Article

Unravelling the Relationship Between Microseisms and Spatial Distribution of Sea Wave Height by Statistical and Machine Learning Approaches

Andrea Cannata [1,2,*], Flavio Cannavò [2], Salvatore Moschella [1], Giuseppe Di Grazia [2],
Gabriele Nardone [3], Arianna Orasi [3], Marco Picone [3], Maurizio Ferla [3] and Stefano Gresta [1]

[1] Università Degli Studi di Catania, Dipartimento di Scienze Biologiche, Geologiche e Ambientali-Sezione di Scienze Della Terra, Corso Italia 57, I-95129 Catania, Italy; moschellasalvo@gmail.com (S.M.); gresta@unict.it (S.G.)

[2] Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo-Sezione di Catania, Piazza Roma 2, 95125 Catania, Italy; flavio.cannavo@ingv.it (F.C.); giuseppe.digrazia@ingv.it (G.D.G.)

[3] Istituto Superiore per la Protezione e la Ricerca Ambientale, CN-COS, via Vitaliano Brancati 48, 00144 Rome, Italy; gabriele.nardone@isprambiente.it (G.N.); arianna.orasi@isprambiente.it (A.O.); marco.picone@isprambiente.it (M.P.); maurizio.ferla@isprambiente.it (M.F.)

* Correspondence: andrea.cannata@unict.it; Tel.: +39-095-7195700

Abstract: Global warming is making extreme wave events more intense and frequent. Hence, the importance of monitoring the sea state for marine risk assessment and mitigation is increasing day-by-day. In this work, we exploit the ubiquitous seismic noise generated by energy transfer from the ocean to the solid earth (called microseisms) to infer the sea wave height data provided by hindcast maps. To this aim, we use a combined approach based on statistical analysis and machine learning. In particular, a random forest model shows very promising results in the spatial and temporal reconstruction of sea wave height by microseisms. The observed dependence of input importance from the distance sea grid cell-seismic station suggests how the reliable monitoring of the sea state in a wide area by microseisms needs data recorded by dense networks, comprising stations evenly distributed along the coastlines.

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., [1,2]. 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 [3].

Among the different instruments used to monitor the sea state, e.g., [4], 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 [5]. 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 [6–8]. 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 [5,9,10].

Finally, seismic signals recorded by seismometers can also be used as a sea "wavemeter" [11]. 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., [12]. Based on both spectral content and source mechanism, it is possible to distinguish two main types of microseism: primary and secondary. The former, also called "single-frequency" microseisms, shares the same spectral content as ocean waves (0.05–0.077 Hz; e.g., [13]) and is likely generated by interactions of ocean waves with the sloping seafloor and/or by ocean waves breaking/shoaling against the shoreline, e.g., [14]. The latter, also known as "double-frequency" microseisms, is characterized by most energy in the band 0.1–0.2 Hz, e.g., [13], shows higher amplitude than the primary microseism, and its source is associated with interactions of waves of the same frequency travelling in opposite directions, e.g., [15,16]. Finally, a short-period secondary microseism, exhibiting a frequency content between 0.2 and 0.4 Hz, e.g., [13], can also be identified. Its source is generally related to local wave–wave interactions [17,18].

On the basis of its source mechanism, microseisms have been used as a proxy of the sea state. For instance, it has been used to reconstruct the spatial and temporal distribution of sea ice around Antarctica [19]; but, above all, the relationship between microseism amplitude and sea wave height has been explored. For instance, [20] made use of seismic and buoy data acquired in California to empirically derive site-specific seismic-to-wave transfer functions, allowing reliable estimations of ocean wave height to be obtained from broadband seismometer data. [21] presented the first numerical model of seismic noise, allowing all the classes of sea states to be reproduced, which give rise to significant seismic noise levels. By comparing seismic data and sea wave height recorded by buoys near the Ligurian coast (Italy), [22,23] calibrated empirical equations allowing microseism energy to be turned into sea wave height.

In this paper, by both 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 [2], the latter data have been provided by 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.

2. Materials and Methods

2.1. Data

To analyze microseisms, seismic signals recorded by the vertical component of six stations, located close to the Eastern Sicilian coast and belonging to the seismic permanent network run by Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo, Sezione di Catania (INGV-OE), were used (Figure 1 and Table 1). The seismometers equipping these stations are broadband three-component Trillium 40 s (Nanometrics™). The digitizers, acquiring at a sampling rate of 100 Hz, are Trident (Nanometrics™).

In particular, the time interval 2010–2017, offering a very good data continuity in most of the considered stations, was chosen.
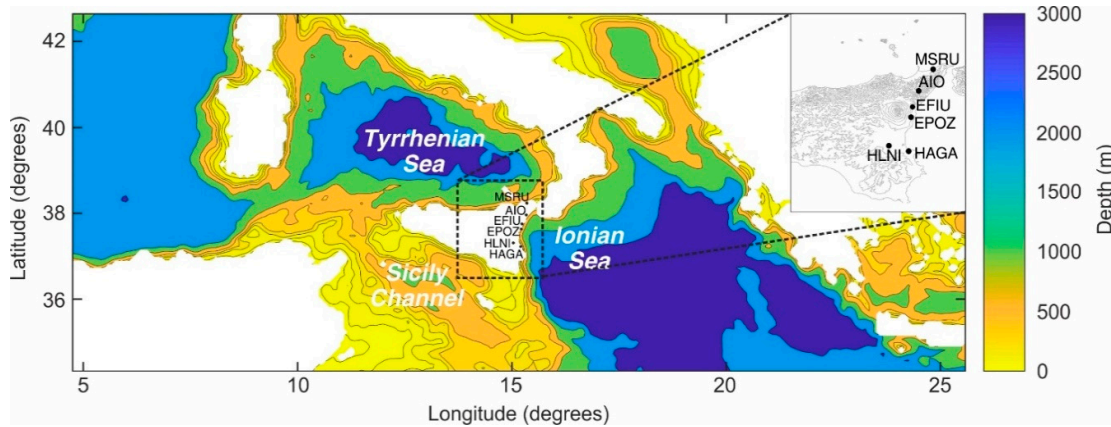
Figure 1. Map of a portion of the Mediterranean Sea, showing bathymetry (data from http://marine.copernicus.eu/) and the location of the six seismic stations used in this work (black dots and labels). The inset in the right upper corner shows the digital elevation model of Eastern Sicily, with the location of the seismic stations.

Table 1. Coordinates of the seismic stations used in this work.

Station Name	Longitude	Latitude	Altitude (m a.s.l.)
MSRU	15.5083	38.2638	401
AIO	15.2278	37.9689	715
EFIU	15.2101	37.7897	98
EPOZ	15.1885	37.6718	119
HAGA	15.1552	37.2853	126
HLNI	14.872	37.3485	147

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 4.5.4 [24]. In particular, the hourly significant wave height data of a portion of the Mediterranean Sea (shown in Figure 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. The significant wave height is defined in terms of moments of the omnidirectional spectrum, $S(f)$, where f is the frequency. The nth moment of the spectrum is defined as

$$m_n = \int_0^\infty f^n S(f) \, df. \tag{1}$$

Wave height is defined as follows:

$$Hs = 4\sqrt{m_0}. \tag{2}$$

It is demonstrated that Hs is approximately equal to the average of the highest one-third of the waves.

2.2. Spectral and Amplitude Analysis

To characterize the spectral content of the seismic signals recorded by the vertical component of the six considered stations, the short-time Fourier transform was carried out as follows: (i) The signal was divided into 81.92 s long non-overlapping windows and a spectrum was computed for each window; (ii) a daily spectrum was calculated by applying Welch's overlapped segment averaging estimator [25] on all the spectra of the same day; (iii) finally, all the daily spectra were gathered and visualized as a spectrogram (Figure 2a–f). In addition, a single spectrum was computed for each station as the median of all the daily spectra (Figure 2g–l).
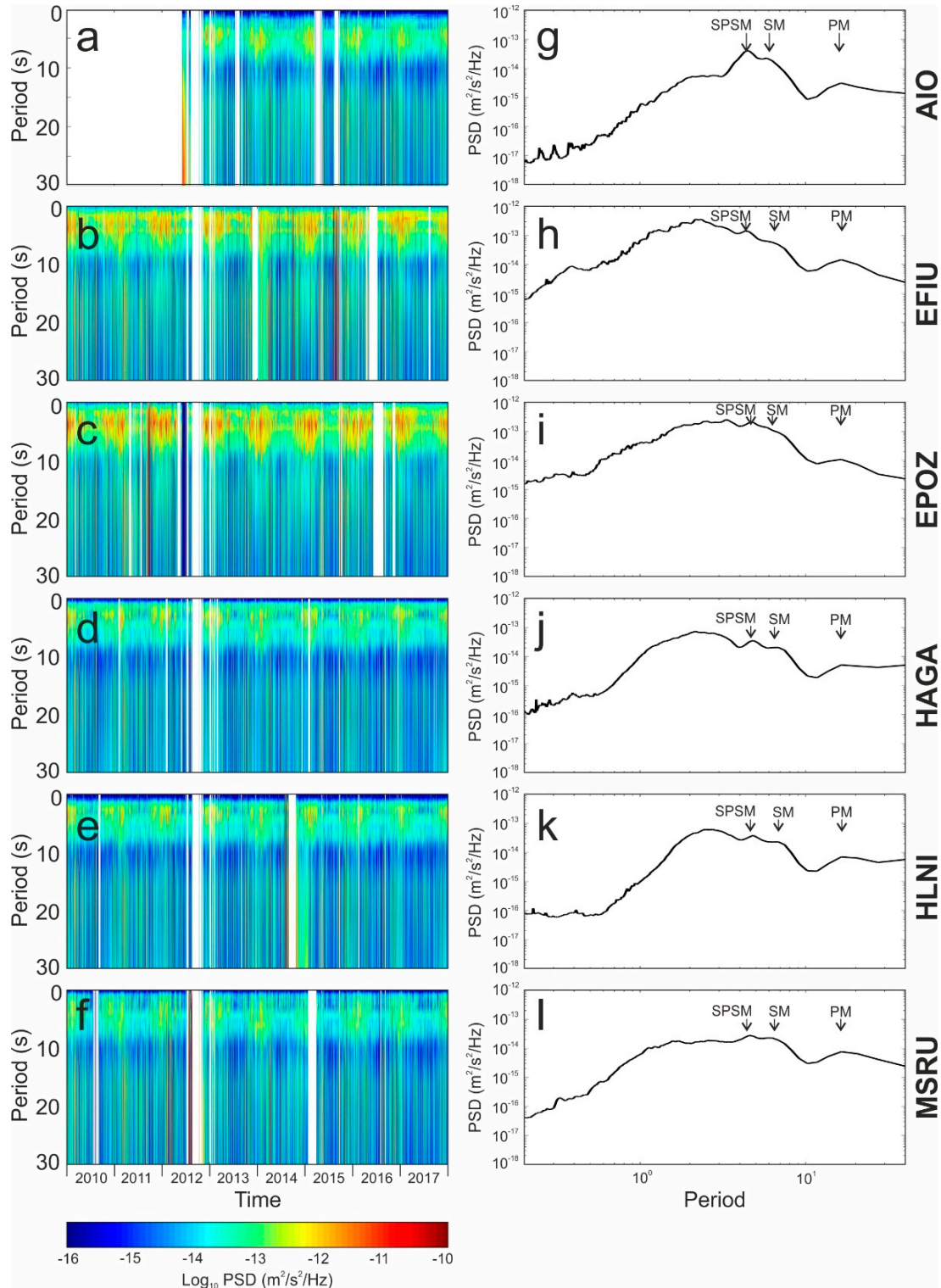
Figure 2. (**a–f**) Spectrograms of the seismic signal recorded by the vertical component of the six considered stations. (**g–l**) Median spectra of the seismic signal recorded by the vertical component of the six considered stations. Arrows and labels in (**g–l**) indicate the spectral peaks corresponding 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 hourly root-mean-square (RMS) amplitude of the seismic signal filtered in different frequency bands: (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. In Figure 3, the variation over time of the seismic RMS amplitude in band (vi) is shown.
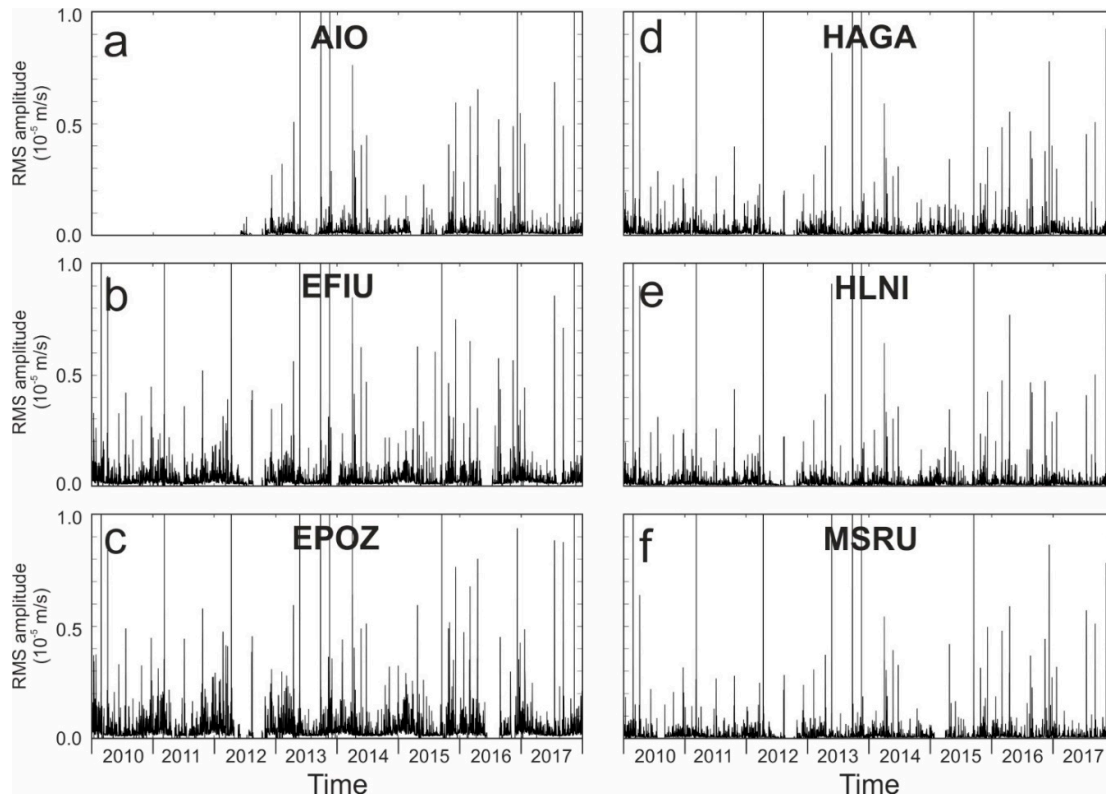

Figure 3. RMS amplitudes of the seismic signal recorded by the vertical component of the six considered stations (a: AIO; b: EFIU; c: EPOZ; d: HAGA; e: HLNI; f: MSRU), and filtered in the band 0.05–0.4 Hz, comprising short-period secondary, secondary, and primary microseisms.

2.3. Correlation Analysis

Following the idea of several authors, e.g., [26–29], 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. As such time series should correlate well at or near the source region of the microseism [28], this analysis provides an idea of the location of the main sources of microseisms recorded by the considered seismic stations.

Following [28,19], in place of using the more common Pearson correlation coefficient, we made use of the Spearman correlation coefficient. The former should be applied to explore the linear dependence between normally distributed parameters. As both significant wave height and seismic RMS amplitude do not have a normal distribution, as well as their relationship is likely nonlinear [27,28], the Spearman correlation coefficient was preferred. In addition, unlike the Pearson correlation coefficient, the Spearman correlation coefficient is less affected by outliers. The Spearman correlation coefficient, defined as a nonparametric measure of rank correlation, can be computed as follows [28]:

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

(3)

where d_i is the difference between the ranks of the two parameters and n is the number of observations. 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 totaling 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 taken into account.

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. 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 5).
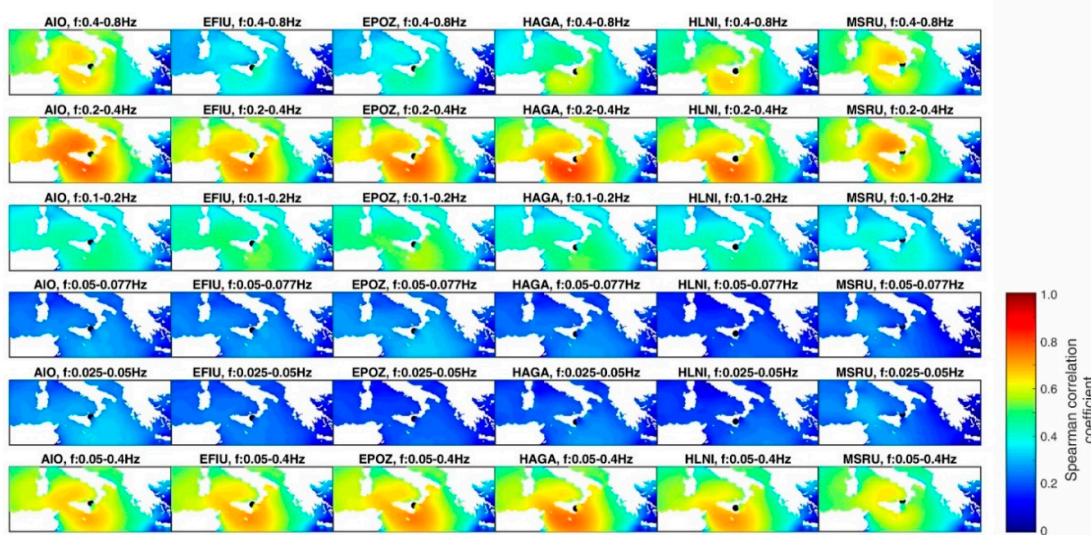

Figure 4. Maps of a portion of the Mediterranean Sea, showing the spatial distribution of Spearman's correlation coefficient computed between significant sea wave height and seismic RMS amplitudes, calculated on the vertical component of the six considered stations and filtered in different frequency bands (see plot titles). The black dots indicate the locations of the considered stations.
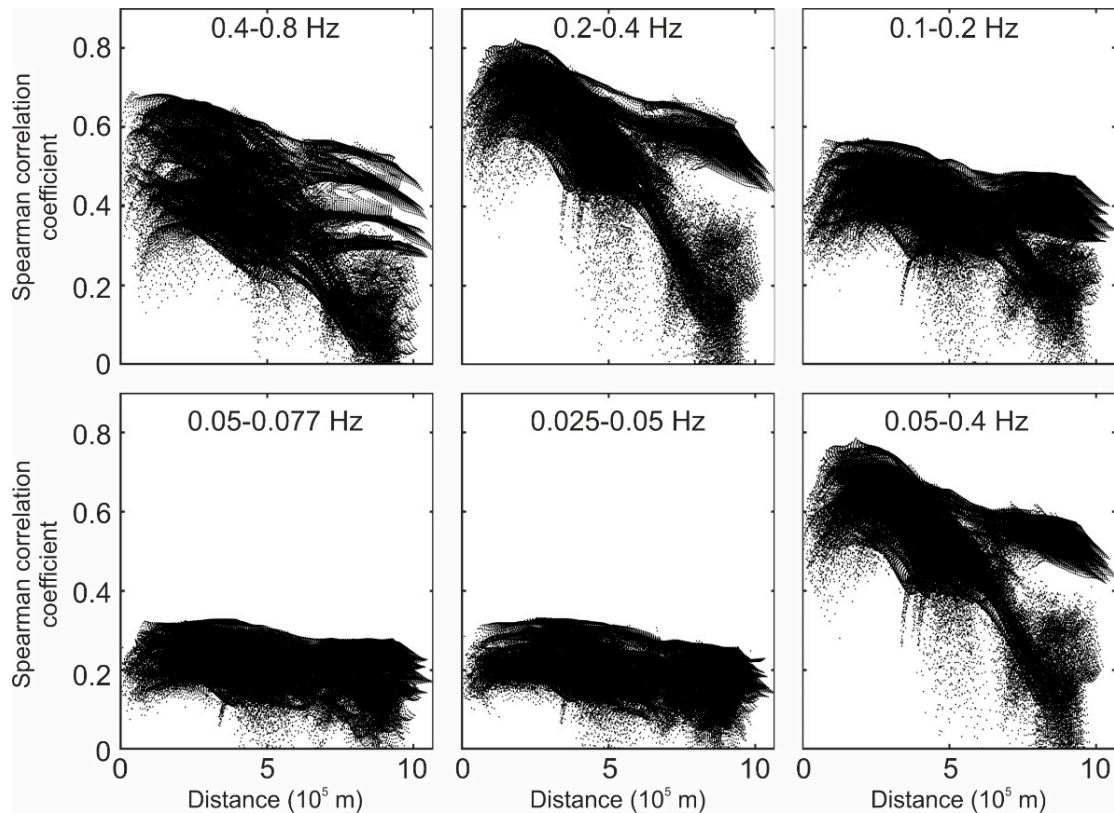
Figure 5. Cross-plots showing the relationship between Spearman's correlation coefficient, computed between significant sea wave height and seismic RMS amplitudes at different frequency bands (see plot titles), and the distance between the seismic station and the sea grid cell.

2.4. Machine Learning

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 [30,31]. RF has had several applications in geoscience, such as data-driven modeling of mineral prospectivity [32], lithological classification of underexplored areas by geophysical and remote sensing data [33], and atmospheric chemistry transport models [34].

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 following method divided into three main steps: (a) Training; (b) validation; (c) testing.

First, to evaluate the capability of generalization of the final model [35], 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. Such a technique consists of [31]:

1. Dividing the original input dataset into two complementary subsets;
2. building a model on one subset (called the "training set") (step a);
3. validating the model performance on the other subset ("validation set") (step b);
4. repeating (i–iii) "k" times, selecting different time intervals as the training set and validation set each time.

In the applied k-fold cross-validation, the input dataset was partitioned into 10 distinct and complementary pairs of training and validation sets (k = 10). 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 6 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.
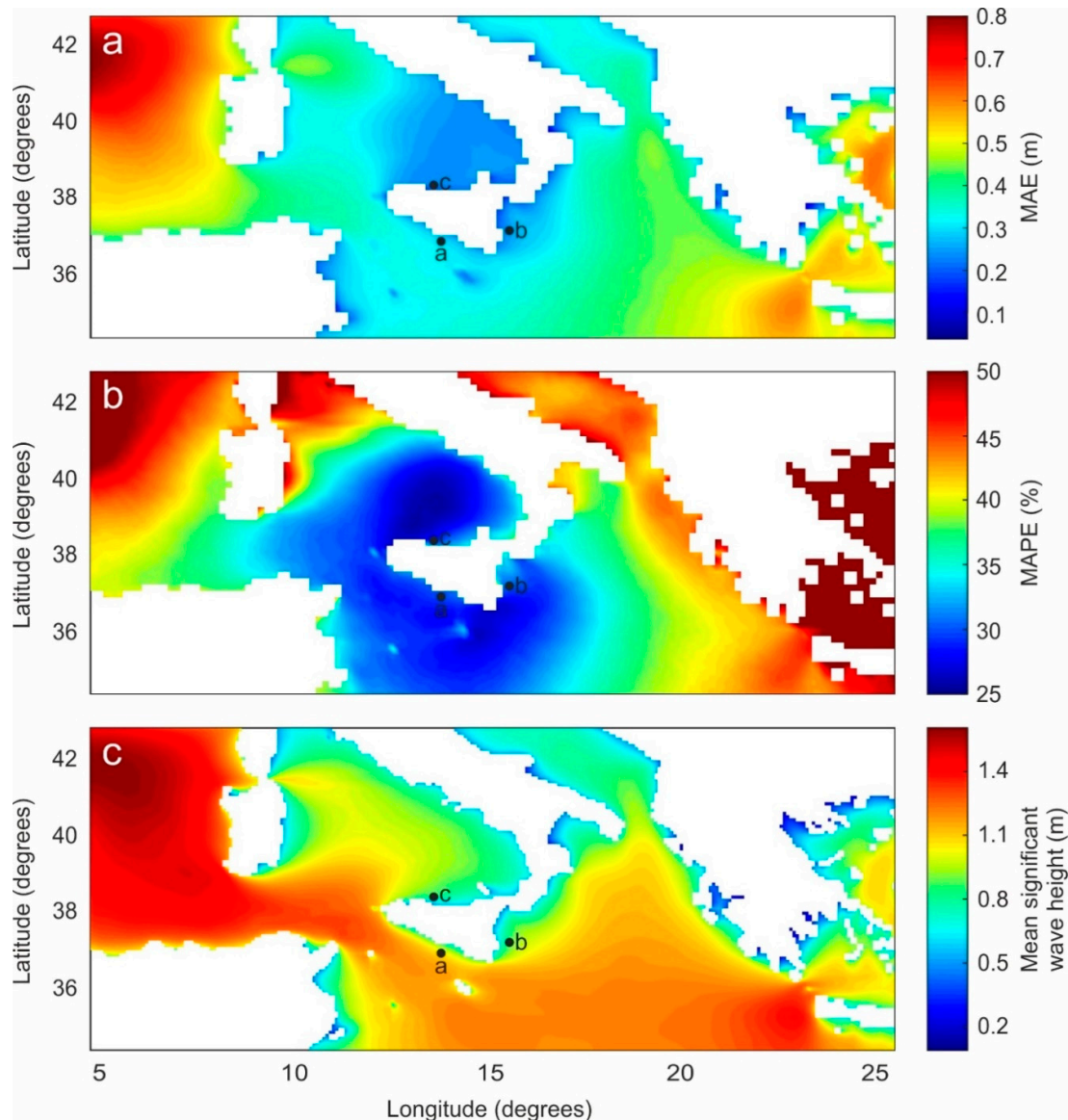

Figure 6. Spatial distribution of (**a**) mean absolute error (MAE), (**b**) mean absolute percentage error (MAPE), and (**c**) mean significant wave height computed during the interval 2010–2017. The black dots with labels "a," "b," and "c" indicate the locations of the grid cells, which the cross-plots in Figure 7 are referred to.

The final model was trained with the whole dataset January 2010–October 2017 and tested on the test set November–December 2017 (step (c) "testing"). In particular, comparisons between hindcast maps and predicted maps were carried out (three days are reported in Figure 8 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 7a), Ionian Sea (Figure 7b), and Tyrrhenian Sea (Figure 7c) (see Figure 6 for the locations of the three grid cells) were also obtained.
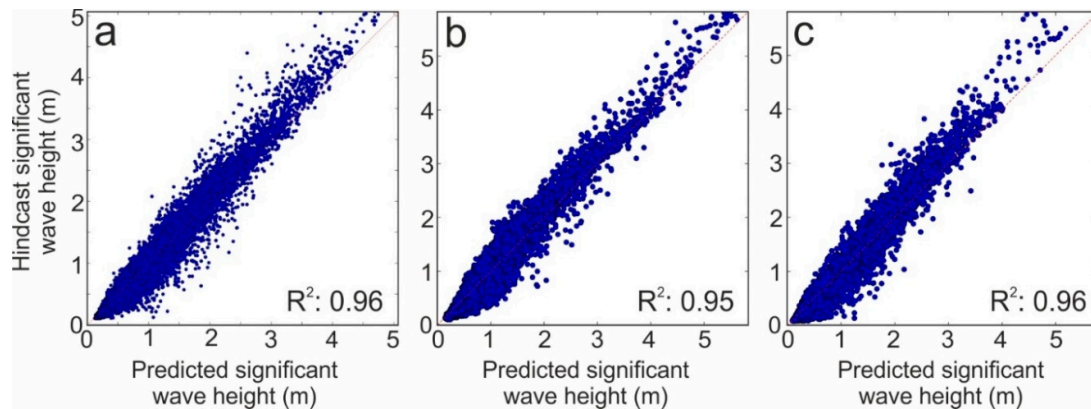

Figure 7. Scatter plots showing the hindcast significant wave height versus the predicted significant wave height from January 2010 to October 2017 for three grid cells, whose locations are indicated in Figure 6. The red dashed line in (**a**–**c**) is the y = x line. The value of the determination coefficient (R^2) is also reported in the plots.
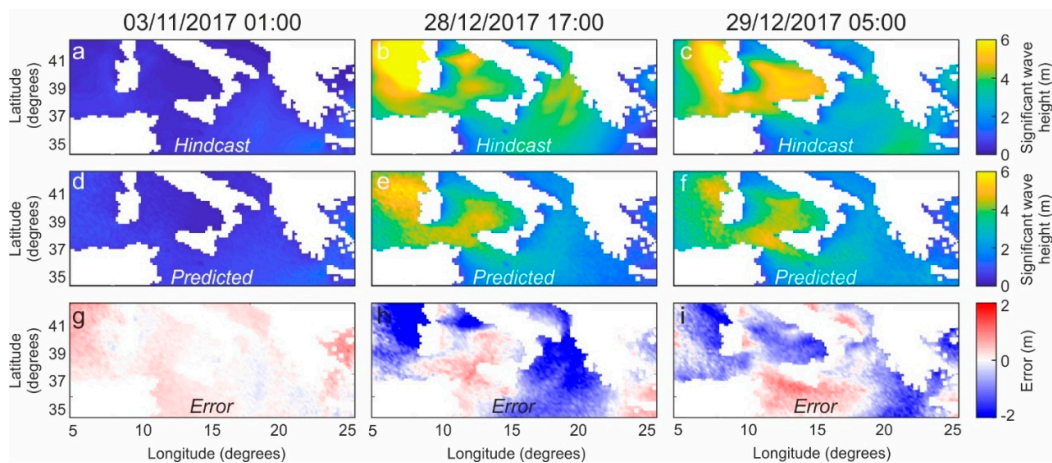

Figure 8. (**a**–**c**) Hindcast maps of significant sea wave height, (**d**–**f**) predicted maps of significant sea wave height based on microseism data, and (**g**–**i**) corresponding spatial distribution of the prediction error, computed as the difference between the predicted maps and the hindcast 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.

Finally, the RF technique also provides an index of input importance (indicated in %), which is shown by the histograms of Figure 9a–c. Aggregation through summation of the input importance was also carried out to rank the microseism bands (Figure 9d–f) and the seismic station (Figure 9g–i) for significant wave height reconstruction purposes.
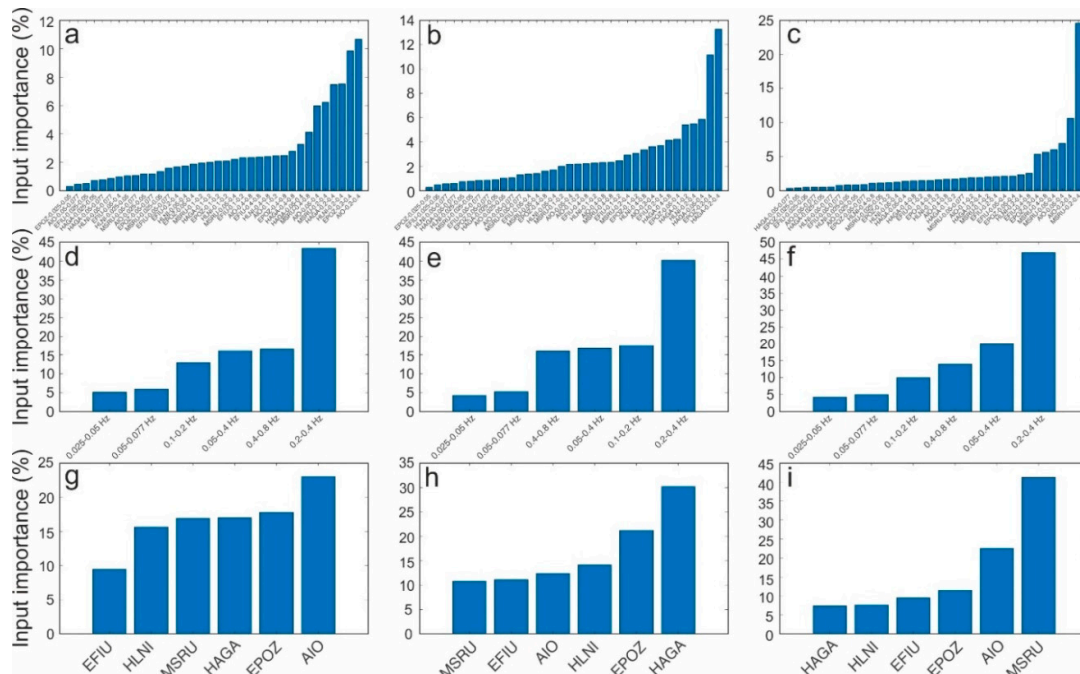
Figure 9. (**a**–**c**) Index of importance of all the inputs taken into account for the three grid cells labelled "a," "b," and "c" in Figure 6, respectively. (**d**–**f**) Aggregation through summation of the input importance allowing the ranking of the frequency bands for three grid cells labelled "a," "b," and "c" in Figure 6, 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 6, respectively.

3. Results

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 2g–l). Furthermore, as observed in the seismic recordings around the world, e.g., [36], 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 2). In particular, the short-period secondary microseism peak is the strongest among the three in all the considered stations (Figure 2g–l). With the exception of AIO station, 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 and minima during the summers (Figures 2a–f and 3).

As for the correlation maps, variable Spearman correlation values were obtained at the different stations and frequency bands. In particular, in agreement with [29], 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 and 5). 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 5 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). 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.

Regarding the machine learning analysis, as suggested by the spatial distribution of MAE and MAPE (Figure 6), by the cross-plots of hindcast significant wave height versus the predicted significant wave height (Figure 7) and by the comparisons between hindcast maps and predicted maps (Figure 8), promising results have been obtained. In particular, 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 6a), which is also characterized by very low mean values of significant wave height (Figure 6c). 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 6b).

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 6) reflects the results obtained by the correlation analysis (Figure 9d–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 9d–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 9g–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 9h). Regarding point "c," the station showing the highest importance is that closest to the Tyrrhenian coastline, MSRU (Figure 9i).

4. Discussion

Spectra of the microseism recorded by stations installed in Eastern Sicily show very high amplitudes of the short-period secondary microseism (Figure 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) 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., [17,18]. According to [17], the clear split between the spectral peaks of short-period secondary microseisms and secondary microseisms, observed in most of the computed spectra (Figure 2g–l), 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 2g–l). As indicated by the high correlation values with the sea wave height (Figures 4 and 5), such high seismic amplitudes at high frequencies (>0.4 Hz) can be related to energy transfer from the hydrosphere/atmosphere to the solid earth. According to previous studies, e.g., [37,38], 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 [39].

Seasonal modulation, identified in both spectrograms and seismic RMS amplitude time series (Figures 2a–f and 3), has been observed in all the areas worldwide with temperate latitudes, e.g., [40,41], 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., [41], 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., [42,19].

By gathering all the information provided by the correlation maps (Figures 4 and 5), 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., [17]. 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., [17,12].

However, as also stated by [43], 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., [44,45], assuming that losses from both scattering and transfer to the solid earth are not too effective [46]. In addition, [47] 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 6a,b). This finding, together with the observed dependence of both Spearman's correlation coefficient (Figure 5) and input importance (Figure 9h,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. By doing so, microseisms can become a very useful tool for significant wave height 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.

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

5. Conclusions

We analyzed the microseism recorded by seismic stations, located close to the Eastern Sicily coastlines, and quantitatively compared it to significant wave height data, provided by the hindcast maps, by using both statistical and machine learning approaches. These are the main findings:

- Short-period secondary microseisms showed the highest amplitude values among the considered frequency bands. In addition, high seismic amplitudes were observed at frequencies above the short-period secondary microseism band (0.4–0.8 Hz), likely related to energy transfer from the hydrosphere/atmosphere to the solid earth.
- Seasonal modulations of microseism amplitudes, due to the more efficient energy transfer from the sea to the solid earth in winters, when the seas are stormier, were identified.
- Correlation analysis suggests how the microseism recorded by the considered stations is mostly generated by sources located close to the stations, and then to the Eastern Sicilian coastlines.
- Machine learning analysis shows promising results in the spatial reconstruction of sea state in terms of significant wave height by microseism data. In the future, a reliable reconstruction of the sea state in wide areas will need microseism data recorded by dense networks, comprising stations evenly distributed along the coastlines.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: Seismic data are provided by the Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo-Sezione di Catania. Bathymetric and significant sea height data are provided by http://marine.copernicus.eu/services-portfolio/access-to-products/.

References

1. Von Storch, H.; Emeis, K.; Meinke, I.; Kannen, A.; Matthias, V.; Ratter, B.M.; Stanev, E.; Weisse, R.; Wirtz, K. Making coastal research useful – cases from practice. *Oceanol.* **2015**, *57*, 3–16.
2. Ferretti, G.; Barani, S.; Scafidi, D.; Capello, M.; Cutroneo, L.; Vagge, G.; Besio, G. Near real-time monitoring of significant sea wave height through microseism recordings: An application in the Ligurian Sea (Italy). *Ocean Coast. Manag.* **2018**, *165*, 185–194.
3. Reguero, B.G.; Losada, I.J.; Méndez, F.J. A recent increase in global wave power as a consequence of oceanic warming. *Nat. Commun.* **2019**, *10*, 205.
4. Holthuijsen, L.H. Waves in Oceanic and Coastal Waters by Leo H. Holthuijsen. *Waves in Oceanic and Coastal Waters* 2007.
5. Orasi, A.; Picone, M.; Drago, A.; Capodici, F.; Gauci, A.; Nardone, G.; Inghilesi, R.; Azzopardi, J.; Galea, A.; Ciraolo, G.; et al. HF radar for wind waves measurements in the Malta-Sicily Channel. *Meas.* **2018**, *128*, 446–454.
6. Fu, L.-L.; Cazenave, A. *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*, Satell. Elsevier: Amsterdam, The Netherlands, 2000.
7. Wiese, A.; Staneva, J.; Schultz-Stellenfleth, J.; Behrens, A.; Fenoglio-Marc, L.; Bidlot, J.-R. Synergy between satellite observations and model simulations during extreme events. *Ocean Sci. Discuss.* **2018**, *14*, 1503–1521.
8. Musa, Z.; Popescu, I.; Mynett, A. 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.
9. Wyatt, L.R.; Green, J.J. Measuring high and low waves with HF radar. *OCEANS 2009-EUROPE* **2009**, 1–5.
10. Wyatt, L.R.; Green, J.J.; Middleditch, A. Signal Sampling Impacts on HF Radar Wave Measurement. *J. Atmospheric Ocean. Technol.* **2009**, *26*, 793–805.
11. Zopf, D.O.; Creech, H.C.; Quinn, W.H. Wavemeter: A Land-Based System for Measuring Nearshore Ocean Waves. *Mar. Technol. Soc. J.* **1976**, *10*, 19–25.
12. Bromirski, P.D.; Duennebier, F.K. The near-coastal microseism spectrum: Spatial and temporal wave climate relationships. *J. Geophys. Res. Space Phys.* **2002**, *107*, 1–20.
13. Anthony, R.E.; Aster, R.C.; McGrath, D. Links between atmosphere, ocean, and cryosphere from two decades of microseism observations on the Antarctic Peninsula. *J. Geophys. Res. Earth Surf.* **2017**, *122*, 153–166.
14. Hasselmann, K. A statistical analysis of the generation of microseisms. *Rev. Geophys.* **1963**, *1*, 177.
15. Longuet-Higgins, M.S. A theory of the origin of microseisms. *Philos. Trans. R. Soc. London. Ser. A, Math. Phys. Sci.* **1950**, *243*, 1–35.
16. Oliver, J.; Page, R. Concurrent Storms of Long and Ultralong Period Microseisms. *Bull. Seism. Soc. Am.* **1963**, *53*, 15–26.

17. Bromirski, P.D.; Duennebier, F.K.; Stephen, R. Mid-ocean microseisms. *Geochem. Geophys. Geosystems* **2005**, *6*, 6.

18. Chen, Y.-N.; Gung, Y.; You, S.-H.; Hung, S.-H.; Chiao, L.-Y.; Huang, T.-Y.; Chen, Y.-L.; Liang, W.-T.; Jan, S. Characteristics of short period secondary microseisms (SPSM) in Taiwan: The influence of shallow ocean strait on SPSM. *Geophys. Res. Lett.* **2011**, *38*, 1–5.

19. Cannata, A.; Cannavo', F.; Moschella, S.; Gresta, S.; Spina, L. Exploring the link between microseism and sea ice in Antarctica by using machine learning. *Sci. Rep.* **2019**, *9*, 13050–15.

20. Bromirski, P.D.; Flick, R.E.; Graham, N. Ocean wave height determined from inland seismometer data: Implications for investigating wave climate changes in the NE Pacific. *J. Geophys. Res. Space Phys.* **1999**, *104*, 20753–20766.

21. Ardhuin, F.; Stutzmann, E.; Schimmel, M.; Mangeney, A. Ocean wave sources of seismic noise. *J. Geophys. Res. Space Phys.* **2011**, *116*, 1–21.

22. Ferretti, G.; Zunino, A.; Scafidi, D.; Barani, S.; Spallarossa, D. On microseisms recorded near the Ligurian coast (Italy) and their relationship with sea wave height. *Geophys. J. Int.* **2013**, *194*, 524–533.

23. Ferretti, G.; Scafidi, D.; Cutroneo, L.; Gallino, S.; Capello, M. Applicability of an empirical law to predict significant sea-wave heights from microseisms along the Western Ligurian Coast (Italy). *Cont. Shelf Res.* **2016**, *122*, 36–42.

24. Behrens, A. Documentation of a web based source code library for WAM; 2013. Available online: https://github.com/mywave/WAM/tree/master/documentation (accessed on 24 February 2020).

25. Welch, P. 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 Electroacoustics* **1967**, *15*, 70–73.

26. Bromirski, P.D. Vibrations from the "Perfect Storm." *Geochem. Geophys. Geosystems* **2001**, *2*, 2.

27. Essen, H.-H.; Krüger, F.; Dahm, T.; Grevemeyer, I. On the generation of secondary microseisms observed in northern and central Europe. *J. Geophys. Res. Space Phys.* **2003**, *108*, 1–15.

28. Craig, D.; Bean, C.; Lokmer, I.; Möllhoff, M. Correlation of Wavefield-Separated Ocean-Generated Microseisms with North Atlantic Source Regions. *Bull. Seism. Soc. Am.* **2016**, *106*, 1002–1010.

29. Xiao, H.; Xue, M.; Yang, T.; Liu, C.; Hua, Q.; Xia, S.; Huang, H.; Le, B.M.; Yu, Y.; Huo, D.; et al. 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* **2018**, *123*, 3923–3942.

30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

31. Kuhn, M.; Johnson, K. *Applied predictive modeling*; Springer: New York, USA, 2013; ISBN 9781461468493.

32. Carranza, E.J.M.; Laborte, A.G. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Comput. Geosci.* **2015**, *74*, 60–70.

33. Kuhn, S.; Cracknell, M.; Reading, A.M. Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia. *Geophys.* **2018**, *83*, B183–B193.

34. Keller, C.A.; Evans, M. Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. *Geosci. Model Dev.* **2019**, *12*, 1209–1225.

35. Lange, N.; Bishop, C.M.; Ripley, B.D. Neural Networks for Pattern Recognition. *J. Am. Stat. Assoc.* **1997**, *92*, 1642.

36. Aster, R.C.; McNamara, D.E.; Bromirski, P.D. Global trends in extremal microseism intensity. *Geophys. Res. Lett.* **2010**, *37*, 1–5.

37. Gal, M.; Reading, A.M.; Ellingsen, S.; Gualtieri, L.; Koper, K.D.; Burlacu, R.; Tkalčić, H.; Hemer, M. The frequency dependence and locations of short-period microseisms generated in the Southern Ocean and West Pacific. *J. Geophys. Res. Solid Earth* **2015**, *120*, 5764–5781.

38. Möllhoff, M.; Bean, C. Seismic Noise Characterization in Proximity to Strong Microseism Sources in the Northeast Atlantic. *Bull. Seism. Soc. Am.* **2016**, *106*, 464–477.

39. Cannata, A.; Di Grazia, G.; Montalto, P.; Ferrari, F.; Nunnari, G.; Patane', D.; Privitera, E. New insights into banded tremor from the 2008–2009 Mount Etna eruption. *J. Geophys. Res. Space Phys.* **2010**, *115*, 1–22.

40. Aster, R.C.; McNamara, D.E.; Bromirski, P.D. Multidecadal Climate-induced Variability in Microseisms. *Seism. Res. Lett.* **2008**, *79*, 194–202.

41. Stutzmann, E.; Schimmel, M.; Patau, G.; Maggi, A. Global climate imprint on seismic noise. *Geochem. Geophys. Geosystems* **2009**, *10*, 10.

42. Grob, M.; Maggi, A.; Stutzmann, E. Observations of the seasonality of the Antarctic microseismic signal, and its association to sea ice variability. *Geophys. Res. Lett.* **2011**, *38*, 38.

43. Essen, H.-H.; Klussmann, J.; Herber, R.; Grevemeyer, I. 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* **1999**, *51*, 33–45.

44. Cessaro, R.K. Sources of primary and secondary microseisms. *Bull. Seismol. Soc. Am.* **1994**, *84*, 142–148.

45. Chevrot, S.; Sylvander, M.; Benahmed, S.; Ponsolles, C.; Lefèvre, J.M.; Paradis, D. Source locations of secondary microseisms in western Europe: Evidence for both coastal and pelagic sources. *J. Geophys. Res. Solid Earth* **2007**, *112*, 1–19.

46. Bromirski, P.D.; Stephen, R.; Gerstoft, P. Are deep-ocean-generated surface-wave microseisms observed on land? *J. Geophys. Res. Solid Earth* **2013**, *118*, 3610–3629.

47. Gualtieri, L.; Stutzmann, E.; Juretzek, C.; Hadziioannou, C.; Ardhuin, F. Global scale analysis and modelling of primary microseisms. *Geophys. J. Int.* **2019**, *218*, 560–572.