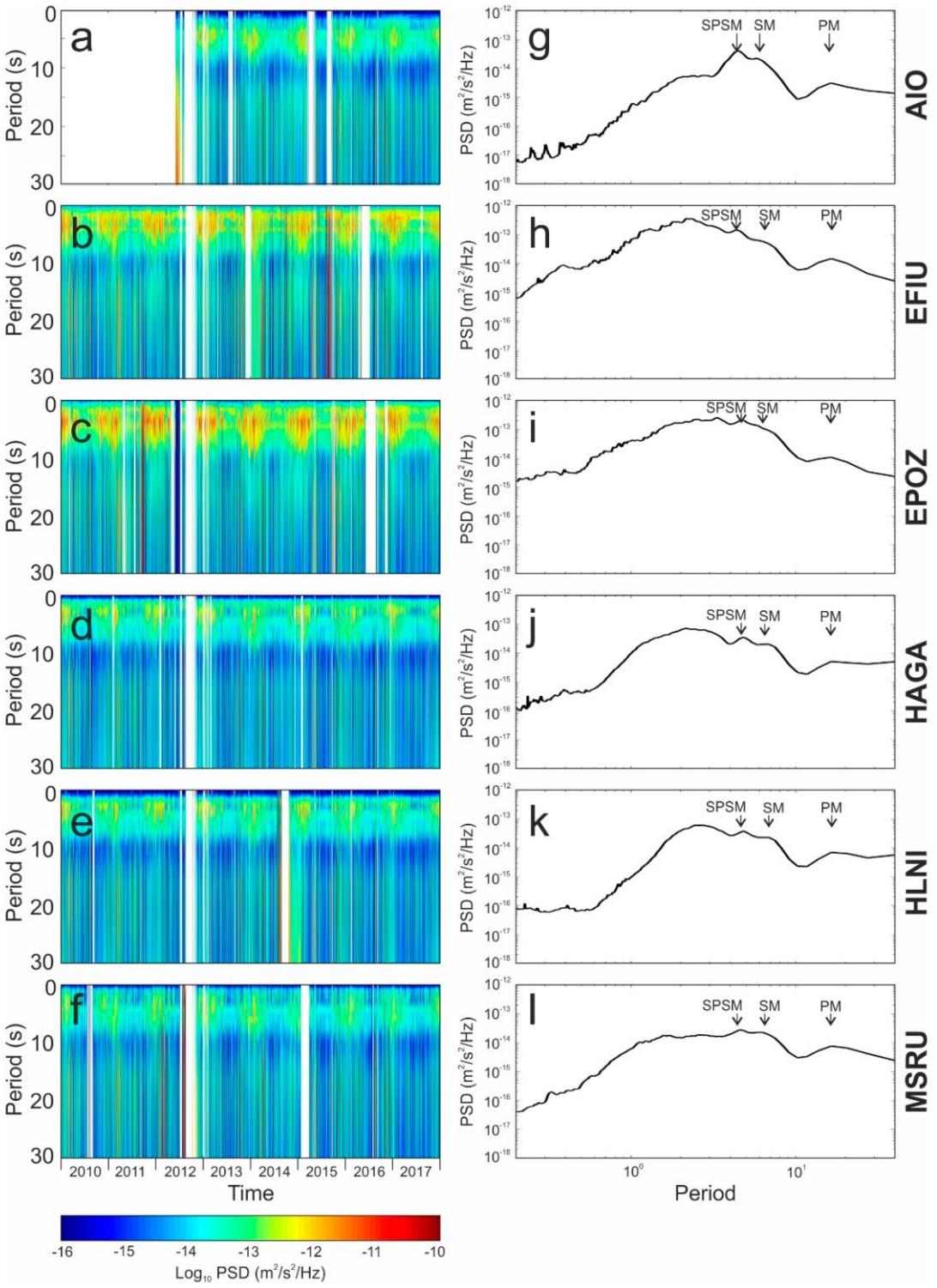

Figure 4.2: (a–f) Spectrograms of the seismic signal recorded by the vertical component of the six considered stations. (g–l) Median spectra of the seismic signal recorded by the vertical component of the six considered stations. Arrows and labels in (g–l) indicate the spectral peaks corresponding to short-period secondary microseisms (SPSM), secondary microseisms (SM) and primary microseisms (PM).

The time variability of the microseism amplitude was investigated by calculating the the hourly root-mean-square (RMS) amplitude of the seismic signal filtered in different frequency bands, using the equation 2.1: (i) 0.050–0.077, (ii) 0.1–0.2, (iii) 0.2–0.4, (iv) 0.025–0.050, (v) 0.4–0.8, and (vi) 0.05–0.4 Hz. In particular, the band (i) corresponds to the primary microseism, (ii) to the secondary microseism, (iii) to the short-period secondary microseism, (iv) has a frequency lower than the primary microseism, (v) is characterized by a frequency higher than the short-period secondary microseism, and, finally, (vi) comprises primary, secondary, and short-period secondary microseisms. Bands (iv) and (v) have been introduced to verify the presence of sea dynamics-related seismic signals at frequencies different from the classical primary, secondary, and short-period secondary microseism bands.

4.3.2 CORRELATION ANALYSIS

Following the idea of several authors (e.g. Bromirski et al., 2001; Essen et al., 2003; Craig et al., 2016; Xiao et al., 2018), maps showing the spatial variability of correlation coefficients, computed between the time series of seismic RMS amplitudes and the time series of significant wave height, obtained in each grid cell of the hindcast maps, were calculated. In particular, for the correlation analysis, we calculate the Spearman correlation coefficient, as explained in paragraph 2.3.2.

As such time series should correlate well at or near the source region of the microseism (Craig et al., 2016), this analysis provides an idea of the location of the main sources of microseisms recorded by the considered seismic stations.

As both seismic RMS amplitudes and significant wave heights are characterized by a sampling period of 1 h (much longer than any physically possible delays between

the two time series), a simple zero-lag correlation was computed. Each considered time series of seismic RMS amplitude and significant wave height has a duration of 8 years and an hourly sampling period, thus totalling to ~70,000 samples per series. As for the spatial resolution, 101 × 251 grid cells belonging to the hindcast maps (corresponding to 25,351 correlation coefficients computed for each station and each frequency band) were considered.

The correlation maps between significant wave height and seismic RMS amplitudes, computed at distinct seismic stations and different frequency bands, are shown in Figure 4.3. Furthermore, in order to highlight relationships between the correlation coefficient and distance from the station recording the microseism to the sea grid cell providing the significant wave height data, cross-plots showing the distance in the x-axis and the correlation coefficient in the y-axis were obtained (Figure 4.4).

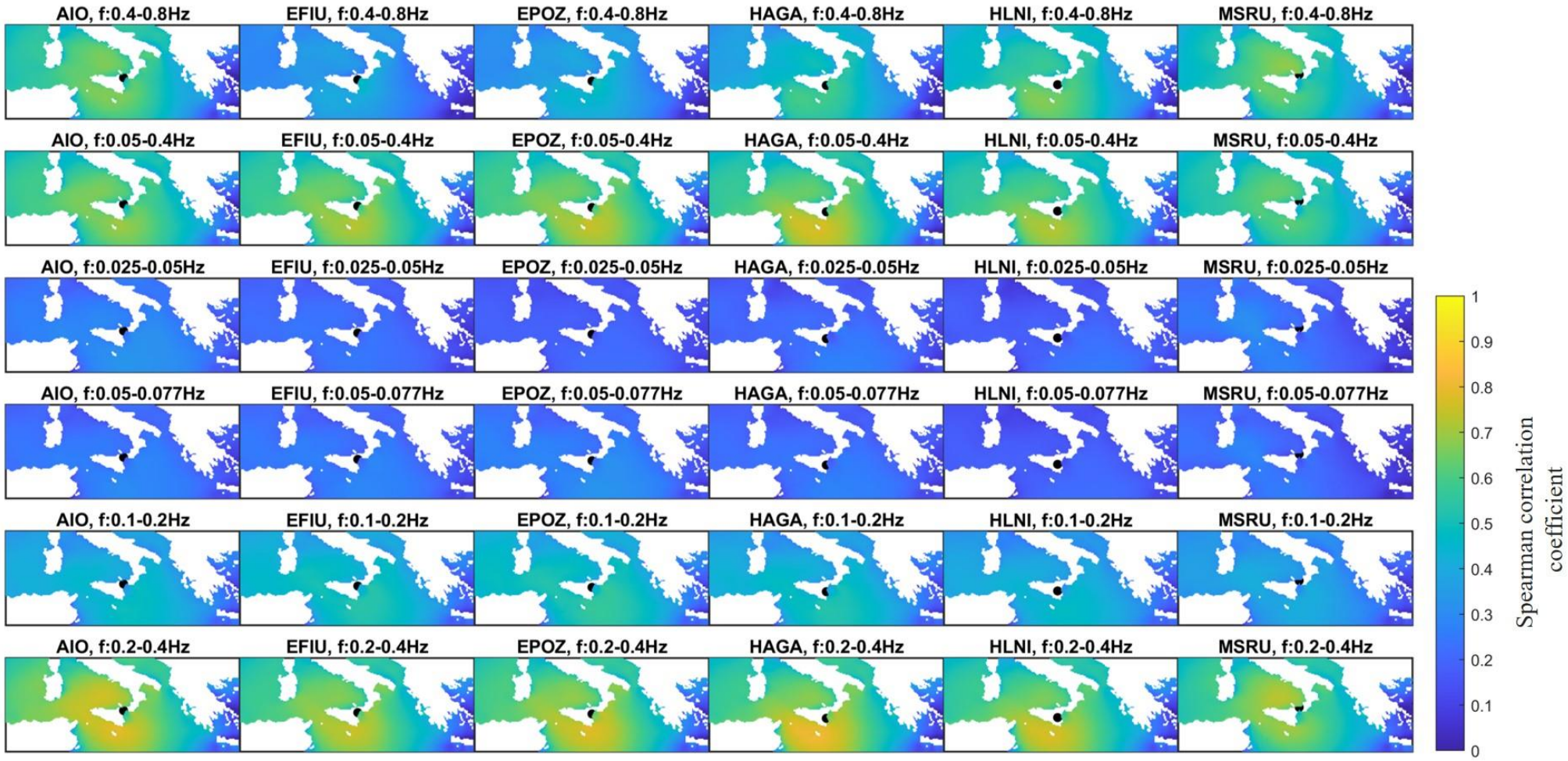

Figure 4.3: Maps of a portion of the Mediterranean Sea, showing the spatial distribution of Spearman's correlation coefficient computed between significant sea wave height and seismic RMS amplitudes, calculated on the vertical component of the six considered stations and filtered in different frequency bands (see plot titles). The black dots indicate the locations of the considered stations.

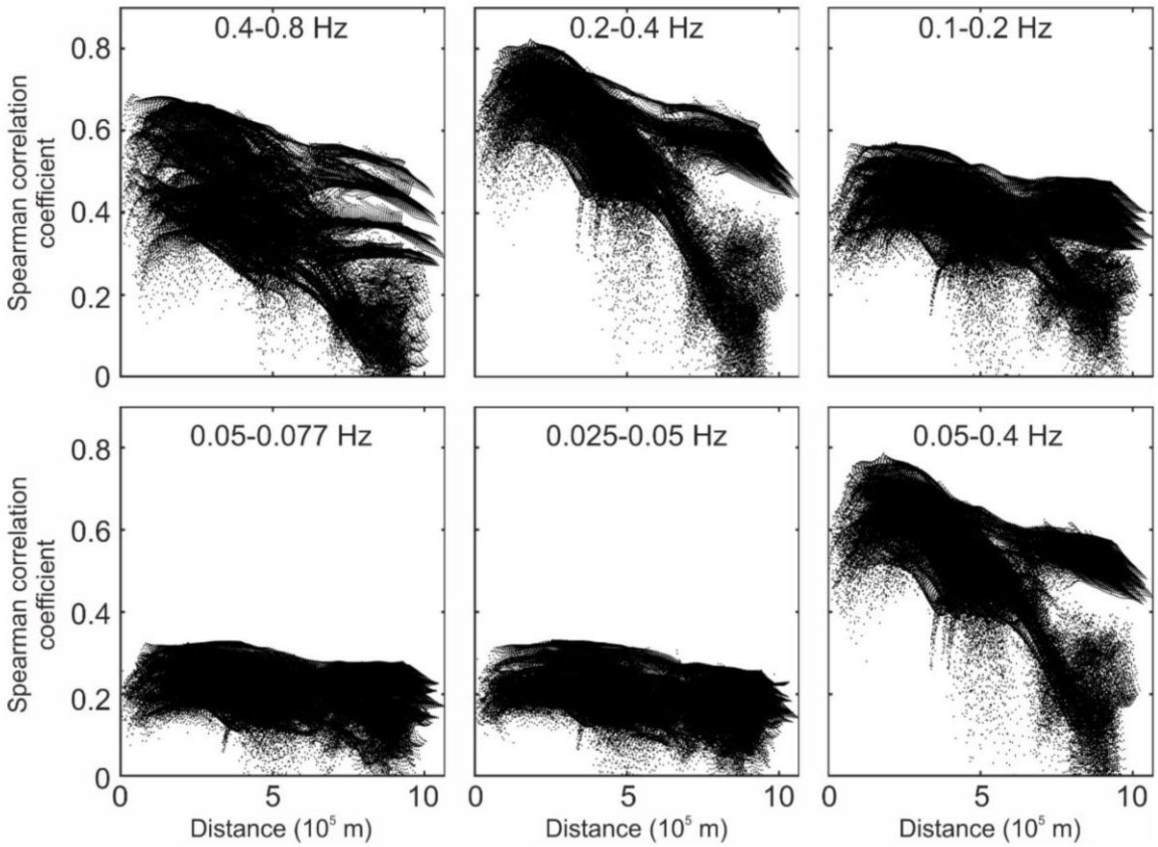

Figure 4.4: Cross-plots showing the relationship between Spearman's correlation coefficient, computed between significant sea wave height and seismic RMS amplitudes at different frequency bands (see plot titles), and the distance between the seismic station and the sea grid cell.

4.3.3 MACHINE LEARNING APPLICATIONS

In order to explore the possibility of predicting the spatial distribution of sea wave height by microseisms, the random forest (RF) technique was applied. RF is a supervised machine learning technique, capable of both classification and regression, making use of an ensemble of decision trees to allow for superior performance and lower sensitivity to over-fitting compared to single classifiers (Breiman, 2001; Kuhn & Johnson, 2013). We tried to link the microseisms recorded by different seismic stations in distinct frequency bands to the spatial distribution of significant wave height. In particular, we trained a regression model by using the seismic RMS amplitude time series as an input and the temporally corresponding hindcast maps of significant wave height as the output. To avoid overfitting problems in building the model, we applied the same method used in chapter 3.

First, to evaluate the capability of generalization of the final model (Lange et al., 1997), we deleted from the 8 year long time series (2010–2017) two consecutive months of data (November–December 2017), reserving them for the final test. Then, by using the remaining data, a k-fold cross-validation technique was applied to evaluate the spatial variability of the model predicting capability using the same methodology applied in chapter 4. The parameters computed to evaluate the predicting capability of the model were: (i) The mean absolute error (MAE) between the hindcast significant sea wave height and the predicted one; (ii) the mean absolute percentage error (MAPE), given by the MAE divided by the mean value of sea wave height in the considered sea grid cell, multiplied by 100. The maps shown in Figure 4.5 exhibit the space distribution of the average values of MAE and MAPE, obtained

by the 10-fold cross-validation technique, together with the space distribution of the mean significant wave height.

The final model was trained with the whole dataset January 2010–October 2017 and tested on the test set November–December 2017. In particular, comparisons between hindcast maps and predicted maps were carried out (three days are reported in Figure 4.7 as examples). Moreover, the cross-plots of hindcast significant wave height versus the predicted significant wave height for three grid cells located nearby the coastlines of the Sicily Channel (Figure 4.6a), Ionian Sea (Figure 4.6b), and Tyrrhenian Sea (Figure 4.6c) (see Figure 4.5 for the locations of the three grid cells) were also obtained.

Finally, the RF technique also provides an index of input importance (indicated in %), which is shown by the histograms of Figure 4.8a–c. Aggregation through summation of the input importance was also carried out to rank the microseism bands (Figure 4.8d–f) and the seismic station (Figure 4.8g–i) for significant wave height reconstruction purposes.

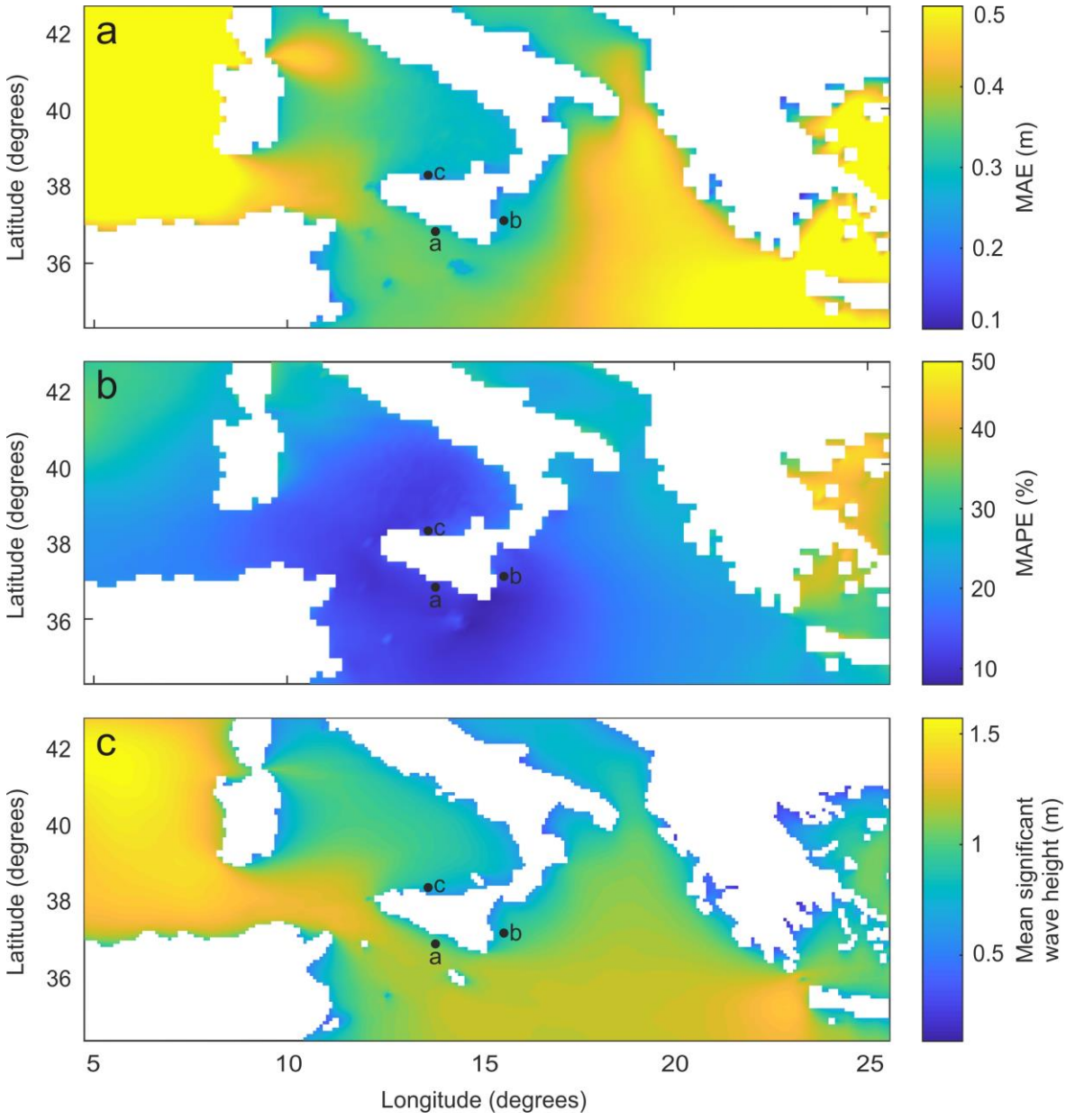

Figure 4.5: Spatial distribution of (a) mean absolute error (MAE), (b) mean absolute percentage error (MAPE), and (c) mean significant wave height computed during the interval 2010–2017. The black dots with labels “a,” “b,” and “c” indicate the locations of the grid cells, which the cross-plots in Figure 4.6 are referred to.

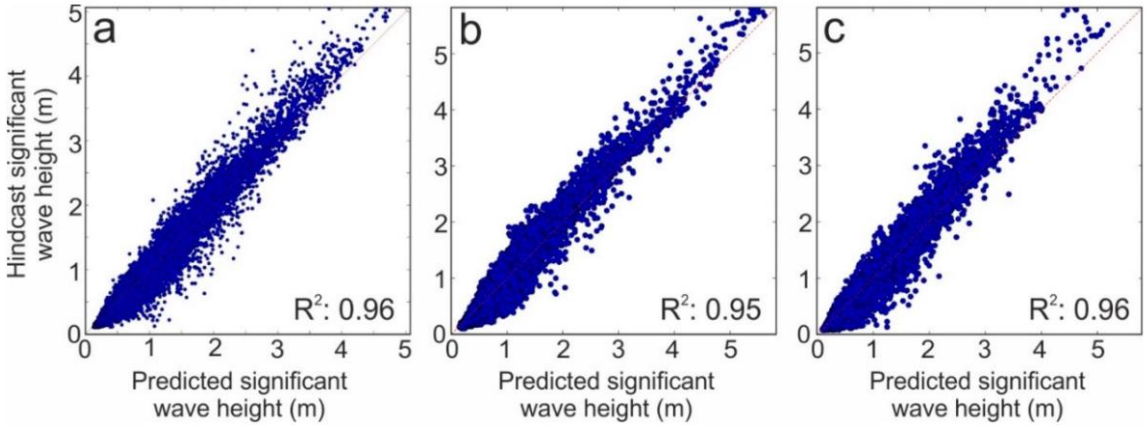

Figure 4.6: Scatter plots showing the hindcast significant wave height versus the predicted significant wave height from January 2010 to October 2017 for three grid cells, whose locations are indicated in Figure 4.5. The red dashed line in (a–c) is the y = x line. The value of the determination coefficient (R^2) is also reported in the plots.

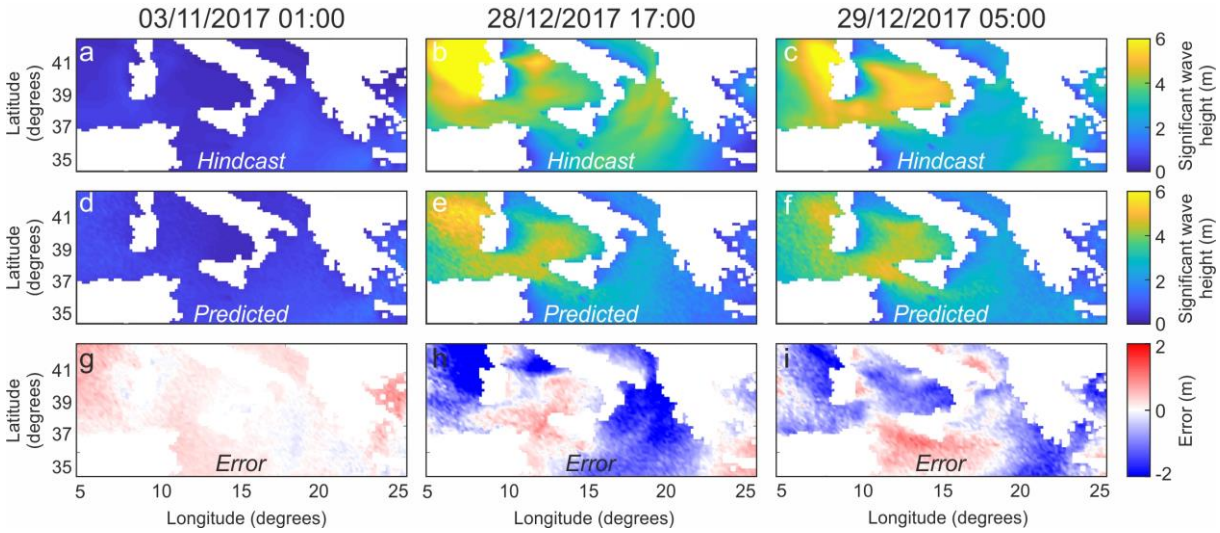

Figure 4.7: (a–c) Hindcast maps of significant sea wave height, (d–f) predicted maps of significant sea wave height based on microseism data, and (g–i) corresponding spatial distribution of the prediction error, computed as the difference between the predicted maps and the hindcast maps, during three selected time intervals: (a,d,g) 03/11/2017 01:00, (b,e,h) 28/12/2017 17:00, and (c,f,i) 29/12/2017 05:00.

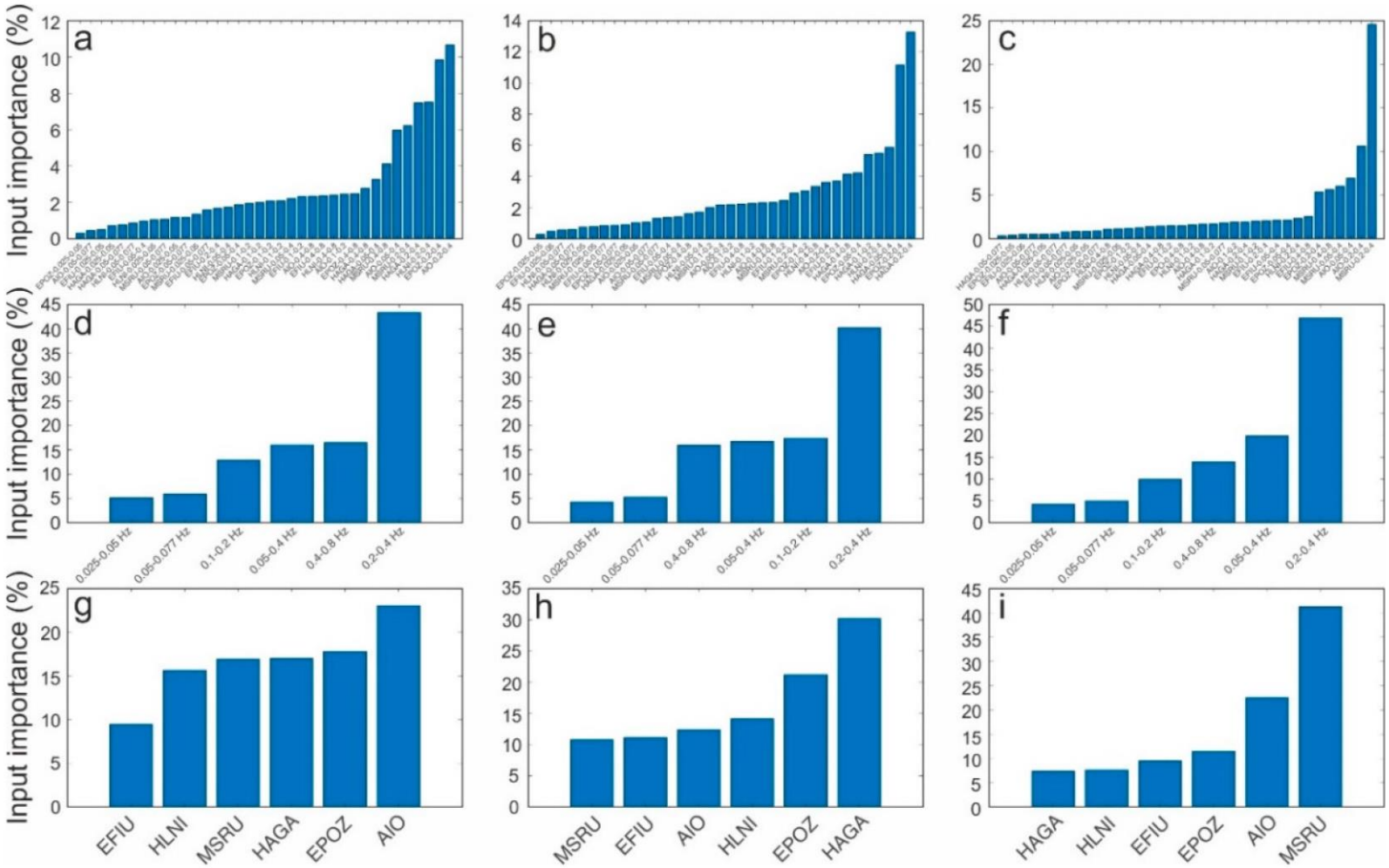

Figure 4.8: (a–c) Index of importance of all the inputs taken into account for the three grid cells labelled “a,” “b,” and “c” in Figure 4.5, respectively. (d–f) Aggregation through summation of the input importance allowing the ranking of the frequency bands for three grid cells labelled “a,” “b,” and “c” in Figure 4.5, respectively. (g–i) Aggregation through summation of the input importance allowing the ranking of seismic stations for three grid cells labelled “a,” “b,” and “c” in Figure 4.5, respectively.

4.4 RESULTS

4.4.1 RESULTS OF THE SPECTRAL AND RMS AMPLITUDE ANALYSES

As shown in results obtained in the chapter 4, microseism recorded in Eastern Sicily shows the highest spectral amplitude in the bands 2.5–5.0 and 5–10 s (SPSM and SM, respectively) at all the considered stations while the band 13–20 s (PM) exhibits a much weaker spectral amplitude (Figure 4.2).

With the exception of AIO station, high values of spectral amplitude can also be observed at frequencies above the short-period secondary microseism band (< 2.5 s). Moreover, evident amplitude seasonal modulation is shown in Figure 4.2, with maxima reached during the winter (December– February) and minima during the summer (June–August).

4.4.2 RESULTS OF THE CORRELATION ANALYSIS

As for the correlation maps, variable Spearman correlation values were obtained at the different stations and frequency bands. In particular, in agreement with Xiao et al. (2018), the highest correlation values (up to 0.8) were observed in the short-period secondary microseism band (0.2–0.4 Hz), and successively (up to 0.7) in the highest considered frequency band (0.4–0.8 Hz) (Figures 4.3 and 4.4). The primary microseism band (0.05–0.077 Hz), as well as the lower frequency band (0.025–0.05 Hz), show very low correlation values (up to 0.3). Intermediate correlation values were computed in the secondary microseism band (0.1–0.2 Hz). Finally, high values of correlation are also observed in the band comprising primary, secondary, and short-period secondary microseisms (0.05–0.4 Hz). In addition, Figure 4.4 clearly shows a

decrease in the correlation values with increasing distance from the station recording the microseism to the sea grid cell providing the significant wave height data.

Furthermore, it is evident how the spatial distribution of correlation is affected by the station location (Figure 4.3). Indeed, the correlation maps, obtained by the microseism recorded by the northernmost station (MSRU), exhibit the highest correlation values in the Tyrrhenian Sea. On the other hand, the southernmost stations (HLNI and HAGA) show the highest correlation values in both the Ionian Sea and Sicily Channel.

4.4.3 RESULTS OF MACHINE LEARNING APPLICATIONS

Regarding the machine learning analysis, as suggested by the spatial distribution of MAE and MAPE (Figure 4.5a,b), by the cross-plots of hindcast significant wave height versus the predicted significant wave height (Figure 4.6) and by the comparisons between hindcast maps and predicted maps (Figure 4.7), it is evident how the lowest values of MAE and MAPE were obtained in the Mediterranean portion closest to Sicily and then to the seismic stations providing the microseism information. In particular, the areas with low values of MAE (down to ~0.1 m) mainly coincide with the coastlines of Sicily (especially the Ionian and Tyrrhenian coastlines), as well as with the southernmost part of the Tyrrhenian Sea (Figure 4.5a), which is also characterized by very low mean values of significant wave height (Figure 4.5c). On the other hand, the areas with the lowest MAPE (minimum values equal to ~25%) comprise the southernmost part of the Tyrrhenian Sea, a part of the Ionian Sea and Sicily channel (Figure 4.5b).

The estimated aggregated importance of the different frequency bands in the three considered points nearby the Sicilian coastlines (see black dots with labels “a,” “b,” and “c” in Figure 4.3) reflects the results obtained by the correlation analysis (Figure 4.8d–f). Indeed, the frequency band with the highest importance in the sea state prediction is the one characterized by the highest Spearman correlation coefficient, that is, the short-period secondary microseism band (0.2–0.4 Hz; Figure 4.8d–f). Furthermore, the values of aggregated importance of the stations generally exhibit a dependence on the distance of the considered sea point from the station providing microseism data (Figure 4.8g–i). This is especially evident for points “b” and “c.” As for the former, the stations with the highest importance are those closest to such a point, HAGA, EPOZ, and HLNI (Figure 4.16h). Regarding point “c,” the station showing the highest importance is that closest to the Tyrrhenian coastline, MSRU (Figure 4.8i).

4.5 DISCUSSION AND CONCLUSION

Spectra of the microseism recorded by stations installed in Eastern Sicily show very high amplitudes of the short-period secondary microseism (Figure 4.2). This can be due to the fact that the considered stations are very close to the sea coastlines (maximum distance equal to ~ 19 km for HLNI). Also, the observed seasonal amplitude modulations (Figures 4.2a-f) are a common feature of the microseism recorded at temperate latitudes. Finally, clear split between the spectral peaks of short-period secondary microseisms and secondary microseisms, observed in most of the computed spectra (Fig. 4.2g-l), is a common feature for stations located nearby

the shoreline. Detailed explanations for this behaviour shown by microseism are given in chapter 3.

Spectral analysis also shows relatively high amplitudes at frequencies above the short-period secondary microseism band (Figures 4.2g–l). As indicated by the high correlation values with the sea wave height (Figures 4.3, 4.4), and explained in detail in chapter 4, such high seismic amplitudes at high frequencies (> 0.4 Hz) can be related to energy transfer from wind-driven ocean waves breaking on the shoreline.

By gathering all the information provided by the correlation maps (Figures 4.3, 4.4), it can be inferred that the microseisms recorded by the considered stations are mostly generated by sources located close to the stations, and then to the Eastern Sicilian coastlines. This was expected for the short-period secondary and partially for primary microseisms. Indeed, the former is characterized by high frequencies and then by quick attenuation with distance (e.g. Bromirski et al., 2005). As for the latter, shallow water conditions are necessary for primary microseism generation. Focusing on primary microseisms at the low frequency of 0.05 Hz and following the Airy linear wave theory, most primary microseism generations should take place at water depths less than ~150 m (e.g. Bromirski et al., 2002, 2005).

However, as also stated by Essen et al. (1999), it must be underlined that the correlation analysis (i) shows only the dominant source areas over the entire analyzed time period 2010–2017, (ii) has no temporal resolution, and (iii) is characterized by a limited space resolution of 1/24°. Hence, we cannot rule out the possibility that part of the recorded microseism is generated in open sea far away from the Sicilian coasts. This could be the case for part of the secondary microseism, whose source can partially be associated with wave–wave interaction mechanisms in the deep ocean

(e.g. Cessaro et al., 1994; Chevrot et al., 2007), assuming that losses from both scattering and transfer to the solid earth are not too effective (Bromirski et al., 2013). In addition, Gualtieri et al. (2019) have recently shown that primary microseisms at low frequency (~ 0.05 Hz) can have source regions located nearby coastlines thousands of kilometers away from the seismic stations.

Machine learning analysis shows promising results in the spatial reconstruction of sea state in terms of significant wave height by microseism. Indeed, low values of both MAE (down to ~0.1 m) and MAPE (down to ~25%) were computed in a portion of the Mediterranean Sea close to Sicily and then to the seismic stations providing microseism information (Figure 4.5a,b).

The underestimation of the predicted sea wave heights for values above 3 - 4 m (Fig. 4.6) is due to the fact that the hindcast dataset used by ML algorithm contain more number of samples characterized by an average sea wave heights values (1 -3 m) and a poor number of samples with heights higher than 4 -5 m.

This finding, together with the observed dependence of both Spearman's correlation coefficient (Figure 4.4) and input importance (Figure 4.8g,h,i) from the distance sea grid cell-seismic station, suggests that a reliable reconstruction of the sea state in a wide area needs microseism data recorded by dense networks, comprising stations evenly distributed along the coastlines.

5. CONCLUSIONS

5.1 GENERAL SUMMARY

Microseism is considered as the most continuous and ubiquitous seismic signal on Earth, generated by ocean wave energy coupling with the Earth's ground (Longuet-Higgins, 1950; Hasselmann, 1963; Ardhuin et al., 2015).

Understanding the relationship between microseism and sea state was the main goal of the conducted research. To this end, seismic signals recorded in two different areas, Antarctica and Eastern Sicily, have been investigated. In particular, the first study carried out in Antarctica aimed to quantitatively investigate the relationship between microseism recorded along the coasts of Antarctica and the sea ice concentration in the whole Southern Ocean, while the second study focused on the Eastern Sicily and the main purpose was to explore the relationship between microseism recorded along the eastern coast of Sicily and the sea wave height data, provided by both buoys in Ionian and Tyrrhenian Seas and hindcast maps of Mediterranean Sea. The first study was performed to highlight the effect of sea ice on microseism and to try to spatially and temporally reconstruct the sea ice concentration starting from microseism recordings; the second study, instead, concerned the relations between microseism and sea wave height measured in Mediterranean Sea (in particular Ionian and Tyrrhenian Seas), aimed at reconstructing the spatial and temporal variability of sea wave height starting from microseism.

In both studies, the three main bands of microseism have been analysed separately: 2.5-5.0 s (Short Period Secondary Microseism, SPSM), 5-10 s (Secondary Microseism, SM) and 13-20 s (Primary Microseism, PM). Concerning the microseism characterization, most of its energy turned out to be contained in the

SPSM and SM bands, while the band 13–20 s (PM) showed a much weaker energy. Microseism exhibited seasonal variability in both Eastern Sicily and Antarctica. In the former area, the amplitude maxima were observed during December–February (winter) and the minima during June–August (summer), as expected in all the temperate latitude regions. Conversely, in Antarctica, because of the sea ice, a different pattern was identified with the maxima during February–April (austral fall) and the minima during October–December (austral spring-summer).

The microseism bands, showing the strongest correlation with sea state is the SPSM. Clear anti-correlation patterns between SPSM microseism and sea ice concentration were found in Antarctica, while high correlation was observed between SPSM microseism and sea wave height in Eastern Sicily.

The innovative methods, based on up-to-date machine learning techniques, were able to reconstruct the spatial-temporal distribution of sea state with fairly low error. Combining the results coming from the correlation maps and the machine learning techniques, we showed a decrease in sea state sensitivity of microseism, due to the increasing distance from the station recording the seismic signal. The closer the station, the more information concerning the state of the nearby sea area are contained in the recorded microseism.

5.2 SUGGESTED FUTURE WORK AND PERSPECTIVES

Considering the studies performed in the framework of this PhD, future works could aim at extending the dataset, as well as improving the methods of monitoring of sea state by microseism.

A future development of this study will be the inclusion of time series of horizontal seismic component amplitudes and the wavefield features as input in the machine learning modelling.

Moreover, as pointed out by Ferretti et al. (2018), the teleseisms (i.e., earthquakes localized at a distance greater than 1000 km from the recording stations) with frequency content that is similar to that of microseism recordings could affect the results of the predicted sea ice concentration. Therefore, a future development of this study will be the removal of the time windows that include teleseism signals.

Focusing on the machine learning analyses, once the regression models have been determined and if the seismic data are available, such models will allow to reconstruct the sea ice concentration in Antarctica and the sea wave height in the Ionian and Tyrrhenian Seas in periods when satellite images and buoys data, in the former and latter cases respectively, are not available, with wide applications in many fields, first of all the climate studies. The inclusion of the afore mentioned new inputs will likely improve the capability to reconstruct the spatial-temporal distribution of sea ice concentration and sea wave height. Finally, the finding on importance of the distance between seismic station and coastline is important to build a reliable experimental monitoring system of the sea conditions based on microseism recordings. Hence, microseism can become a very useful tool for sea state monitoring, even as effective as other instruments routinely used for that purpose. Indeed, seismic stations have

lower costs of both installation and maintenance with respect to wave buoys. In addition, seismic signals are acquired continuously with a sampling rate from tens to hundreds of Hz, and then have a much higher temporal resolution than radar altimeters and synthetic aperture radar. As for the spatial resolution, it depends on the number of the stations, installed nearby the coastlines. Finally, a further advantage of the seismic signals is represented by the fact that, in most cases, it is not necessary to install a seismic network specifically for microseism monitoring, but it is possible to make use of the broadband stations today installed worldwide to monitor seismic and volcanic activities.

Finally, in the future, efforts will be taken in the reconstruction of wave directional information, which is essential in the overall description of the sea state.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the facilities of IRIS (Incorporated Research Institutions for Seismology) Data Services, and the IRIS Data Management Center, for the access to waveforms of 20 seismic stations, the NSIDC (National Snow & Ice Data Center) for the information about temporal and spatial variability of sea ice concentration in the Southern Ocean, the technicians of INGV (Istituto Nazionale di Geofisica e Vulcanologia) - Osservatorio Etneo for enabling the acquisition of seismic data, the ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) for providing the buoy data and the E.U. Copernicus Marine Service Information for the bathymetric data and the hindcast maps of significant sea height. My thanks also go to the developers of the software used in this thesis: the Antarctic Mapping Tools for MATLAB, Matplotlib package for Python and the Seizmo – Passive Seismology Toolbox for MATLAB.

I would like to thank my supervisor Professor Stefano Gresta for giving me the opportunity to approach the scientific world and to my tutors, Professor Andrea Cannata, Dr. Giuseppe Di Grazia and Dr. Flavio Cannavò, for their essential contribution in the development of this thesis. I especially thank Professor Andrea Cannata for the continuous support and for his patience and motivation, his guidance helped me in all the time of research.

Finally, I would like to thank you all the people who believed in me and gave me the moral support to carry out this thesis.

REFERENCES

- Almendros, J., Ibáñez, J. M., Alguacil, G., and Del Pezzo, E., 2002. *Array analysis using circular-wave-front geometry: an application to locate the nearby seismovolcanic source*. Geophys. J. Int. 136, 159–170. DOI: 10.1046/j.1365-246x.1999.00699.x.
- Altman, N. S., 1992. *An introduction to kernel and nearest-neighbor nonparametric regression.* Am. Stat. 46, 175–185. DOI: 10.1080/00031305.1992. 10475879.
- Aminzadeh, F. and de Groot, P., 2006., *Neural Networks and Other Soft Computing Techniques with Applications in the Oil Industry*. EAGE Publications.
- Angus, J. E. 1994., *The probability integral transform and related results*. SIAM review 36(4), 652–654.
- Anthony, R. E., Aster, R.C., Wiens, D., Nyblade, A., Anandakrishnan, S., Huerta, A., Winberry, P., Wilson, T., Rowe, C. 2015., *The seismic noise environment of Antarctica*. Seismol. Res. Lett. 86, 89–100, DOI:https://DOI.org/10.1785/0220140109.
- Anthony, R.E.; Aster, R.C.; McGrath, D. 2017., *Links between atmosphere, ocean, and cryosphere from two decades of microseism observations on the Antarctic Peninsula*. J. Geophys. Res. Earth Surf., 122, 153–166.
- Ardhuin, F., Stutzmann, E., Schimmel, M., Mangeney, A., 2011. *Ocean wave sources of seismic noise.* J. Geophys. Res. Space Phys., 116, 1–21.

- Ardhuin, F., Balanche, A., Stutzmann, E., and Obrebski, M., 2012. *From seismic noise to ocean wave parameters: general methods and validation*. J. Geophys. Res. 117:C05002. DOI: 10.1029/2011JC007449.
- Ardhuin, F., Gualtieri, L., and Stutzmann, E., 2015. *How ocean waves rock the Earth: two mechanisms explain microseisms with periods 3 to 300 s.* Geophys. Res. Lett. 42, 765–772. DOI: 10.1002/2014GL062782.
- Arndt, J. E., Schenke, H.S., Jakobsson, M., Nitsche, F.O., Buys, G., Goleby, B., Rebesco, M., Bohoyo, F., Hong, J., Black, J., Greku, R., Udintsev, G., Barrios, F., Reynoso-Peralta, W., Taisei, M., Wigley, R. 2013., *The International Bathymetric Chart of the Southern Ocean (IBCSO) Version 1.0 A new bathymetric compilation covering circum-Antarctic waters*. Geophys. Res. Lett. 40(9), 1–7. DOI: https://DOI.org/10.1002/grl.50413.
- Asten, M.W., 1978. *Geological control on the three-component spectra of Rayleigh-wave microseisms*. Bulletin of the Seismological Society of America ; 68 (6): 1623–1636.
- Asten, M. W., & Henstridge, J. D., 1984. *Array estimators and the use of microseisms for reconnaissance of sedimentary basins*. GEOPHYSICS 49: 1828-1837. DOI: https://doi.org/10.1190/1.1441596.
- Aster, R. C., McNamara, D. E., and Bromirski, P. D. 2008., *Multidecadal climate induced variability in microseisms.* Seismol. Res. Lett. 79, 194–202. DOI: 10.1785/ gssrl.79.2.194.
- Aster, R. C., McNamara, D. E., and Bromirski, P. D. 2010., *Global trends in extremal microseism intensity*. Geophys. Res. Lett. 37:L14303. DOI: 10.1029/ 2010GL043472.

- Barzilai, A., VanZandt, T., Kenny, T., 1998. *Technique for measurement of the noise of a sensor in the presence of large background signals*. Review of Scientific Instruments 69, 2767 (1998). DOI: https://doi.org/10.1063/1.1149013.
- Battaglia, J., & Aki, K., 2003. *Location of seismic events and eruptive fissures on the Piton de la Fournaise volcano using seismic amplitudes*. J. Geophys. Res., 108(B8), 2364, DOI:10.1029/2002JB002193.
- Bencivenga, M., Nardone, G., Ruggiero, F., and Calore, D., 2012. *The Italian data buoy network (RON)*. Adv. Fluid Mech. 74, 321–332. DOI: 10.2495/AFM12 0291.
- Bergen, K.J., Chen, T., Li, Z., 2019., *Preface to the Focus Section on Machine Learning in Seismology*. Seismol. Res. Lett. DOI: 10.1785/0220190018.
- Bertelli, T., 1872. *Osservazioni sui Piccolo Movimenti dei Pendoli in Relazione ad Alcuni Fenomeni Meteorologiche*. Bullettino Meteorologico Dell'Osservatorio Dell Collegio Romano, 101.
- Bertelli, T., 1875. *Della Realta' dei Moti Microsismici ed Osservazioni sui Medesimi fatte nell'anno 1873–1874 nel Collegio alla Querce Presso Firenze.* Atti dell'Accademia Pontificia de' nuovi Lincei, Tomo XXVIII – Anno XXVIII, 334–375.
- Bertelli, T., 1878. *Riassunto delle Osservazioni Microsismiche fatte nei Collegio alla Querce di Firenze*. Atti dell'Accademia Pontificia de' nuovi Lincei, Pt. III (Anno XXI, Sessione 2a), 193–243.

- Behrens, A., 2013. *Documentation of a web based source code library for WAM*; https://github.com/mywave/WAM/tree/master/documentation (accessed on 24 February 2020).
- Bishop, C., 2006. *Pattern recognition and machine learning*. Springer.
- Bregman, Y. & Rabin, N., 2019. *Aftershock identification using diffusion maps*. Seismol. Res.Lett. DOI: 10.1785/0220180291.
- Breiman, L., 2001. *Random forests*. Mach. Learn. 45, 5–32. DOI: 10.1023/A: 1010933404324.
- Brenguier, F, Campillo, M, Hadziioannou, C, Shapiro, N.M., Nadeau, R.M., Larose, E., 2008. *Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations*. Science 321, 1478–1481, https://DOI.org/10.1126/science.1160943.
- Bromirski, P. D., Flick, R. E., and Graham, N., 1999. *Ocean wave height determined from inland seismometer data: implications for investigating wave climate changes in the NE Pacific*. J. Geophys. Res. 104, 20753–20766. DOI: 10.1029/1999JC900156.
- Bromirski, P.D., 2001. *Vibrations from the "Perfect Storm"*. Geochem. Geophys. Geosyst., 2, 2.
- Bromirski, P. D., and Duennebier, F. K., 2002. *The near-coastal microseism spectrum: spatial and temporal wave climate relationships*. J. Geophys. Res. 107:2166. DOI: 10.1029/2001JB000265.
- Bromirski, P. D., Duennebier, F. K., and Stephen, R. A. 2005., *Midocean microseisms*. Geochem. Geophys. Geosyst. 6:Q04009. DOI: 10.1029/ 2004GC000768.

- Bromirski, P. D., Stephen, R. A., and Gerstoft, P., 2013. *Are deep-ocean-generated surface-wave microseisms observed on land?* J. Geophys. Res. Solid Earth 118, 3610–3629. DOI: 10.1002/jgrb.50268.
- Brooks, L. A., Townend, J., Gerstoft, P., Bannister, S., and Carter, L., 2009. *Fundamental and higher-mode Rayleigh wave characteristics of ambient seismic noise in New Zealand*. Geophys. Res. Lett. 36:L23303. DOI: 10.1029/2009GL040434.
- Brzak, K., Gu, Y.G., Ökeler, A., Steckler, M, Lerner-Lam, A., 2009. *Migration imaging and forward modeling of microseismic noise sources near southern Italy*. Geoch., Geophy., Geosys. Volume 10, Issue 1. DOI: https://doi.org/10.1029/2008GC002234.
- Burley, N. & Suddeth, R., 2013. *Wave climate and energy dissipation near Santa Cruz Island, California*. Center for Watershed Sciences. https://watershed.ucdavis.edu/education/classes/eastern-santa-cruz-island/pages/flogs.
- Caliński, T. & Harabasz, J., 1974. *A dendrite method for cluster analysis*. Communications in Statistics-theory and Method 3(1), 1–27.
- Campillo, M., Paul, A., 2003. *Long-Range Correlations in the Diffuse Seismic Coda*. Science 299, 547 (2003). DOI: 10.1126/science.1078551.
- Cannata, A., Di Grazia, G., Montalto, P., Ferrari, F., Nunnari, G., Patanè, D., 2010. *New insights into banded tremor from the 2008–2009 Mount Etna eruption.* J. Geophys. Res. 115:B12318. DOI: 10.1029/2009JB007120.
- Cannata, A., Larocca, G., Del Carlo, P., Giudice, G., Giuffrida, G., Liuzzo, M., Zuccarello, L., Di Grazia, G., Gambino, S., Privitera, E., Delladio, A.,

Grigioni, P., 2017a. *Characterization of seismic signals recorded in Tethys Bay, Victoria Land (Antarctica): data from atmosphere-cryosphere-hydrosphere interaction*. Annals of Geophysics, 60, 5, 2017; S0555; DOI: 10.4401/ag-7408.

- Cannata, A., Cannavò, F., Montalto, P., Ercoli, M., Mancinelli, P., Pauselli, C., Leto, G., 2017b. *Monitoring crustal changes at volcanoes by seismic noise interferometry: Mt. Etna case of study*. J. Volc. Geoth. Res. 337, 165–174, https://DOI.org/10.1016/j.jvolgeores.2017.03.023.
- Cannata, A., Cannavò, F., Moschella, S., Gresta, S., and Spina, L., 2019. *Exploring the link between microseism and sea ice in Antarctica by using machine learning*. Sci. Rep. 9:13050. DOI: 10.1038/s41598-019-49586-z.
- Cannata, A., Cannavò, F., Moschella, S., Di Grazia, G., Nardone, G., Orasi, A., Picone, M., Ferla, M., Gresta, S., 2020. *Unravelling the Relationship Between Microseisms and Spatial Distribution of Sea Wave Height by Statistical and Machine Learning Approaches*. Remote Sens. 2020, 12, 761; DOI:10.3390/rs12050761.
- Cannavò, F., Cannata, A., Cassisi, C., Di Grazia, G., Montalto, P., Prestifilippo, M., Privitera, E., Coltelli, M., Gambino, S., 2017. *A multivariate probabilistic graphical model for real-time volcano monitoring on Mount Etna.* J. Geophys. Res. Solid Earth 122, https://DOI.org/10.1002/2016JB013512.
- Capon, J., 1973. *Signal processing and frequency-wavenumber spectrum analysis for a large aperture seismic array*. Methods Comput. Phys. 13, 1–59. DOI: 10. 1016/b978-0-12-460813-9.50007-2.

- Carranza, E.J.M.; Laborte, A.G., 2015. *Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines)*. Comput. Geosci., 74, 60–70.
- Cessaro, R.K., Chan, W.W., 1989. *Wide-angle triangulation array study of simultaneous primary microseism sources.* J. Geophy. Res.: Solid EarthVolume 94, Issue B11. DOI: https://doi.org/10.1029/JB094iB11p15555.
- Cessaro, R.K., 1994. *Sources of primary and secondary microseisms*. Bull. Seismol. Soc. Am., 84, 142–148.
- Chen, Y.N., Gung, Y., You, S.H., Hung, S.H., Chiao, L.Y., Huang, T.Y., 2011. *Characteristics of short period secondary microseisms (SPSM) in Taiwan: the influence of shallow ocean strait on SPSM*. Geophys. Res. Lett. 38:L04305. DOI: 10.1029/2010GL046290.
- Chen, Y., 2018. *Automatic microseismic event picking via unsupervised machine learning*, Geophys. J. Int. 212, 88–102.
- Chen, Y, Zhang, M, Bai, M., Chen, W., 2019. *Improving the Signal-to-Noise Ratio of Seismological Datasets by Unsupervised Machine Learning*. Seismol. Res. Lett.Volume 90, Number 4.
- Chevrot, S., Sylvander, M., Benahmed, S., Ponsolles, C., Lefevre, J. M., and Paradis, D., 2007. *Source locations of secondary microseisms in western Europe: evidence for both coastal and pelagic sources*. J. Geophys. Res. 112:B11301. DOI: 10.1029/2007JB005059.
- Craig, D., Bean, C., Lokmer, I., and Möllhoff, M., 2016. *Correlation of wavefieldseparated ocean-generated microseisms with North Atlantic Source regions*. Bull. Seism. Soc. Am. 106, 1002–1010. DOI: 10.1785/0120150181.

- Curtis, A., Gersoft, P., Sato, H., Snieder, R., Wapenaar, K., 2006. *Seismic interferometry—turning noise into signal*. The Leading Edge Volume 25, Issue 9. DOI: https://doi.org/10.1190/1.2349814.
- De Caro, M., Monna, S., Frugoni, F., Beranzoli, L., and Favali, P., 2014. *Seafloor seismic noise at Central Eastern Mediterranean sites*. Seismol. Res. Lett. 85, 1019–1033. DOI: 10.1785/0220130203.
- De Plaen, RSM, Cannata, A., Cannavo', F., Caudron, C, Lecocq T., Francis, O., 2019. *Temporal Changes of Seismic Velocity Caused by Volcanic Activity at Mt. Etna Revealed by the Autocorrelation of Ambient Seismic Noise*. Front. Earth Sci. 6:251. DOI: 10.3389/feart.2018.00251.
- Dell'Aversana, P., 2014. *Integrated Geophysical Models: Combining Rock Physics with Seismic, Electromagnetic and Gravity Data*. EAGE Publications.
- Dell'Aversana, P., Ciurlo, B., Colombo, S., 2018. *Integrated Geophysics and Machine Learning for Risk Mitigation in Exploration Geosciences*. EAGE Conference 2018.
- Dodge, Y., 2006. *The Oxford Dictionary of Statistical Terms* (OUP, Oxford, ed. 6,2006).
- Drucker, H., Burges, C., Kaufman, L., Smola, A., and Vapnik, V., 1997. *Support vector regression machines*. Adv. Neural Inform. Proc. Syst. 28, 779–784.
- Ebeling, C.W., 2012. *Inferring Ocean Storm Characteristics from Ambient Seismic Noise: A Historical Perspective*. Advances in Geophysics, Volume 53. DOI: http://dx.doi.org/10.1016/B978-0-12-380938-4.00001-X.

- Efron, B., 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia, PA: Soc. for Ind. and Appl. Math.
- EMODnet Bathymetry Consortium., 2018. *EMODnet Digital Bathymetry (DTM 2018)*, in EMODnet Bathymetry Consortium. Available at: https://DOI.org/10. 12770/18ff0d48-b203-4a65-94a9-5fd8b0ec35f6.
- Essen, H.-H.; Klussmann, J.; Herber, R.; Grevemeyer, I., 1999. *Does microseisms in Hamburg (Germany) reflect the wave climate in the North Atlantic?* Spiegelt die in Hamburg gemessene Mikroseismik das Wellenklima im Nordatlantik wider? Dtsch. Hydrogr. Zeitschrift, 51, 33–45.
- Essen, H.-H., Krüger, F., Dahm, T., and Grevemeyer, I., 2003. *On the generation of secondary microseisms observed in northern and central Europe*. J. Geophys. Res. Space Phys. 2003, 1–15. DOI: 10.1029/2002JB002338.
- Ferretti, G., Zunino, A., Scafidi, D., Barani, S., and Spallarossa, D., 2013. *On microseisms recorded near the Ligurian coast (Italy) and their relationship with sea wave height*. Geophys. J. Int. 194, 524–533. DOI: 10.1093/gji/ggt114. Ferretti, G.; Scafidi, D.; Cutroneo, L.; Gallino, S.; Capello, M., 2016. *Applicability of an empirical law to predict significant sea-wave heights from microseisms along the Western Ligurian Coast (Italy).* Cont. Shelf Res., 122, 36–42.7.
- Ferretti, G., Barani, S., Scafidi, D., Capello, M., Cutroneo, L., Vagge, G., 2018. *Near real-time monitoring of significant sea wave height through microseism recordings: an application in the Ligurian Sea (Italy).* Ocean Coast.Manag. 165, 185–194. DOI: 10.1016/j.ocecoaman.2018.08.023.

- Fetterer, F., Knowles, K., Meier, W. N., Savoie, M. & Windnagel, A. K., 2017. *Sea Ice Index, Version 3. [1993–2017]*. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center, https://DOI.org/10.7265/N5K072F8 (2017, updated daily).
- Friedrich, A., Krüger, F., Klinge, K., 1998. *Ocean-generated microseismic noise located with the Gräfenberg array*. Journal of Seismology volume 2, pages47–64(1998).
- Fu, L.L.; Cazenave, A., 2000. *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*, Satell. Elsevier: Amsterdam, The Netherlands.
- Gal, M.; Reading, A.M.; Ellingsen, S.; Gualtieri, L.; Koper, K.D.; Burlacu, R.; Tkalčić, H.; Hemer, M., 2015. *The frequency dependence and locations of short-period microseisms generated in the Southern Ocean and West Pacific*. J. Geophys. Res. Solid Earth, 120, 5764–5781.
- Geurts, P., Ernst, D. & Wehenkel, L., 2006. *Extremely randomized trees*. Machine learning 63(1), 3–42.
- Gilmore, M., 1946. *Microseisms and ocean storms*. Bulletin of the Seismological Society of America, 36, 89–119.
- Gilmore, M., Hubert, W., 1948. *Microseisms and Pacific typoons*. Bulletin of the Seismological Society of America, 38, 195–228.
- Greene, C. A., Gwyther, D. E. & Blankenship, D. D., 2017. *Antarctic Mapping Tools for Matlab.* Computers & Geosciences 104, 151–157, https://DOI.org/10.1016/j.cageo.2016.08.003.

- Grevemeyer, I., Herber, R., and Essen, H., 2000. *Microseismological evidence for a changing wave climate in the northeast Atlantic Ocean.* Nature 408, 349–352. DOI: 10.1038/35042558.
- Grob, M.; Maggi, A.; Stutzmann, E., 2011. *Observations of the seasonality of the Antarctic microseismic signal, and its association to sea ice variability.* Geophys. Res. Lett. 2011, 38, 38.
- Gu, Y.G., Dublanko, C., Lerner-Lam, A., Brzak, M.S., 2007. *Probing the sources of ambient seismic noise near the coasts of southern Italy*. Geophy. Res. Lett. Volume 34, Issue 22. DOI: https://doi.org/10.1029/2007GL031967.
- Gualtieri, L., Stutzmann, E., Capdeville, Y., Ardhuin, F., Schimmel, M., Mangeney, A., 2013. *Modelling secondary microseismic noise by normal mode summation*. Geophys. J. Int. 193, 1732–1745. DOI: 10.1093/gji/ggt090.
- Gualtieri, L., Stutzmann, E., Capdeville, Y., Farra, V., Mangeney, A., and Morelli, A., 2015. *On the shaping factors of the secondary microseismic wavefield*. J. Geophys. Res. 120, 6241–6262. DOI: 10.1002/2015jb012157.
- Gualtieri, L., Stutzmann, E., Juretzek, C.,Hadziioannou, C., and Ardhuin, F., 2019. *Global scale analysis and modelling of primary microseisms.* Geophys. J. Int. 218, 560–572. DOI: 10.1093/gji/ggz161.
- Guo, Z., Xue, M., Aydin, A., Ma, Z., 2020. *Exploring source regions of single- and double-frequency microseisms recorded in eastern North American margin (ENAM) by cross-correlation*. Geophys. J. Int. (2020) 220, 1352–1367. DOI: 10.1093/gji/ggz470.
- Gutenberg, B., 1947. *Microseisms and weather forecasting*. Journal of Meteorology, 4, 21–28.

- Gutenberg, B., 1958. *Two types of microseisms*. J. Geophys. Res. 63(3),595–597.
- Hasselmann, K. A., 1963. *Statistical analysis of the generation of microseisms*. Rev. Geophys. Space Phys. 1, 177–210. DOI: 10.1029/RG001i002p00177.
- Haubrich, R. A., & McCamy, K., 1969. *Microseisms: coastal and pelagic sources*. Rev. Geophys. Space Phys. 7, 539–571. DOI: 10.2183/pjab.93.026.
- Havskov, J. & Alguacil, G., 2016. *Instrumentation in Earthquake Seismology*. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-21314-9.
- Hirn, A., Nercessian, A., Sapin, M., Ferrucci, F., andWittlinger, G., 1991. *Seismic heterogeneity of Mt. Etna: structure and activity*. Geophys. J. Int. 105, 139–153.DOI: 10.1111/j.1365-246x.1991.tb03450.x.
- Hjortenberg, E., Nikolaev, A. N., 1990. *Microseisms caused by factors internal and external to the earth's crust*. Phys. Earth Planet. Interiors 63,145 –283.
- Ho, T., 1995. *Random Decision Forests*. Proceedings of the Third Int'l Conf.Document Analysis and Recognition, Montreal, 278–282.
- Ho, T., 1998. *The random subspace method for constructing decision forests*. IEEE Trans. Pattern Anal. Mach. Intell. 13, 340–354.
- Holland, P. R., 2014. *The seasonality of Antarctic sea ice trends*. Geophys. Res. Lett. 41, 4230–4237, https://DOI.org/10.1002/2014GL060172.
- Holthuijsen, L.H., 2007. *Waves in Oceanic and Coastal Waters.* Cambridge University Press, 2007.

- Hong, T.K., Lee, J., Lee, G., Lee. J., Park, S., 2020. *Correlation between Ambient Seismic Noises and Economic Growth*. Seismol. Res. Lett. XX, 1–12. DOI: 10.1785/0220190369.
- Hunter, J. D., 2007. *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering 9, 90–95, https://DOI.org/10.1109/MCSE.2007.55.
- Jia, K., Liu, J., Tu, Y., Li, Q., Sun, Z., Wei, X., 2019. *Land use and land cover classification using Chinese GF-2 multispectral data in a region of the North China Plain.* Front. Earth Sci. 13, 327–335. DOI: 10.1007/s11707-018-0734-8.
- Jiao, P., and Alavi, A. H., 2019. *Artificial intelligence in seismology: advent, performance and future trends*. Geosci. Front. 19:30198. DOI: 10.1016/j.gsf.2019.10.004.
- Jordan, M.I., Mitchell, T.M., 2015. *Machine learning: Trends, perspectives, and prospects*. Science, 349, no. 6245, pp. 255-260.
- 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, https://DOI.org/10.1002/2016JB013017.
- Juretzek, C., & Hadziioannou, C., 2017. *Linking source region and ocean wave parameters with the observed primary microseismic noise*. Geophys. J. Int. 211,1640–1654. DOI: 10.1093/gji/ggx388.
- Jurkevics, A., 1988. *Polarization analysis of three-component array data*. Bull. Seismol. Soc. Am 78, 1725–1743.
- Kanji, J. K., 2006. *100 Statistical Tests*. London, Sage Publications, 2006.

- Kästle, E. D., El-Sharkawy, A., Boschi, L.,Meier, T., Rosenberg, C., Bellahsen, N., Weidle, C., 2018. *Surface wave tomography of the Alps using ambient-noise and earthquake phase velocity measurements*. J. Geoph. Res.: Solid Earth, 123, 1770–1792.https://DOI.org/10.1002/2017JB014698.
- Kedar, S., Longuet-Higgins, M. S., Graham, F. W. N., Clayton, R., Jones, C., 2008. *The origin of deep ocean microseisms in the North Atlantic Ocean*. Proc. R. Soc. A 464, 777–793. DOI: 10.1098/rspa.2007.0277.
- Keller, C.A., Evans, M., 2019. *Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10*. Geosci. Model Dev., 12, 1209–1225.
- Kimman, W. P., Campman, X., and Trampert, J., 2012. *Characteristics of seismic noise: fundamental and higher mode energy observed in the Northeast of the Netherlands*. Bull. Seismol. Soc. Am. 102, 1388–1399. DOI: 10.1785/0120110069.
- Kirkwood, C., Cave, M., Beamish, D., Grebby, S., and Ferreira, A., 2016. *A machine learning approach to geochemical mapping*. J. Geochem. Explorat. 167,49–61. DOI: 10.1016/j.gexplo.2016.05.003.

  Koch, F., Wiens, D.A., Euler, G.G., Aster, R.C., Nyblade, A., Anandakrishnan, S., Huerta, A.D., Wilson, T.J., 2011. *Tracking the effect of sea ice cover on microseismic noise using two seismic arrays in Antarctica.* Abstract S31C-2247, presented at 2013 Fall Meeting, AGU, San Francisco, Calif., 5–9 Dec (2011).
- Kong, Q., Allen, R. M., Schreier, L., and Kwon, Y.-W., 2016. *MyShake: a smartphone seismic network for earthquake early warning and beyond*. Sci.

Adv. 2:e1501055. DOI: 10.1126/sciadv.1501055.

- Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., and Gerstoft, P., 2018. *Machine learning in seismology: turning data into insights.* Seismol. Res. Lett. 90, 3–14. DOI: 10.1785/0220180259.
- Kuhn, M., and Johnson, K., 2013. *Applied Predictive Modeling*. New York: Springer.
- Kuhn, S., Cracknell, M. J., and Reading, A. M., 2018. *Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: a demonstration study from the Eastern Goldfields of Australia*. Geophysics 82, 183–193. DOI: 10.1190/geo2017-0590.1.
- Kutner, M. H., Nachtsheim, C. & Neter, J., 2004. *Applied linear regression models* (McGraw-Hill/Irwin, 2004).
- Lange, N.; Bishop, C.M.; Ripley, B.D. 1997., *Neural Networks for Pattern Recognition.* J. Am. Stat. Assoc. 1997, 92,1642.
- Lary, D. J., Alavi, A. H., Gandomi, A. H. & Walker, A. L., 2016. *Machine learning in geosciences and remote sensing*. Geoscience Frontiers 7, 3–10.
- LeCun, Y., Yoshua, B., Hinton, G., 2015. *Deep learning*. Nature, 521, no. 7553, pp. 436.
- Lecocq, T., F. Ardhuin, F. Collin, Camelbeeck, T., 2020a. *On the Extraction of Microseismic Ground Motion from Analog Seismograms for the Validation of Ocean-Climate Models*. Seismol. Res. Lett. XX, 1–13, DOI: 10.1785/0220190276.

- Lecocq T. et al., 2020b. *Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures*. Science. DOI: 10.1126/science.abd2438.
- Lee, W. S., Sheen, D. H., Yun, S. & Seo, K. W., 2011. *The Origin of Double-Frequency Microseism and Its Seasonal Variability at King Sejong Station, Antarctica.* Bull. Seismol. Soc. Am. 101, 1446–1451, https://DOI.org/10.1785/012010014.
- Lepore, S., and Grad, M., 2018. *Analysis of the primary and secondary microseisms in the wavefield of the ambient noise recorded in northern Poland*. Acta Geophys. 66, 915–929. DOI: 10.1007/s11600-018-0194-2.
- Li, J., Heap, A. D., Potter, A., and Daniell, J. J., 2011. *Application of machine learning methods to spatial interpolation of environmental variables*. Environ. Model. Softw. 26, 1647–1659. DOI: 10.1016/j.envsoft.2011.07.004.
- Li, Y., and Cheng, B., 2009. *An improved K-Nearest neighbor algorithm and its application to high resolution remote sensing image classification.* Proceedings of the IEEE Geoinformatics Conference, Fairfax, VA, 1–4.
- Liaw, A., andWiener, M., 2002. *Classification and regression by random Forest*. R News 2, 18–22.
- Longuet-Higgins, M. S.,1950. *A theory of the origin of microseisms*. Philos. Trans. R. Soc. Lond. Ser. A 243, 1–35. DOI: 10.1098/rsta.1950.0012.
- Louppe, G., 2014. *Understanding Random Forests: From Theory to Practice*, PhD Thesis, University of Liege.
- MacAyeal, D.R., Okal E.A., Aster, R.C., Bassis, J.N., Brunt, K.M., Cathles, L.M., Drucker, R., Fricker, H.A., Kim, Y., Martin, S., Okal, M.H., Sergienko,

O.V., Sponsler, M.P., Thom, E., 2006. *Transoceanic wave propagation links iceberg calving margins of Antarctica with storms in tropics and Northern Hemisphere*. Geophy. Res. Lett. Volume 33, Issue 17. DOI: https://doi.org/10.1029/2006GL027235.

- MacQueen, J., 1967. *Some methods for classification and analysis of multivariate observations*. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1(14), 281–297.
- Malfante, M., Dalla Mura, M., Metaxian, J., Mars, J. I., Macedo O., Inza, A., 2018. *Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives.* IEEE Signal Processing Magazine, vol. 35, no. 2, pp. 20-30, March 2018, DOI: 10.1109/MSP.2017.2779166.
- Mansouri, E., Feizi, F., Rad, A. J., and Arian, M., 2018. *Remote-sensing data processing with the multivariate regression analysis method for iron mineral resource potential mapping: a case study in the Sarvian area, central Iran*. Solid Earth 9:373. DOI: 10.5194/se-9-373-2018.
- McLachlan, G. J., 2004. *Discriminant Analysis and Statistical Pattern Recognition* (Wiley Interscience, 2004).
- Milne, J., 1883. *Earth Pulsations.* Nature. 28, 367–370. DOI: https://doi.org/10.1038/028367e0.
- Mitchell, B. J., 1995. *Anelastic structure and evolution of the continental crust and upper mantle from seismic surface wave attenuation*. Rev. Geophys. 33(4), 441–462.

- Möllhoff, M.; Bean, C., 2016. *Seismic Noise Characterization in Proximity to Strong Microseism Sources in the Northeast Atlantic*. Bull. Seism. Soc. Am. 2016, 106, 464–477.
- Moschella, S., Cannata, A., Cannavò, F., Di Grazia, G., Nardone, G., Orasi, A., Picone, M., Ferla, M., Gresta, S., 2020. *Insights Into Microseism Sources by Array and Machine Learning Techniques: Ionian and Tyrrhenian Sea Case of Study*. Front. Earth Sci. 8:114. DOI: 10.3389/feart.2020.00114.
- Murphy, K. P., 2012. *Machine learning: a probabilistic perspective*. The MIT Press.
- Musa, Z.; Popescu, I.; Mynett, A., 2015. *A review of applications of satellite SAR , optical, altimetry and DEM data for surface water modelling, mapping and parameter estimation*. Hydrol. Earth Syst. Sci. 2015, 19, 3755–3769.
- Myers, J. L. & Well, A. D. 2003. *Research Design and Statistical Analysis*. (Mahwah, Lawrence Erlbaum, 2003).
- Noi, P. T., and Kappas, M., 2018. *Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery*. Sensors 18:18. DOI: 10.3390/s18010018.
- Nishida, K., Kawakatsu, H., Fukao, Y., Obara, K., 2008. *Background Love and Rayleigh waves simultaneously generated at the Pacific Ocean floors*. Geophys. Res. Lett. Volume35, Issue16. DOI: https://doi.org/10.1029/2008GL034753.
- 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, L11601, https://DOI.org/10.1029/2012GL051896.

- Oliver, J., and Page, R., 1963. *Concurrent storms of long and ultralong period microseisms*. Bull. Seismol. Soc. Am. 53, 15–26.
- Orasi, A., Picone, M., Drago, A., Capodici, F., Gauci, A., Nardone, G., 2018. *HF radar for wind waves measurements in the Malta-Sicily Channel*. Measurement 128, 446–454. DOI: 10.1016/j.measurement.2018.06.060.
- Patanè, D., Privitera, E., Ferrucci, F., and Gresta, S., 1994. *Seismic activity leading to the 1991-1993 eruption of Mt. Etna and its tectonic implications.* Acta Vulcanol. 4, 47–55.
- Peterson, J.R., 1993. *Observations and modeling of seismic background noise*. US Geological Survey, January 1993. DOI: 10.3133/ofr93322.
- Porritt, R. W., Miller, S.M., O'Driscoll, L.J., Harris, C.W., Roosmawati, N., da Costa, L.T., 2016. *Continent–arc collision in the Banda Arc imaged by ambient noise tomography*. Earth and Planetary Science Letters 449, 246–258 ().
- Povak, N. A., Hessburg, P. F., McDonnell, T. C., Reynolds, K. M., Sullivan, T. J., Salter, R. B., 2014. *Machine learning and linear regression models to predict catchment-level base cation weathering rates across the southern Appalachian Mountain region, USA*. Water Resour. Res. 50, 2798–2814. DOI:10.1002/2013wr014203.
- Pratt, M. J., Wiens, D. A., Winberry, J. P., Anandakrishnan, S., and Euler, G. G., 2017. *Implications of Sea ice on Southern Ocean microseisms detected by a seismic array in West Antarctica*. Geophys. J. Int. 209, 492–507. DOI: 10.1093/gji/ggx007.

- Prieto, G. A., Lawrence, J. F. & Beroza, G. C., 2009. *Anelastic Earth structure from the coherency of the ambient seismic field*. J. Geophys. Res. 114, B07303, https://DOI.org/10.1029/2008JB006067.
- Reguero, B.G.; Losada, I.J.; Méndez, F.J., 2019. *A recent increase in global wave power as a consequence of oceanic warming*. Nat. Commun., 10, 205.
- Reza Pourghasemi, H., Jirandeh, A. G., Pradhan, B., Xu, C., and Gokceoglu, C., 2013. *Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province*. Iran. J. Earth Syst. Sci. 122, 349–369. DOI: 10.1007/s12040-013-0282-2.
- Rivet, D., Campillo, M., Sanchez-Sesma, F., Shapiro, N. M., and Singh, S. K., 2015. *Identification of surface wave higher modes using a methodology based on seismic noise and coda waves*. Geophys. J. Int. 203, 856–868. DOI: 10.1093/gji/ggv339.
- Rost, S., and Thomas, C., 2002. *Array seismology: methods and applications*. Rev. Geophys. 40:1008. DOI: 10.1029/2000RG000100.
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J., Johnson, P.A., 2017. *Machine learning predicts laboratory earthquakes.* Geophys. Res. Lett. 44, 9276–9282, https://DOI.org/10.1002/2017GL074677.
- Rousseuw, P. & Silhouettes, J., 1987. *A Graphical Aid to the Interpretation and Validation of Cluster Analysis*. Computational and Applied Mathematics 20, 53–65 DOI: https://DOI.org/10.1016/0377-0427(87)90125-7.
- Sabra, K.G., Gersoft, P.. Roux, P., Kuperman, W.A., Fehler, M.C. 2005. *Surface wave tomography from microseisms in Southern California*. Volume32, Issue14. DOI: https://doi.org/10.1029/2005GL023155.

- Saito, T., 2010. *Love-wave excitation due to the interaction between a propagating ocean wave and the sea-bottom topography*. Geophy. J. Int., Volume 182, Issue 3, September 2010, Pages 1515–1523 DOI: https://doi.org/10.1111/j.1365-246X.2010.04695.x.
- Samuel, A.L. 1959. *Some Studies in Machine Learning Using the Game of Checkers*. IBM Journal of Research and Development 44:1.2: 210–229.
- Schweitzer, J., Fyen, J., Mykkeltveit, S., Gibbons, S. J., Pirli, M., Kühn, D., 2012. *Seismic arrays.* IASPEI New Manual of Seismological Observatory Practice 2 (NMSOP-2), ed. P. Bormann (Potsdam: IASPEI). DOI: 10.2312/GFZNMSOP-2_ch9.
- Shapiro, N.M., Campillo, M., Stehly, L., Ritzwoller, M.H., 2005. *High-Resolution Surface-Wave Tomography from Ambient Seismic Noise.* Science, Vol. 307, Issue 5715, pp. 1615-1618. DOI: 10.1126/science.1108339.
- Sibson, R., 1981. *A brief description of natural neighbour interpolation*. Interpreting Multivariate Data (ed. Barnett, V.) 21–36 (Wiley, New York, 1981).
- Slater, T., Sheperd, A., McMillan, M., Muir, A., Gilbert, L., Hogg, A.E., Konrad, H, Parrinello, T., 2018. *A new digital elevation model of Antarctica derived from CryoSat-2 altimetry.* The Cryosphere 12, 1551–1562, https://DOI.org/10.5194/tc-12-1551-2018.
- Steele, K. E., and Mettlach, T., 1993. *NDBC wave data – current and planned*. Measurement and Analysis - Proceedings of the Second International Symposium (Reston: ASCE), 198–207.

- Stehly, L., Campillo, M., Shapiro, N.M., 2006. *A study of the seismic noise from its long-range correlation properties*. J. Geophy. Res. Volume111, IssueB10. DOI: https://doi.org/10.1029/2005JB004237.
- Steffen, K., Key, J., Cavalieri, D.J., Comiso, J., Gloersen, P., St.Germain, K., Rubinstein, I., 1992. *The estimation of geophysical parameters using passive microwave algorithms*. In Microwave Remote Sensing of Sea Ice (ed. Carsey, F.) 202–231 (American Geophysical Union, Washington, 1992).
- Stopa, J., Ardhuin, F., Stutzmann, E., Lecocq, T. *Sea State Trends and Variability: Consistency Between Models, Altimeters, Buoys, and Seismic Data (1979-2016).* J. Geophy. Res. Oceans, Wiley-Blackwell, 2019, 10.1029/2018JC014607.
- Stutzmann, E., Schimmel, M., Patau, G., and Maggi, A., 2009. *Global climate imprint on seismic noise*. Geochem. Geophys. Geosyst. 10:Q11004. DOI: 10.1029/2009GC002619.
- Tanimoto, T., 2005. *The oceanic excitation hypothesis for the continuous oscillations of the Earth*. Geophys. J. Int. 160(1). DOI: http://dx.doi.org/10.1111/j.1365-246X.2005.02491.x.
- Tanimoto, T., 2007. *Excitation of microseisms*. Geophys. Res. Lett. 34: L05308. DOI: http://dx.doi.org/10.1029/2006GL029046.
- Tanimoto, T., Heki, K., and Artru-Lambin, J., 2015. *Interaction of Solid Earth, Atmosphere, and Ionosphere*. Treatise in Geophysics, 2nd Edn, ed. G. Schubert (Amsterdam: Elsevier), 421–444. DOI: 10.1016/b978-044452748-6.00075-4.

- Trnkoczy, A., 2012. *Understanding and parameter setting of STA/LTA trigger algorithm*. IASPEI New Manual of Seismological Observatory Practice 2 (NMSOP-2), ed. P. Bormann (Potsdam: IASPEI), 1–20. DOI: 10.2312/GFZ.NMSOP-2_IS_8.1.
- Trugman, D.T. & Shearer, P.M., 2018. *Strong correlation between stress drop and peak ground acceleration for recent M 1–4 earthquakes in the San Francisco Bay area*. Bull. Seism. Soc. Am. DOI: 10.1785/0120170245.
- Tsai, V. C., and McNamara, D. E., 2011. *Quantifying the influence of sea ice on ocean microseism using observations from the Bering Sea, Alaska.* Geophys. Res. Lett. 38:L22502. DOI: 10.1029/2011GL049791.
- Turner, R. J., Gal, M., Hemer, M. A., Reading, A. M., 2020. *Impacts of the cryosphere and atmosphere on observed microseisms generated in the Southern Ocean*. J. of Geophy. Res.: Earth Surface, 125, e2019JF005354. DOI: https://doi.org/10.1029/2019JF005354.
- Vapnik, V. N., 2000. *The Nature of Statistical Learning Theory*. New York: Springer.
- Von Storch, H.; Emeis, K.; Meinke, I.; Kannen, A.; Matthias, V.; Ratter, B.M.; Stanev, E.; Weisse, R.; Wirtz, K., 2015. *Making coastal research useful – cases from practice.* Oceanol., 57, 3–16.
- Wang, K., Ellsworth, W. L., Beroza, G. C., Williams, G., Zhang, M., Schroeder, D., Rubinstein, J., 2019. *Seismology with dark data: image based processing of analog records using machine learning for the rangely earthquake control experiment*. Seismol. Res. Lett. DOI:10.1785/0220180298.

- Webb, S.C., 1998. *Broadband seismology and noise under the ocean*. Reviews of Geophysics. Volume 36, Issue 1. DOI: https://doi.org/10.1029/97RG02287.
- Welch, P. D., 1967. *The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms*. IEEE Trans. Audio Electroacoust. 15, 70–73. DOI: 10.1109/TAU.1967.1161901.
- Wiese, A., Staneva, J., Schultz-Stellenfleth, J., Behrens, A., Fenoglio-Marc, L., Bidlot, J.R., 2018. *Synergy between satellite observations and model simulations during extreme events.* Ocean Sci. Discuss. 2018, 14, 1503–1521.
- Windnagel, A., Brandt, M., Fetterer, F., Meier, W., 2017. *Sea Ice Index Version 3 Analysis.* NSIDC Special Report 19. Boulder CO, USA: National Snow and Ice Data Center, http://nsidc.org/sites/nsidc.org/files/files/NSIDC-special-report-19.pdf.
- Wiszniowski, J., Plesiewicz, B. M., Trojanowski, J., 2014. *Application of real time recurrent neural network for detection of small natural earthquakes in Poland*. Acta Geophys. 62, 469–485. DOI: 10.2478/s11600-013-0140-2.
- Wrona, T., Pan, I., Gawthorpe, R.L., Fossen, H., 2018. *Seismic facies analysis using machine learning*. Geophysics 83(5):1-34 DOI: 10.1190/geo2017-0595.1.
- Wyatt, L.R., Green, J.J., 2009. *Measuring high and low waves with HF radar*. OCEANS 2009-EUROPE 2009, 1–5.
- Wyatt, L.R., Green, J.J., Middleditch, A., 2009. *Signal Sampling Impacts on HF Radar Wave Measurement*. J. Atmospheric Ocean. Technol. 2009, 26, 793–805.

- Xiao, H., Xue, M., Yang, T., Liu, C., Hua, Q., Xia, S., Huang, H., Le, B.M., Yu, Y., Huo, D., 2018. *The Characteristics of Microseisms in South China Sea: Results From a Combined Data Set of OBSs, Broadband Land Seismic Stations, and a Global Wave Height Model*. J. Geophys. Res. Solid Earth, 123, 3923–3942.
- Yang, X. S., Sherratt, S., Dey, N., Joshi, A., 2019. *Fourth International Congress on Information and Communication Technology: ICICT 2019.* London: Springer Nature.
- Zopf, D.O., Creech, H.C., Quinn, W.H. Wavemeter., 1976. *A Land-Based System for Measuring Nearshore Ocean Waves*. Mar. Technol. Soc. J. 1976, 10, 19–25.
- Zwally, H. J., Comiso, J. C., Parkinson, C. L., Cavalieri, D. J., Gloersen, P., 2002. *Variability of Antarctic sea ice 1979–1998*. J. Geophys. Res. 107, 3041, https://DOI.org/10.1029/2000JC000733.