Manuscript prepared for Ocean Sci.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 12 January 2016

Wave extremes characterization using Self-Organizing Maps

Francesco Barbariol[1], Francesco Marcello Falcieri[1], Carlotta Scotton[2], Alvise Benetazzo[1], Sandro Carniel[1], and Mauro Sclavo[1]

[1]Institute of Marine Sciences-Italian National Research Council, Venice (Italy)
[2]University of Padua, Padua (Italy)

Correspondence to: Francesco Barbariol (francesco.barbariol@ve.ismar.cnr.it)

Abstract. The Self-Organizing Map technique (SOM) is considered and extended to assess the extremes of a multivariate sea wave climate at a site. Main purpose is to obtain a more complete representation of the sea states, including the most severe states that otherwise would be missed by SOM. Indeed, it is commonly recognized, and herein confirmed, that SOM is a good regressor of a sample if the frequency of events is high (e.g. for low/moderate sea states), while SOM fails if the frequency is low (e.g. for the most severe sea states). Therefore, we have considered a trivariate wave climate (composed by significant wave height, mean wave period, and mean wave direction) collected continuously at the *Acqua Alta* oceanographic tower (northern Adriatic Sea, Italy) during the period 1979-2008. Three different strategies derived by SOM have been tested in order to capture the most extreme events. The first contemplates a pre-processing of the input dataset aimed at reducing redundancies; the second, based on the post-processing of SOM outputs, consists in a two-steps SOM where the first step is applied to the original dataset, and the second step is applied on the events exceeding a given threshold. A complete graphical representation of the outcomes of two-steps SOM is proposed. Results suggest that the post-processing strategy is more effective than the pre-processing one in order to represent the wave climate extremes. An application of the proposed two-steps approach is also provided, showing that a proper representation of the extreme wave climate leads to enhanced quantification of, for instance, the alonghshore component of the wave energy flux in shallow-water. Finally, the third strategy focuses on the peaks of the storms.

1 Introduction

The assessment of wave conditions at sea is fruitful for many research fields in marine and atmospheric sciences and for human activities in the marine environment. In the last decades, the observational network (mostly relying on buoys, satellites and other probes) has been integrated with numerical models outputs allowing to obtain the parameters of sea states over wider regions. Apart from the collection of wave parameters, the technique adopted to infer the wave climate at those sites is a crucial step in order to provide high quality data and information to the community. In

this context, several statistical techniques have been proposed to provide a reliable representation of the probability structure of wave parameters. While univariate and bivariate probability distribution functions (pdfs) are routinely derived, multivariate pdfs that represent the joint probability structure of more than two wave parameters are not straightforward. For individual waves, for instance, the bivariate joint pdf of wave height and period was derived by Longuet-Higgins (1983) and the bivariate joint pdf of wave height and direction was obtained by Isobe (1988). A trivariate joint pdf of wave height, wave period and direction is due to Kwon and Deguchi (1994). For sea states, attempts have been made to model the joint probability structure of the integral wave parameters. For instance, a joint pdf of the significant wave height and the average zero-crossing wave period was derived by Ochi (1978), and Mathisen and Bitner-Gregersen (1990). De Michele et al. (2007) exploited the "copula" statistical operators to describe the dependence among several random variables, e.g. significant wave hight, storm duration, storm direction and storm interarrival time, deriving their joint probability distributions. The same approach was applied by Masina et al. (2015) to the significant wave height and peak water level in the context of coastal flooding.

Recently, the Self-Organizing Maps technique (SOM) has been successfully applied to represent the multivariate wave climate around the Iberian peninsula (Camus et al., 2011a, b) and the south American continent (Reguero et al., 2013). SOM (Kohonen, 2001) is an unsupervised neural network technique that classifies multivariate input data and projects them onto a uni- or bi-dimensional output space, called map. The SOM technique was originally developed in the 1980s, and has been largely applied in various fields, including oceanography (Liu et al., 2006; Solidoro et al., 2007; Morioka et al., 2010; Camus et al., 2011a; Falcieri et al., 2013). Typical applications of SOM are vector quantization, regression and clustering. SOM gained credit among other techniques with same applications due to its visualization capabilities that allow to get multidimensional information from a 2D lattice. SOM has also the advantages of unsupervised learning, therefore vector quantization is performed autonomously. However, the quantization is strongly driven by the input data density. Indeed, SOM is principally forced by the most frequent conditions, while the most rare (i.e. the extreme events) are often missed. Consequently, it is highly unlike to find extremes properly represented on a SOM map.

In the context of ocean waves, drawing upon the works of Camus et al. (2011a, b) and Reguero et al. (2013), SOM input is generally constituted by a set of wave parameters measured or simulated at a given location and evolving over the time t, e.g. the triplet composed by significant wave height $H_s(t)$, mean wave period $T_m(t)$ and mean wave direction $\theta_m(t)$, even if other variables can be added (examples of five- or six-dimensional inputs can be found in Camus et al. (2011a)). Several activities in the wave field could benefit of the SOM outcomes, such as: selection of typical deep-water sea states for propagation towards the coast to study the longshore currents regime and coastal erosion, identification of typical sea states for wave energy resource assessment and wave farm optimization. In addition the empirical joint and marginal pdfs can be derived from SOM. As accurately shown in

Camus et al. (2011b), besides interesting potentialities, especially in visualization, some drawbacks in using SOM for wave analysis have emerged with respect to other classification techniques. Indeed, the largest H_s are missed by SOM because such extreme events are both rare (few comparisons in the "competitive" stage of the SOM learning) and distant from the others in the multidimensional space of input data (poorly influenced during the "cooperative" stage).

Moving from this evidence, the scientific question being asked is: how can we employ SOM with its visualization capabilities to improve representation of the extremes of a multivariate wave climate at a location? To answer this question we have followed three different strategies. Firstly, we have pre-preprocessed the SOM input data using the Maximum Dissimilarity Algorithm (MDA) in order to reduce the redundancies of the frequent low and moderate sea states, as done by Camus et al. (2011a). Indeed, MDA is a technique that reduces the density of inputs by preserving only the most representative (i.e. the most distant from each other in a Euclidean sense). Doing so, the most severe sea states are expected to gain weight in the learning process. We have called this strategy MDA-SOM. Then, we have focused on the post-processing of the SOM outputs. In this context, we have applied a two-steps SOM approach (herein called TSOM), by firstly running SOM to get a reliable representation of the low/moderate (i.e. the most frequent) wave climate, and then by running a second SOM on a reduced input sample. This new sample has been obtained by taking from first-step SOM results the events exceeding a prescribed threshold (e.g. 97^{th} percentile of H_s). To present results of two-steps SOMs, we have proposed a double-sided map, showing on the left the SOM map with the reliable representation of the low/moderate sea states, and on the right the map with the most severe sea states (i.e. the extremes). Then, we have applied SOM to the peak of the storms individuated by means of a Peak-Over-Threshold analysis (calling this strategy POT-SOM) and we have represented results using the double-sided map. An application of the proposed TSOM approach is finally reported: we have exploited the TSOM results to compute the longshore component of the wave energy flux, showing that a more proper representation of the extreme wave climate leads to an enhanced quantification of the energy approaching the shore.

2 Data

The dataset employed for the SOM analysis consists of wave time series gathered at the *Acqua Alta* oceanographic tower, owned and operated by the "Italian National Research Council - Institute of Marine Sciences" (CNR-ISMAR). *Acqua Alta* is located in the northern Adriatic Sea (Italy, northern Mediterranean Sea), approximately 15 km off the Venice coast at 17 m depth (Figure 1) and is a preferential site for marine observations (wind, wave, tide, physical and biogeochemical water properties are routinely retrieved), with a multi-parameters measuring structure on board (Cavaleri, 2000) upgraded over the years. For this study, we have relied on a 30 years (1979-2008) dataset of 3-hourly significant wave height H_s, mean wave period T_m and mean wave direction of propagation

θ_m (measured clockwise from the geographical north), observed using pressure transducers. Preliminarily, data have been preprocessed in order to remove occasional spikes. To this end, at first the time series have been treated with an ad-hoc despiking algorithm (Goring and Nikora, 2002). The complete dataset is therefore constituted of three variables and 50503 sea states.

Basic statistics of the time series (Table 1) points out that sea states at *Acqua Alta* have on average low intensity ($\langle H_s \rangle = 0.62$ m, where $\langle - \rangle$ denotes mean), though occasionally they can reach severe levels: the most intense event ($H_s = 5.23$ m, $T_m = 5.36$ s, $\theta_m = 242°$N) occurred on December 9th, 1992 during a storm forced by winds coming from north-east. Such severe events are not frequent, as confirmed by the 99^{th} percentile of H_s which is 2.68 m. Nevertheless they populate the wave time series at *Acqua Alta* and constitute the most interesting part of the sample, for instance for extreme analysis. Mean wave period is on average 4.1 s, while mean wave direction is 260°N, indeed most of the waves propagate towards the western quadrants.

This is represented more in detail in the pdf of θ_m (Figure 2, bottom panel), which shows that the most frequent directions of propagation are indeed in the range 180< θ_m <360°N (western quadrants), with peaks at 247.5 and 315°N. Directions associated with the most intense sea states ($H_s > 4.5$ m) can be obtained from the bivariate diagram ($H_s - \theta_m$) representing the joint pdf of H_s and θ_m (Figure 2, top panel): 247.5, 270 and 315°N. Mild sea states and calms ($H_s < 1.5\langle H_s \rangle$, following Boccotti (2000)) are the most frequent conditions at *Acqua Alta*, with 80% of occurrence during the 30 years of observations. They mainly propagate towards the western quadrants too, though the principal propagation directions of such seas states is north-west. In this context, the most frequent sea states at *Acqua Alta* are represented by $\{H_s, \theta_m\}$=$\{$0.25 m, 315°N$\}$. Storms in the area (denoted as sea states with $H_s \geq 1.5\langle H_s \rangle$) are generated by the dominant winds, i.e. the so called Bora and Sirocco winds (Signell et al., 2005; Benetazzo et al., 2012). Bora is a gusty katabatic and fetch-limited wind that blows from north-east; it generates intense storms along the Italian coast of Adriatic Sea characterized by relatively short and steep waves. Sirocco is a wet wind that blows from south-east; it is not fetch-limited and it generates longer and less steep waves than Bora which come from the southern part of the basin. Denoting conventionally as Bora the events with $180 \leq \theta_m \leq 270°$N, and as Sirocco the events with $270 < \theta_m \leq 360°$N, it follows that Bora storms have an occurrence of 12%, while Sirocco storms of 8%. The most frequent $\{H_s, T_m\}$ occurred in the Bora and Sirocco quadrants are shown in the bivariate ($H_s - T_m$) diagrams (Figure 3), and are $\{$0.15 m, 3.6 s$\}$ and $\{$0.35 m, 3.8 s$\}$, respectively, being Sirocco the most frequent among the two. The associated marginal pdfs (Figure 3) point out that Sirocco winds are responsible for the most of the calms, in particular for sea states with $H_s < 1$ m, while Bora for the most energetic sea states. Nevertheless, the pdf of H_s shows that Sirocco events with H_s in the range of 4-5 m can occur as well as Bora events. Bora is also associated with the shortest period waves observed: indeed, the pdfs of T_m almost coincide for waves shorter than 5.5 s, while for longer waves the probability level of Bora mean periods abruptly drops to values much smaller than those of Sirocco (which remains

to non negligible levels until 9 s). The consequence of shorter and higher Bora waves, with respect to Sirocco, is steeper waves (3% against 2% on average, respectively).

3 Self-Organizing Maps

3.1 Theoretical background

In this Section, we recall SOM features that are functional to the study. For more comprehensive readings we refer to Kohonen (2001) and other references cited in the following.

SOM is an unsupervised neural network technique that classifies multivariate input data and projects them onto a uni- or bi-dimensional output space, called map. Typically a bi-dimensional lattice is produced as output map. The global structure of the lattice is defined by the map shape that can be sheet, cylindrical or toroidal. The local structure of the lattice is defined by the shape of the elements, called units, that are typically either rectangular or hexagonal. The output map produced by a SOM on wave input data (e.g. as in Camus et al. (2011a)) furnishes an immediate picture of the multivariate wave climate and allow to identify, among the others, the most frequent sea states along with their significant wave height, mean direction of propagation and mean period.

The core of SOM is represented by the learning stage. Therefore, the choice of functions and parameters that control learning is crucial to obtain reliable maps. In SOM, the classification of input data is performed by means of competitive-cooperative learning: at each iteration, the elements of the output units compete among themselves to be the winning or Best-Matching Units (BMUs), i.e. the closest to the input data according to a prescribed metric (competitive stage), and they organize themselves due to lateral inhibition connections (cooperative stage). Usually, given that the chosen metric is a Euclidean distance, inputs have to be normalized before learning (e.g. by imposing unit variance or [0,1] range for all the input variables) and de-normalized once finished. The lateral inhibition among the map units is based upon the map topology and upon a neighboring function, that expresses how much a BMU affects the neighboring ones at each step of the learning process. During the learning process, the neighboring function reduces its domain of influence according to the decrease of a radius, from an initial to a final user-defined value. Learning can be performed sequentially, i.e. presenting the input data one at a time to the map, as done by the original incremental SOM algorithm. A more recent algorithm performs a batchwise learning, presenting the input data set all at once to the map (Kohonen et al., 2009). While the sequential algorithm requires the accurate choice of a learning rate function, which decreases during the process, the batch algorithm does not. At the beginning of the learning stage, the map has to be initialized: randomly, or preferably as an ordered 2D sequence of vectors obtained from the eigenvalues and eigenvectors of the covariance matrix of the data. In both SOM algorithms the learning process is performed over a prescribed number of iterations that should lead to an asymptotic equilibrium. Even if Kohonen (2001) argued that convergence is not a problem in practice, the convergence of the learning process to an optimal

solution is however an unsolved issue (convergence has been formally proved only for the univariate case, (Yin, 2008)). The reason is that, unlike other neural network techniques, SOM does not perform a gradient descent along a cost function that has to be minimized (Yin, 2008). Hence, in order to achieve reliable maps, the degree of optimality has to be assessed in other ways, e.g. by means of specific error metrics. The most common ones are the mean quantization error and the topographic error (Kohonen, 2001). The former is the average of the Euclidean distances between each input data and its BMU, and is a measure of the goodness of the map in representing the input. The latter, is the percentage of input data that have first and second best matching units adjacent in the map and is a measure of the topological preservation of the map.

3.2 SOM set-up

In this paper, the SOM technique has been applied by means of the *SOM toolbox* for *MATLAB* (Vesanto et al., 2000), that allows for most of the standard SOM capabilities, including pre- and post-processing tools. Among the techniques available, we have chosen the batch algorithm, because together with a linear initialization permits repeatable analyses, i.e. several SOM runs with the same parameters produce the same result (Kohonen et al., 2009). This is not a general feature of SOM, as the non-univoque character of both random initialization and selection of the data in the sequential algorithm lead to always different, though consistent, SOM maps (Kohonen, 2001).

Parameters controlling the SOM topology and batch-learning have been accurately examined and their values have been chosen as the result of a sensitivity analysis aimed at attaining the lowest mean quantization and topographic errors. Therefore, we have chosen bi-dimensional squared SOM outputs, that are sheet-shaped and with hexagonal cells. This kind of topology has been preferred to others (e.g. rectangular lattice, toroidal shape, rectangular cells, etc) because the maps so produced had the best topological preservation (low topographic error) and visual appearance. The maps size is 13x13 (169 cells), hence each cell represents approximately 300 sea states on average, if the complete dataset is considered. The lateral inhibition among the map units is provided by a cut-gaussian neighborhood function, that ensures a certain stiffness to the map (Kohonen, 2001) during the batch learning process (1000 iterations). At the same time, to allow the map to widely span the dataset, the neighborhood radius has been set to 7 at the beginning, i.e. more than half the size of the map, and then it linearly decreased to 1 during a single phase learning process.

Input data have been normalized so that the minimum and maximum distance between two realizations of a variable are 0 and 1, respectively. To this end, according to Camus et al. (2011a), the following normalizations have been used:

$$H = \frac{H_s - \min(H_s)}{\max(H_s) - \min(H_s)}; \qquad T = \frac{T_m - \min(T_m)}{\max(T_m) - \min(T_m)}; \qquad \theta = \theta_m/180 \tag{1}$$

Doing so, H and T range in [0,1], while θ ranges in [0,2). To take into account the circular character of θ_m in distance evaluation, following Camus et al. (2011a) we have considered the Euclidean-

circular distance as the metric for SOM learning. In this context, the distance d_{ij} between input data $\{H_i, T_i, \theta_i\}$ and SOM unit $\{\overline{H}_j, \overline{T}_j, \overline{\theta}_j\}$ is defined as:

$$d_{ij} = \{(H_i - \overline{H}_j)^2 + (T_i - \overline{T}_j)^2 + [\min(|\theta_i - \overline{\theta}_j|,\ \ 2 - |\theta_i - \overline{\theta}_j|)]^2\} \quad (2)$$

The Euclidean-circular distance has been therefore implemented in the scripts of *SOM toolbox* for *MATLAB* where distance is calculated.

4 SOM strategies to characterize wave extremes

In this Section, results of the standard SOM approach (applied one time, hence called single-step SOM) and results of the different strategies proposed to improve extremes representation are presented. The performances of single-step SOM, MDA-SOM and TSOM are assessed by comparing the wave parameters time series and their empirical marginal pdfs to the time series reconstructed from the results of the different strategies and relative pdfs, respectively. POT-SOM is treated separately because a direct comparison with the other strategies using the described methods is not possible. At the end of the Section, an application of TSOM is reported.

4.1 Single-step SOM

Single-step SOM has been applied using the set-up illustrated in Section 3.2. The SOM output map in Figure 4 merges all the information about the trivariate wave climate at *Acqua Alta* (H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction) including the frequency of occurrence (F: outer hexagons color) of each $\{H_s, T_m, \theta_m\}$ triplet. Hence, one can have an immediate sight on the wave climate features and on the empirical joint pdf thanks to visual capabilities of SOM's output. Gradual and continuous change in wave parameters over the cells points out that the topological preservation is quite good, as confirmed by the 22% topographic error.

According to the map, the most frequent sea states are represented by the triplet $\{0.17$ m, 3.5 s, 323°N$\}$, which substantially resembles the information that one could have gather from the bivariate (H_s-T_m) and (H_s-θ_m) diagrams (Figure 3), though these two diagrams are not formally related one to the other. Most cells show wave propagation directions pointing towards the western quadrants, as also displayed in the joint and marginal pdfs of θ_m (Figure 2). The cells denoting sea states forced by land winds (pointing toward east) are clustered in the top-left corner of the map and have low frequencies of occurrence (individual and cumulated). The frequency of occurrence of calms is 80%, while that of Bora storms is 12% and that of Sirocco storms is 8% (using definition of calms, Bora and Sirocco storm events given in Section 2). Hence, the integral distribution of the observed events over H_s and θ_m is retained by SOM. Sea states with the longest wave periods are clustered in the top-right corner of the map.

The most severe sea states of the map are clustered in the top-right part of the map, but are limited to H_s values smaller than 2.75 m. Indeed, the triplet with the highest H_s produced by SOM is $\{2.75$

m, 5.9 s, 270°N$\}$. However, Table and diagrams in Section 2 have shown that H_s can exceed 5.0 m at *Acqua Alta*. Therefore, sea states with $H_s > 2.75$ m are represented by cells with lower H_s. This is clear in Figure 5, where a sequence of observed events, including one with $H_s > 4.0$ m, has been compared to the sequence reconstructed after SOM, i.e. for each sea state of the sequence the triplet assumes the values of the corresponding BMU. In Figure 5 sea states with $H_s > 2.75$ m are represented by the cell with the highest H_s, i.e. cell $\#118$ (first row, tenth column, assuming the cells numbering starts at the top-left cell and proceeds from top to bottom over map rows and then from left to right over map columns), hence H_s are limited to 2.75 m, whereas the peak of the most severe storm in Figure has $\{$4.46 m, 6.7 s, 275°N$\}$. Quantitatively, for this particular event, single-step SOM underestimates the peak of 32% H_s, 12% T_m and 2% θ_m. Although H_s appears to be the most affected (T_m and θ_m after SOM are in a better agreement with the original data), all the variables processed by SOM experience a tightening of the original ranges of variation as it is shown in Figure 6 displaying the marginal empirical pdfs of H_s, T_m and θ_m after SOM. Generally, pdfs provided by SOM are in good agreement with the original ones. However, the range of variation of H_s is reduced from $[0.05, 5.23]$ m to $[0.17, 2.75]$ m, the range of T_m from $[0.5, 10.1]$ s to $[2.4, 7.4]$ s, and the range of θ_m from $[0, 360]$°N to $[41, 323]$°N. The maximum H_s value given by SOM (2.75 m) is pretty close to the 99^{th} percentile value (2.68 m), pointing out that SOM provides a good representation of the wave climate up to the 99^{th} percentile approximately. Nevertheless, the remaining 1% of events not properly described (extending up to 5.23 m) is for some applications the most interesting part of the sample. This confirms that single-step SOM provides an incomplete representation of the wave climate.

4.2 Maximum Dissimilarity Algorithm and SOM (MDA-SOM)

In order to reduce redundancy in the input data and to enable a wider variety of represented sea states, in previous studies (e.g. Camus et al. (2011a)) authors applied the "Maximum Dissimilarity Algorithm" (MDA) before the SOM process. In doing so, a new set of input data for SOM is constituted by sampling the original data in a way that the chosen sea states have the maximum dissimilarity (herein assumed as the Euclidean-circular distance) one from each other. As a result of MDA, a reduction of the number of sea states with low/moderate H_s, i.e. the most frequent at *Acqua Alta*, is observed. Hence, MDA-SOM is expected to provide a better description of the extreme sea states. Nevertheless, as pointed out by Camus et al. (2011a) the reduction of the sample numerosity leads to lower errors in the 99^{th} percentile of H_s (chosen to represent extremes) but also to higher errors in the average of H_s. Therefore, in terms of percentage reduction of the original input dataset, an optimum balance has to be found in order to get good descriptions of the average and of the extreme wave climate.

In the MDA-SOM application, we have pre-processed the input dataset by applying MDA, as described in detail in Camus et al. (2011a). Looking for the best reduction coefficient, the original

dataset has been reduced by means of MDA from the initial 50503 sea states (100%) to 5050 (10%), with step 10%. The absolute errors on $\langle H_s \rangle$ and on the 99^{th} percentile of H_s after MDA-SOM, relative to the original dataset, are summarized in Table 2. The error on $\langle H_s \rangle$, initially 2%, monotonically increases up to 57%, while the error on the 99^{th} percentile of H_s, initially 9%, decreases down to 3% at $50-60\%$ and then increase up to 27%. With widening of the variables' range as principal target (hence a better description of extremes) but without losing the quality on the average climate description, we chose to consider 80% reduction (7% error on $\langle H_s \rangle$, 4% error on 99^{th} percentile H_s). The corresponding MDA-SOM output map displayed in Figure 7 is topologically equivalent to that produced by SOM (Figure 4), except for minor differences on the location of some sea states. However, the most frequent sea state has $\{H_s, T_m, \theta_m\}$={0.28 m, 2.8 s, 328°N}, which still resembles what have emerged from diagrams of Section 2, even if T_m is less in agreement with respect to single-step SOM. Also, the sea state with highest H_s has the triplet equal to {2.8 m, 6.0 s, 275°N}, hence even if the input dataset has been reduced, the representation of extremes is still unsatisfactory.

This is confirmed by the comparison of the original and the reconstructed (after MDA-SOM) time series. In Figure 8, the comparison has been extended to the results of 60% MDA-SOM (smaller error on 99^{th} percentile H_s, see Table 2) and 10% MDA-SOM (maximum input dataset reduction), in order to investigate if MDA-SOM can enhance extreme wave climate representation even accepting a worsening of the average one. Actually, 60% MDA-SOM performs only slightly better than 80% MDA-SOM in describing the chosen events; indeed the highest H_s triplet, which represent the sea states at the peak of the most severe storm, is {2.93 m, 5.8 s, 258°N}. A better reproduction of H_s at this peak is provided by 10% MDA-SOM, though the maximum is however missed and in its proximity the original data are overestimated. Indeed, 60% and 10% MDA-SOMs locally overestimate H_s in the low/moderate sea states.

The marginal empirical pdfs after MDA-SOM are compared in Figure 9 to the pdfs of the original dataset. The distributions are in good agreement and the representation is more complete with respect to single-step SOM, especially concerning H_s. Nevertheless, 10% MDA-SOM distribution for H_s exhibits a larger departure from the original distribution at 1.7 m with respect to single-step SOM. Also 10% MDA-SOM distributions, which provides the widest ranges, locally depart from the reference distributions, in particular for T_m and θ_m. The frequency of occurrence of calms is 81%, while that of Bora storms is 12% and that of Sirocco storms is 7%. Hence, except for a minor change in the frequency of calms and Sirocco events, the overall statistics resembles that one directly derived from the *Acqua Alta* dataset.

4.3 Two-steps SOM (TSOM)

A two-step SOM (TSOM) has been then applied to provide a more complete description of the wave climate at *Acqua Alta*. To this end, the SOM algorithm has been run a first time on the original

dataset, without reductions (first step). Then, outputs have been post-processed: a threshold H_s^* has been fixed, and the cells having $H_s > H_s^*$ have been considered to constitute a new input dataset that is composed of the sea states represented by the cells exceeding the threshold. Hence, a second SOM has been run on the new dataset (second step). Using the same SOM set-up as in the first step, we have obtained a two-sided map (Figure 10): the first map (left panel) provides a good representation of the low/moderate wave climate but fails in the description of the most severe sea states, which are described in the second map (right panel), focusing on the climate over H_s^*. Three thresholds have been tested that correspond to the 95^{th}, 97^{th} and 99^{th} percentile of H_s: 1.80 m, 2.12 m and 2.68 m, respectively. In the following, we have focused on the results with 97^{th} percentile threshold, since they have turned out to be more representative of the extreme wave climate than the others.

Figure 10 depicts TSOM results with $H_s^* = 2.12$ m (97^{th} percentile). The first map, on the left, is the map shown in Figure 4, representing the whole wave climate at *Acqua Alta*. On that map, the six cells with $H_s > 2.12$ m have been encompassed by a black line. Without such cells, the map on the left represents the low/moderate sea states, i.e. the 97% of the whole original dataset constituted by events with H_s below or equal to the 2.12 m threshold. The remaining 3% of events, represented by the encompassed cells, are the most severe events at *Acqua Alta*. The first step SOM associates to such events $2.12 < H_s < 2.75$ m, $5.0 < T_m < 6.5$ s and $249 < \theta_m < 299°$N. Hence, according to SOM, the most severe sea states pertain to a rather narrow directional sector ($50°$) hardly allowing to discriminate between Bora and Sirocco conditions. A more detailed representation of such extremes is provided by the second map in Figure 10, on the right, where extreme Bora and Sirocco events are more widely described by cells. Indeed, a sort of diagonal (from the top-right corner to the bottom-left corner of the map) divides the cells. Bora events are clustered on the left of this diagonal (top-left part of the map), while Sirocco ones on the right of that (bottom-right part of the map). On the diagonal, cells represent sea states that travel towards west. This configuration somehow resembles the one observed in the left map, except for the land sea states, in the top-left corner. The most severe sea sates are clustered in the top-right corner of the map and also, though to a smaller extent, in the bottom-left part of it. The resulting ranges of H_s, T_m and θ_m are $1.94 < H_s < 4.26$ m, $4.4 < T_m < 8.3$ s and $224 < \theta_m < 316°$N, respectively.

The widened ranges of wave parameters provided by TSOM allow a more complete description of the sea states at *Acqua Alta*, including the most severe as it is shown in Figure 11. There, for the sequence of events presented in previous Sections, the reconstructed TSOM time-series is compared to the original one. Also results with 95^{th} and 99^{th} percentile TSOM are plotted, and it clearly appears that the differences among the three tests (i.e. TSOM with H_s threshold on 95^{th}, 97^{th} and 99^{th} percentiles) are very small, in particular for what concerns θ_m. Nevertheless, 95^{th} percentile TSOM yields to a smaller estimate of the highest H_s peak with respect to the others, and 99^{th} percentile TSOM deviates more than the others from the original T_m.

Such differences are also found in the marginal empirical pdfs of the wave parameters, shown in Figure 12. Indeed, $p(H_s)$ and $p(T_m)$ locally differ among the three thresholds and also from the original pdf, in particular in the largest values of H_s and T_m. As expected, the more the threshold is high, the more H_s range widens, extending to higher values. Hence, 99^{th} percentile TSOM provides the more complete representation of the wave climate, at least concerning H_s. Indeed, the widest T_m range is obtained with 97^{th} percentile and the narrowest with 99^{th} percentile TSOM. Instead, $p(\theta_m)$ is equally represented by the three thresholds and is in excellent agreement with the original pdf, though the θ_m range is limited with the respect to the complete circle. In addition, local departure from the original pdfs are still observed, especially for H_s and T_m. The frequency of occurrence of calms is 81%, while that of Bora storms is 11% and that of Sirocco storms is 8%. Hence, except for a minor change in the frequency of calms and Bora events, the overall statistics resembles that one observed at *Acqua Alta*.

4.4 Peak-Over-Threshold SOM (POT-SOM)

As an additional strategy to provide a more complete representation of the wave climate through SOM, we tested a third different approach. SOM was applied initially on the whole dataset, and then on the peaks of the storms defined by means of Peak-Over-Threshold technique. Storms were identified according to the definition of Boccotti (2000): a storm is the sequence of H_s that remains at least 12 hours over a given threshold H_s^* corresponding to 1.5 times the mean H_s. We considered the $\langle H_s \rangle$ at *Acqua Alta* (Table 1) and then, with $H_s^* = 0.93$ m, we individuated 710 storms. The peaks of the storms constitute a new dataset that has been analyzed by means of SOM. At the end, we have obtained a double-sided map that represent at the same time the whole wave climate (on the left) and the "stormy" part of it (on the right).

POT-SOM output map is shown in Figure 13. As expected, stormy events are Bora and Sirocco events: the former are clustered on the upper and middle part of the map, the latter in the lower part of it. The most severe storms, concentrated on the right of the map, are both Bora and Sirocco events. The triplet with the highest H_s is $\{4.27$ m, 6.32 s, 265°N$\}$ and the maximum H_s value is very close to the 99^{th} percentile of H_s of the new dataset, i.e. 4.28 m. Hence, 99% of the stormy events are included within the represented range, resembling what observed for the original dataset analyzed with single-step SOM.

5 Discussion

A summary of the performances of the different SOM strategies is given in Table 3. There, single-step SOM, MDA-SOM with 80% reduction and TSOM with H_s threshold at 97^{th} percentile are compared in their capabilities of representing the wave climate at *Acqua Alta* by means of the cells. POT-SOM is not directly comparable to the other strategies since the dataset used for the second

map is composed by the storm peaks only. As done in the previous sections, the performances are assessed by comparing the reconstructed time series from each strategy with the original ones, and the resulting marginal pdfs with pdfs of the original data. However, here the performances are quantified by statistical parameters (see caption of Table 3 for nomenclature). Generally, the reconstructed time series are in agreement with the original ones, as shown by the high r_{av} (over 0.98) and r_{std} (over 0.89), as well as high CC (over 0.95) and low $RMSE$ (below 0.19 m for H_s, 0.37 s for T_m and 23° for θ_m). Nevertheless, the highest ratios and correlation coefficients, and the lowest RMSE pertain to TSOM. Similar conclusions can be drawn for the pdfs, which are reproduced with very high CC (over 0.95) and $RMSE_{pdf}$ (below 0.04) by all the approaches, but to a greater extent by TSOM. As expected, the most wide range variability among the different strategies concerns H_s. With the only exception of θ_m, whose widest range is provided by MDA-SOM, TSOM turned out to be the most efficient in providing the most complete representation among the tested strategies.

We verified that a higher size single-step SOM (e.g. 25x25, not shown here) can produce a wider range of extremes with respect to that used in the study (i.e. 13x13): the units' maximum H_s is 3.56 m instead of 2.75 m. In the same map configuration (i.e. 25x25), MDA preselection can further widen this range towards extremes, being 3.63 m the units' maximum H_s obtained with an 80% reduction of the sample (using MDA) and 3.66 m the units' maximum H_s with a 40% reduction. This has the effect of reducing the absolute error on 99th percentile of H_s (1% with 80% reduction and 11% with 40% reduction). However, the most extreme sea states are still far to be properly represented (we recall that the most extreme sea state observed had $H_s = 5.23$ m). In addition and most important, if a larger number of elements in the map can improve the SOM performance shown in the paper, it will certainly worsen the readability of the map and the possibility of extracting quantitative information from the map. Indeed, considering for instance the 25x25 map, sea states at a site would be represented by 625 typical sea states: a huge number that is hardly manageable for a practical classification of the wave conditions.

6 Application of TSOM

An application of TSOM is proposed to show that a more detailed representation of the extreme wave climate can enhance the quantification of the longshore component of the shallow-water wave energy flux P (per unit shore length), expressed as (Komar and Inman, 1970):

$$P = E c_g \sin\alpha \cos\alpha \tag{3}$$

where $E = \rho g H_s^2/16$ is the wave energy per unit crest length (being ρ the water density), c_g is the group celerity, and α is the mean wave propagation direction measured counterclockwise from the normal to the shoreline. P is a driving factor for the potential longshore transport, and its dependence upon the wave energy E (which in turn depends on the square of H_s) suggests that an accurate representation of H_s is crucial. As the shoreline in front of *Acqua Alta* tower is almost parallel

to the 20°N direction (i.e., orthogonal to the 290°N direction), the longshore transport is directed towards south-west when P is positive, and directed towards north-east when P is negative. Given the wave energy flux Ec_g, P is maximized when $\alpha = \pm 45°$N, which correspond to $\theta_m = 245°$N and $\theta_m = 335°$N, respectively.

In order to obtain the shallow-water values of wave parameters, following Reguero et al. (2013), we propagated the *Acqua Alta* sea state resulting from TSOM (see maps in Figure 10) from 17-meter to 5-meter depth (a typical closure depth in the region), approximately accounting for the wave transformations, i.e. shoaling, refraction, and wave breaking. In doing so, we assumed straight and parallel bottom contour lines, we neglected wave energy dissipation prior to wave breaking, and we allowed H_s to reach the 80% of the water depth at most (depth-induced wave breaking criterion). Roughly, shoaling mostly affects the Sirocco sea states that are typically associated to longer wavelengths with respect to Bora sea states. In shallow-water, refraction tends to reduce the difference between Bora and Sirocco directions with respect to *Acqua Alta*, as the normal direction to the shoreline, that waves tend to align to, is very close to the boundary (i.e. 270°N) that we assumed to discriminate between the two conditions. Sea states forced by land winds ($20° N < \theta_m < 200°$N) were excluded from the analysis.

The longshore component of the wave energy flux P at 5-meter depth is shown in Figure 14. It is worth noting that the left map represents the longshore component of the wave energy flux resulting from single-step SOM technique (e.g., the left panel of Figure 10). Here, P ranges between -2 kW/m and 8 kW/m, and the highest values are mainly due to Bora events, that are responsible for potential longshore transport towards south-west (even if few Sirocco events with θ_m close to 270°N have the same effect). According to the left map, the transport towards north-east is due to Sirocco events that however cause less intense potential transport. The highest P values are associated with the highest H_s events, clustered on the cells at the top of the Figure 10 left map. The right map of Figure 14 describes the longshore flux component due to the *Acqua Alta* sea states represented by the SOM cells exceeding the 97^{th} percentile H_s threshold (i.e. the six cells bounded by the black line in the left map). The range of P variation widens considerably when the extreme sea states are considered, with values ranging from -20 kW/m to 20 kW/m. As observed in the right map of Figure 10, the sea states exceeding the 97^{th} percentile threshold on H_s are Bora and Sirocco events. The Bora events in the top-left part of the map (except for two cells in the bottom-right corner) contribute to positive, i.e. south-westward, transport, while Sirocco events in the bottom-right part contribute to negative, i.e. north-eastward, transport. The most intense transport is associated with the highest H_s cells at the bottom-left, bottom-right, and top-right corners of the Figure 10 right map. The major difference with respect to single-step SOM estimate concerns the Sirocco sea states, associated with negative P, that had the most intense value extended from -2 kW/m to -20 kW/m.

The mean longshore wave energy flux in shallow water $\overline{P}$, i.e. the average of P weighted on the frequencies of occurrence F over the 30 years of observations, was obtained by taking the absolute

value of P from the maps of Figure 14 and is 0.57 kW/m (Table 4). In order to support this estimate, we compared the 1.71 kW/m estimate of the mean wave energy flux Ec_g at *Acqua Alta* against the 1.5 kW/m value obtained at the same site over 1996-2011 by Barbariol et al. (2013). The contributions to $\overline{P}$ from Bora ($\overline{P}_+$) and Sirocco ($\overline{P}_-$) are 0.45 kW/m and -0.12 kW/m, respectively, pointing out the predominant effect of Bora on the longshore transport over the western side of the Gulf of Venice. For comparison, $\overline{P}$ was also computed using single-step SOM results (see Table 4): in this case, $\overline{P}$ is 0.52 kW/m, $\overline{P}_+$ is 0.41 kW/m and $\overline{P}_-$ is -0.11 kW/m. Hence, with respect to TSOM, the estimate of the mean longshore energy flux is 9.0% lower for $\overline{P}$, 7.5% lower for $\overline{P}_+$ and 16.5% lower for $\overline{P}_-$.

7 Conclusions

In this paper, we have tested different strategies aimed at improving the characterization of multivariate wave climate using SOM. Indeed, we have verified that besides a satisfactory description of the low/moderate wave climate (in agreement with usual uni- and bivariate diagrams), single-step SOM approach misses the most severe sea states, which are hidden in SOM cells with H_s even considerably smaller than the extreme ones.

To our aims, we used the 1979-2008 trivariate wave climate $\{H_s, T_m, \text{and } \theta_m\}$ recorded at *Acqua Alta* tower, and we showed that, for instance, single-step SOM assigned most of the sea states with $H_s > 2.75$ m to the $\{2.75$ m, 5.9 s, 270°N$\}$ class. Hence, the most interesting part of the wave climate was condensed within few cells of the map, also hindering the distinction between Bora and Sirocco events, i.e. the prevailing meteorological conditions in the northern Adriatic Sea. To increase the weight of the most severe and rare events in SOM classification, we tested a strategy based on the pre-processing of the input dataset (i.e. MDA-SOM) and a strategy based on the post-processing of the SOM outputs (i.e. TSOM). Results presented in the study showed that the post-processing technique is more effective than the pre-processing one. Indeed, TSOM allowed a more accurate and complete representation of the sea states with respect to the one furnished by MDA-SOM, because it provided a wider range of the wave parameters (particularly H_s), and more reliable a posteriori reconstructions of time series and marginal pdfs. Nevertheless, some deviations from original pdfs were observed and the range of θ_m was not complete, such that sea states traveling towards the north were not properly described. This requires further studies to improve SOM applications to wave analysis, which are rather promising, thanks to the well recognized visualization capabilities of SOM. In this context, we proposed a double-sided map representation, which provides on the left a description of the whole wave climate that is particularly reliable for the low/moderate events, completed on the right by the description of the extreme wave climate. This novel representation was also employed to provide a SOM classification of the storms peaks, based on the Peak-Over-Threshold approach, on the right (POT-SOM).

Finally, TSOM was applied for the assessment of the potential longshore wave energy flux to show how practical oceanographic and engineering applications can benefit from this novel SOM strategy. Indeed, the mean flux in front of the Venice coast was found to be the 9% higher if evaluated after TSOM instead of SOM.

Acknowledgements. The research was supported by the Flagship Project RITMARE - The Italian Research for the Sea-coordinated by the Italian National Research Council and funded by the Italian Ministry of Education, University and Research within the National Research Program 2011-2015. The authors gratefully acknowledge Luigi "Gigi" Cavaleri for providing wave data at *Acqua Alta* tower and for the fruitful discussions.

References

Barbariol, F., Benetazzo, A., Carniel, S., and Sclavo, M.: Improving the assessment of wave energy resources by means of coupled wave-ocean numerical modeling, Renewable Energy, 60, 462–471, 2013.

Benetazzo, A., Fedele, F., Carniel, S., Ricchi, A., Bucchignani, E., and Sclavo, M.: Wave climate of the Adriatic Sea: a future scenario simulation, Nat. Hazards Earth Syst. Sci, 12, 2065–2076, 2012.

Boccotti, P.: Wave mechanics for ocean engineering, vol. 64, Elsevier Science, 2000.

Camus, P., Cofiño, A. S., Mendez, F. J., and Medina, R.: Multivariate wave climate using self-organizing maps, Journal of Atmospheric and Oceanic Technology, 28, 1554–1568, doi:10.1175/JTECH-D-11-00027.1, 2011a.

Camus, P., Mendez, F. J., Medina, R., and Cofiño, A. S.: Analysis of clustering and selection algorithms for the study of multivariate wave climate, Coastal Engineering, 58, 453–462, doi:10.1016/j.coastaleng.2011.02.003, 2011b.

Cavaleri, L.: The oceanographic tower Acqua Alta - activity and prediction of sea states at Venice, Coastal Engineering, 39, 29–70, doi:10.1016/S0378-3839(99)00053-8, 2000.

De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R.: A multivariate model of sea storms using copulas, Coastal Engineering, 54, 734–751, doi:10.1016/j.coastaleng.2007.05.007, 2007.

Falcieri, F. M., Benetazzo, A., Sclavo, M., Russo, A., and Carniel, S.: Po River plume pattern variability investigated from model data, Continental Shelf Research, 87, 84–95, doi:10.1016/j.csr.2013.11.001, http://dx.doi.org/10.1016/j.csr.2013.11.001, 2013.

Goring, D. G. and Nikora, V. I.: Despiking Acoustic Doppler Velocimeter Data, doi:10.1061/(ASCE)0733-9429(2002)128:1(117), 2002.

Isobe, M.: On joint distribution of wave heights and directions, in: Coastal Engineering Proceedings, 1(21), pp. 524–538, 1988.

Kohonen, T.: Self-Organizing Maps, vol. 30, doi:10.1007/978-3-642-56927-2, http://www.springerlink.com/index/10.1007/978-3-642-56927-2, 2001.

Kohonen, T., Nieminen, I. T., and Timo, H.: On the Quantization Error in SOM vs. VQ: A Critical and Systematic Study, in: Advances in Sel-Organizing Maps, p. 374, Springer, lecture no edn., 2009.

Komar, P. and Inman, D.: Longshore sand transport on beaches, Journal of Geophysical Research, 75, 5914–5927, doi:10.1029/JC075i030p05914, 1970.

Kwon, J. and Deguchi, I.: On the Joint Distribution of Wave Height, Period and Direction of Individual Waves in a Three-Dimensional Random Sea, in: Coastal Engineering Proceedings, 1(24), pp. 370–383, 1994.

Liu, Y., Weisberg, R. H., and He, R.: Sea surface temperature patterns on the West Florida Shelf using growing hierarchical self-organizing maps, Journal of Atmospheric and Oceanic Technology, 23, 325–338, doi:10.1175/JTECH1848.1, 2006.

Longuet-Higgins, M. S.: On the Joint Distribution of Wave Periods and Amplitudes in a Random Wave Field, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 389, 241–258, http://rspa.royalsocietypublishing.org/content/389/1797/241.abstract, 1983.

Masina, M., Lamberti, A., and Archetti, R.: Coastal flooding: A copula based approach for estimating the joint probability of water levels and waves, Coastal Engineering, 97, 37–52,

doi:10.1016/j.coastaleng.2014.12.010, http://linkinghub.elsevier.com/retrieve/pii/S0378383914002270, 2015.

Mathisen, J. and Bitner-Gregersen, E.: Joint distributions for significant wave height and wave zero-up-crossing period, doi:10.1016/S0141-1187(05)80033-1, 1990.

Morioka, Y., Tozuka, T., and Yamagata, T.: Climate variability in the southern Indian Ocean as revealed by self-organizing maps, Climate Dynamics, 35, 1059–1072, doi:10.1007/s00382-010-0843-x, http://link.springer.com/10.1007/s00382-010-0843-x, 2010.

Ochi, M. K.: On long-term statistics for ocean and coastal waves, in: Coastal Engineering Proceedings, 1(16), 1978.

Reguero, B. G., Méndez, F. J., and Losada, I. J.: Variability of multivariate wave climate in Latin America and the Caribbean, Global and Planetary Change, 100, 70–84, doi:10.1016/j.gloplacha.2012.09.005, http://dx.doi.org/10.1016/j.gloplacha.2012.09.005, 2013.

Signell, R. P., Carniel, S., Cavaleri, L., Chiggiato, J., Doyle, J. D., Pullen, J., and Sclavo, M.: Assessment of wind quality for oceanographic modelling in semi-enclosed basins, Journal of Marine Systems, 53, 217–233, 2005.

Solidoro, C., Bandelj, V., Barbieri, P., Cossarini, G., and Fonda Umani, S.: Understanding dynamic of biogeochemical properties in the northern Adriatic Sea by using self-organizing maps and k-means clustering, Journal of Geophysical Research, 112, 1–13, doi:10.1029/2006JC003553, 2007.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J.: SOM Toolbox for Matlab 5, Technical Report A57, 2, 59, doi:http://www.cis.hut.fi/somtoolbox/package/papers/techrep.pdf, 2000.

Yin, H.: The Self-Organizing Maps: Background, Theories, Extensions and Applications, in: Computational Intelligence: a compendium, p. 1179, Springer, studies in edn., 2008.

Table 1. Wave climate at *Acqua Alta* in the period 1979-2008. Mean ($\langle - \rangle$), standard deviation (std), minimum ($\min$), p^{th} percentile (p^{th} perc), and maximum ($\max$) of wave parameters.

	$\langle - \rangle$	std	$\min$	50^{th} perc	95^{th} perc	97^{th} perc	99^{th} perc	$\max$
H_s (m)	0.62	0.57	0.05	0.44	1.80	2.12	2.68	5.23
T_m (s)	4.1	1.1	0.5	3.9	6.0	6.35	7.18	10.1
θ_m (°N)	260	72	1	270	336	343	353	360

Table 2. MDA-SOM, absolute errors of average and 99^{th} percentile of H_s relative to the original dataset (%).

H_s	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%
average	2	4	7	13	15	22	25	32	45	57
99^{th} percentile	9	8	4	5	3	3	5	5	18	27

Table 3. Performance summary of different SOM approaches, through the comparisons of reconstructed to original time series, and resulting to original pdfs. r_{av}: ratio of time series averages, r_{std}: ratio of time series standard deviations, CC: time series cross-correlation coefficient, $RMSE$: time series root mean square error, CC_{pdf}: pdfs cross-correlation coefficient, $RMSE_{pdf}$: pdfs root mean square error).

H_s	r_{av}	r_{std}	CC	$RMSE$ (m)	range (m)	CC_{pdf}	$RMSE_{pdf}$
Single-step SOM	0.98	0.91	0.95	0.18	[0.17,2.75]	1.00	0.04
MDA-SOM (80%)	1.00	0.90	0.95	0.19	[0.21.2.82]	0.99	0.04
TSOM (97^{th} perc)	0.99	0.95	0.96	0.16	[0.17,4.26]	1.00	0.04
T_m	r_{av}	r_{std}	CC	$RMSE$ (s)	range (s)	CC_{pdf}	$RMSE_{pdf}$
Single-step SOM	1.00	0.89	0.95	0.34	[2.4,7.4]	0.99	0.02
MDA-SOM (80%)	1.00	0.90	0.95	0.37	[2.4,7.4]	0.95	0.05
TSOM (97^{th} perc)	1.00	0.90	0.95	0.32	[2.4,8.3]	0.99	0.02
θ_m	r_{av}	r_{std}	CC	$RMSE$ (°N)	range (°N)	CC_{pdf}	$RMSE_{pdf}$
Single-step SOM	1.00	0.92	0.95	23	[41,323]	0.97	0.00
MDA-SOM (80%)	0.99	0.95	0.96	20	[30,328]	0.98	0.00
TSOM (97^{th} perc)	1.00	0.92	0.95	23	[41,323]	0.97	0.00

Table 4. Application of TSOM: assessment of the longshore flux of wave energy in shallow-water P. $\overline{P}$ is the mean over the 1979-2008 period accounting for the absolute value of P, $\overline{P}_+$ is the mean of the positive P, and $\overline{P}_-$ is the mean of the negative P, $\Delta_{TSOM-SOM}$ is the relative difference of values computed after TSOM with respect to values computed after SOM.

	SOM (kW/m)	TSOM (kW/m)	$\Delta_{TSOM-SOM}$ ($\%$)
$\overline{P}$	0.52	0.57	9.0
$\overline{P}_+$	0.41	0.45	7.5
$\overline{P}_-$	-0.11	-0.13	16.5

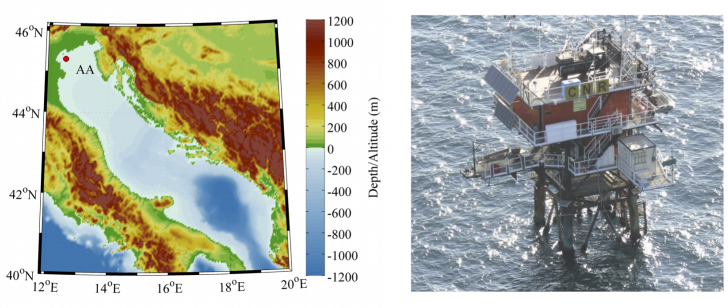

Figure 1. *Acqua Alta* (AA) oceanographic tower location in the northern Adriatic Sea, Italy (left panel). The tower is depicted in the right panel.

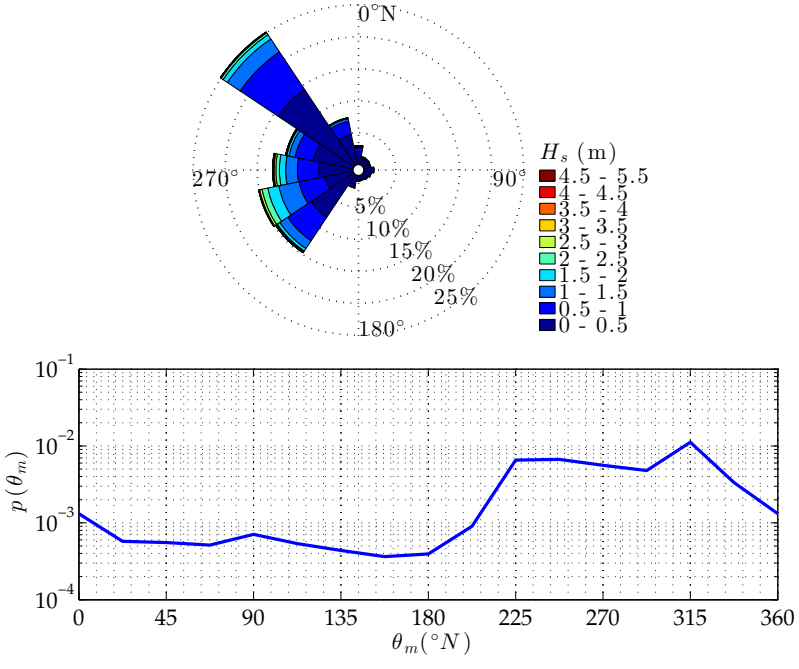

Figure 2. Observed bivariate wave climate at *Acqua Alta*: joint pdf of H_s and θ_m with radially distributed frequencies (top panel), and marginal pdf of θ_m (bottom panel). Resolutions are $\Delta H_s = 0.5$ m and $\Delta\theta_m = 22.5^\circ$.

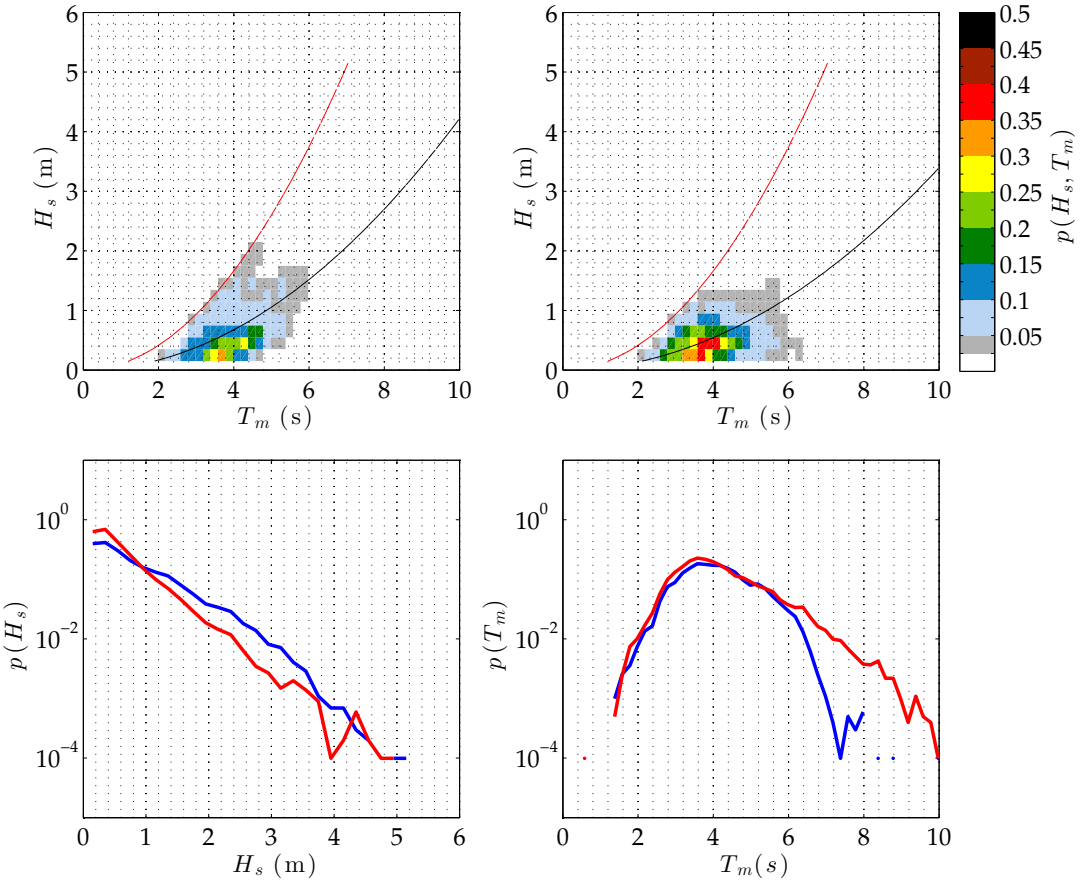

Figure 3. Observed bivariate wave climate at *Acqua Alta*: joint pdf of H_s and T_m for Bora (top-left panel) and Sirocco (top-right panel) sea states, and corresponding marginal pdfs of H_s (bottom-left panel; blue for Bora, red for Sirocco) and T_m (bottom-right panels; blue for Bora, red for Sirocco). Black solid lines in the top panels denote average wave steepness $2\pi H_s/g/T_m^2$ (3% for Bora, 2% for Sirocco, g being gravitational acceleration), red solid lines denote wave breaking limit (7%). Resolutions are $\Delta H_s = 0.2$ m and $\Delta T_m = 0.2$ s.

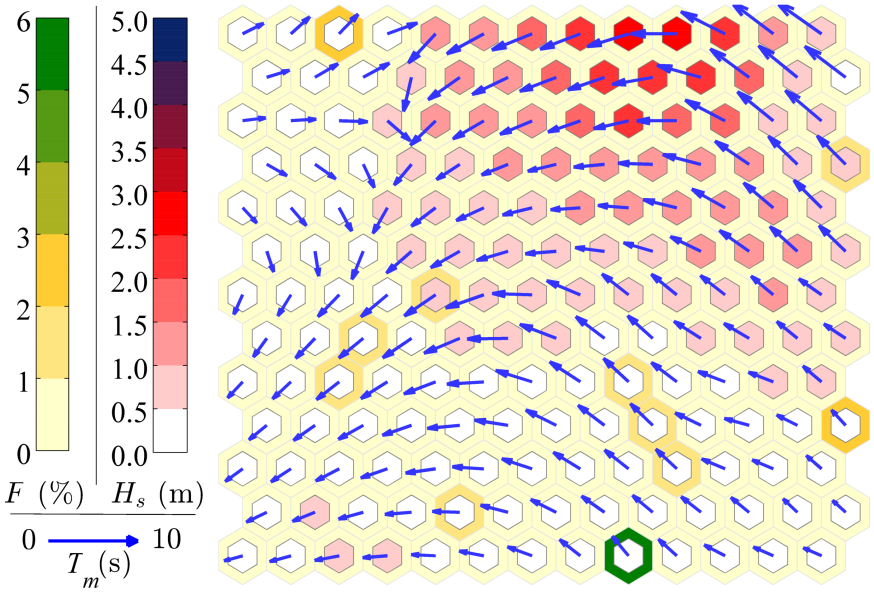

Figure 4. Single-step SOM output map. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Mean quantization error: 0.06; topographic error: 22%.

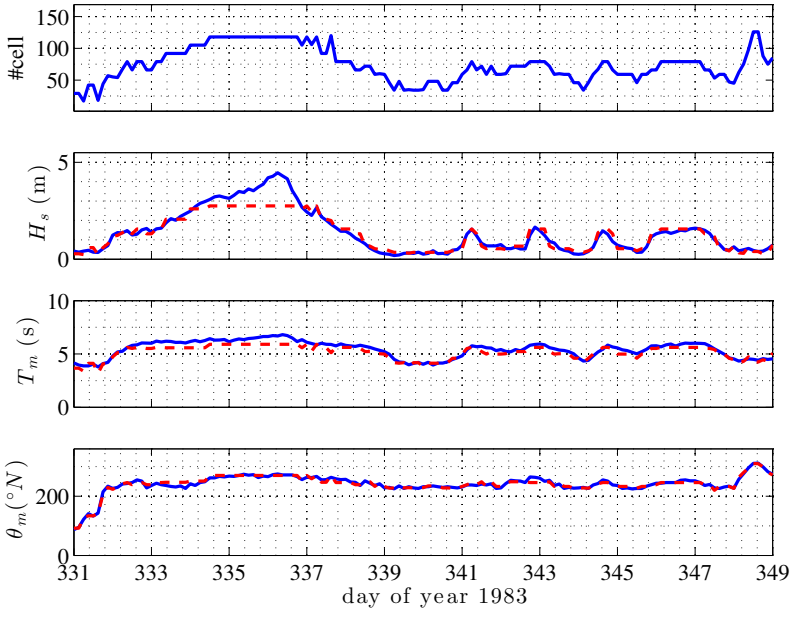

Figure 5. Single-step SOM: BMU cells (top panel) and comparison between original (blue solid lines) and reconstructed (red dashed lines) time series of H_s (central-top panel), T_m (central-bottom panel) and θ_m (bottom panel), for a chosen sequence of events.

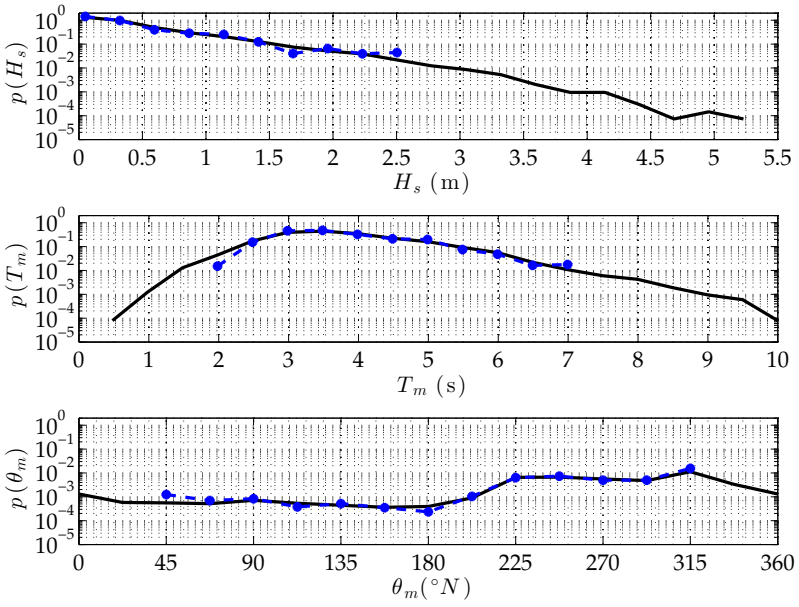

Figure 6. Single-step SOM: comparison of original (black solid line) and resulting (blue dashed-circles) pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for the whole dataset.

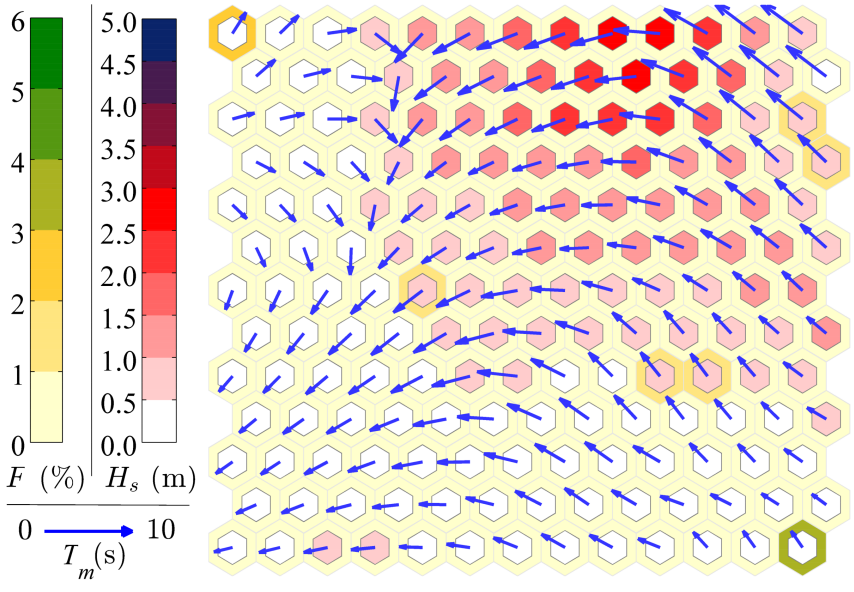

Figure 7. MDA-SOM output map, 80% reduction of the original dataset. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Mean quantization error: 0.06; topographic error: 15%.

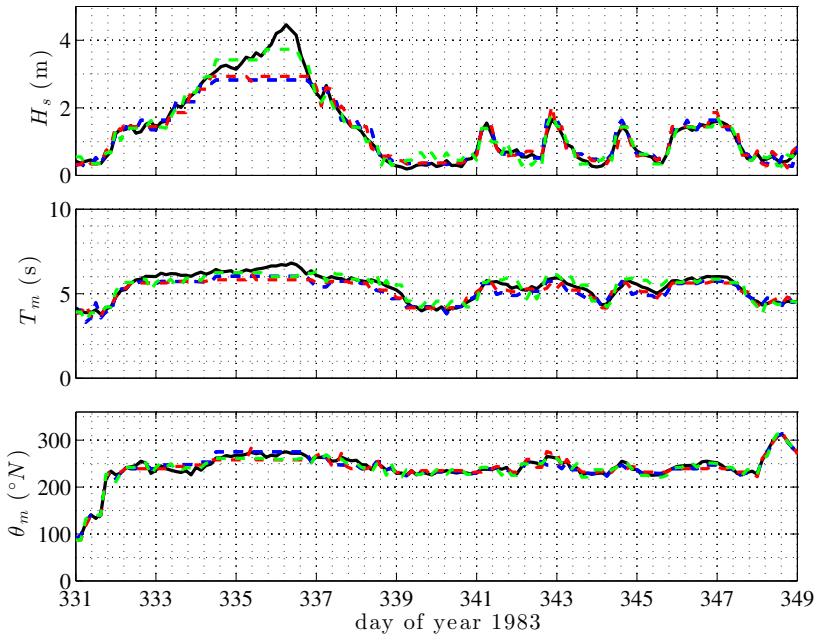

Figure 8. MDA-SOM: comparison between original (black solid lines) and reconstructed time series of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for a chosen sequence of events. Dataset reduction: 80% (blue dashed line), 60% (red dashed line) and 10% (green dashed line).

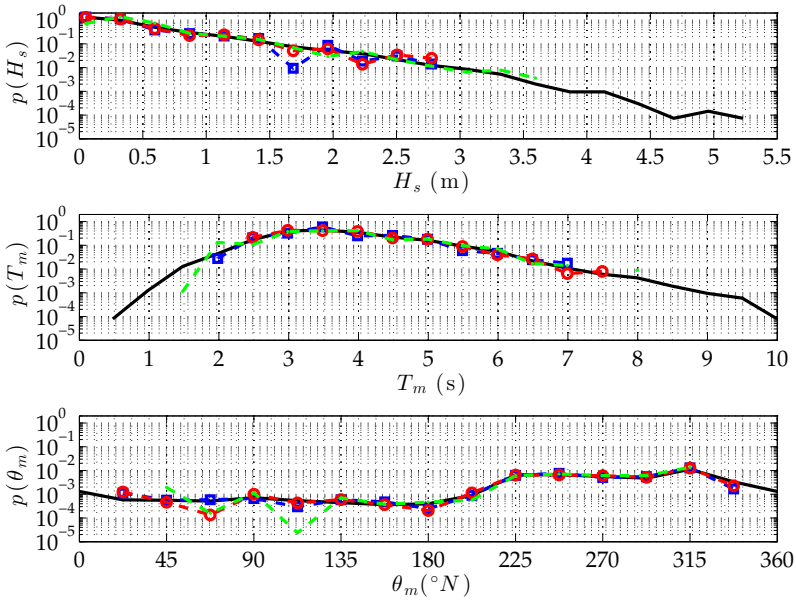

Figure 9. MDA-SOM: comparison between original (black solid lines) and resulting pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for the whole period of observations. Dataset reduction: 80% (blue dashed-squares line), 60% (red dashed-circles line) and 10% (green dashed-dots line).

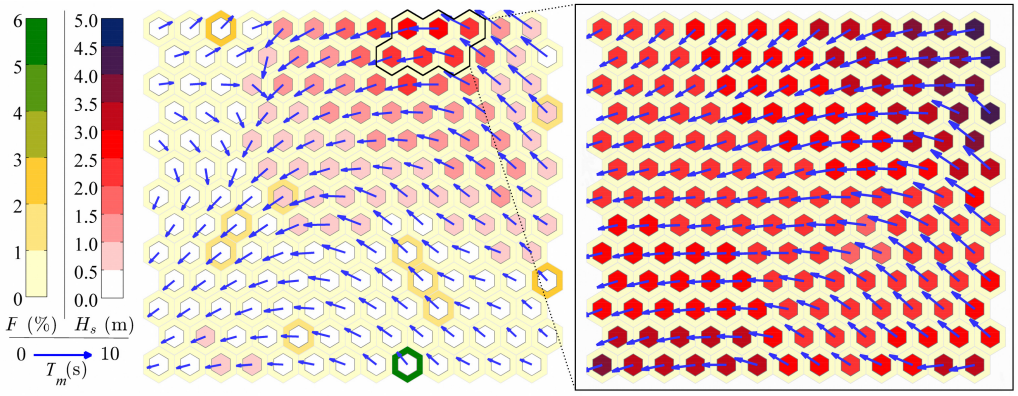

Figure 10. TSOM output map with threshold $H_s^* = 2.12$ m (97^{th} percentile of H_s). H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Wave climate after single-step SOM (left panel) and TSOM extreme wave climate (i.e. over the threshold, right panel and cells within black solid line in the left panel). For the right panel map, mean quantization error: 0.04; topographic error: 6%.

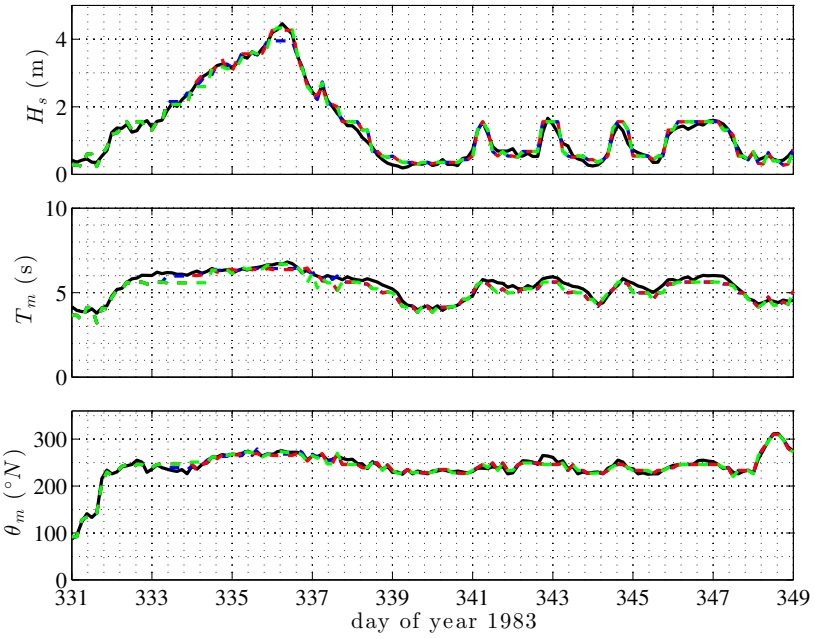

Figure 11. TSOM: comparison between original (black solid lines) and reconstructed time series of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for a chosen sequence of events. Thresholds: 95^{th} (blue dashed line), 97^{th} (red dashed line) and 99^{th} (green dashed line) percentile of H_s.

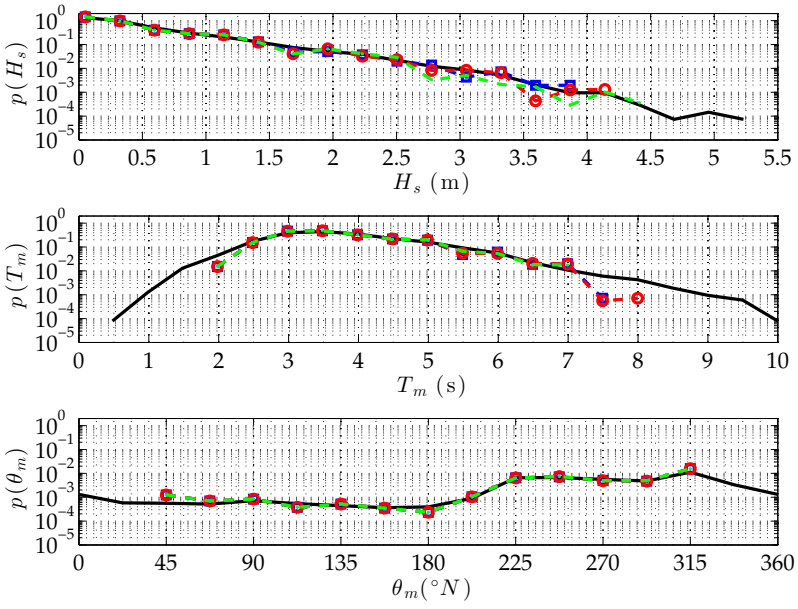

Figure 12. TSOM: comparison of original (black solid line) and resulting pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for the whole dataset. Thresholds: 95^{th} (blue dashed-squares line), 97^{th} (red dashed-circles line) and 99^{th} (green-dots dashed line) percentile of H_s.

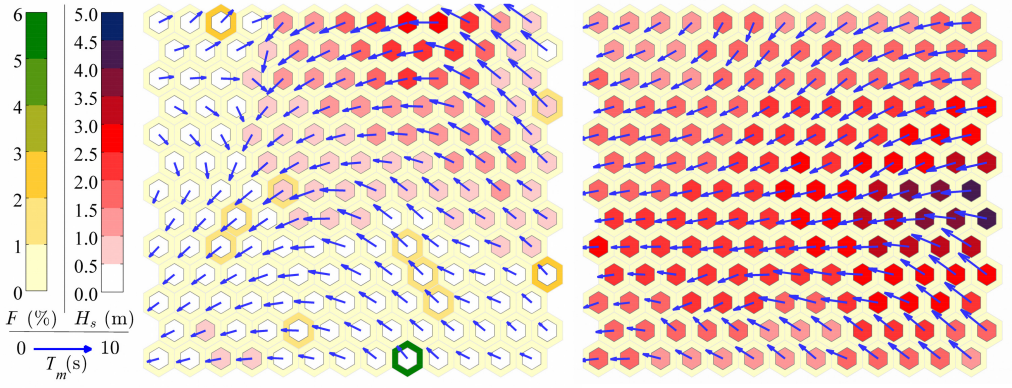

Figure 13. POT-SOM output map. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Wave climate after single-step SOM (left panel) and stormy wave climate (right panel). For the right panel map, mean quantization error: 0.06; topographic error: 12%.

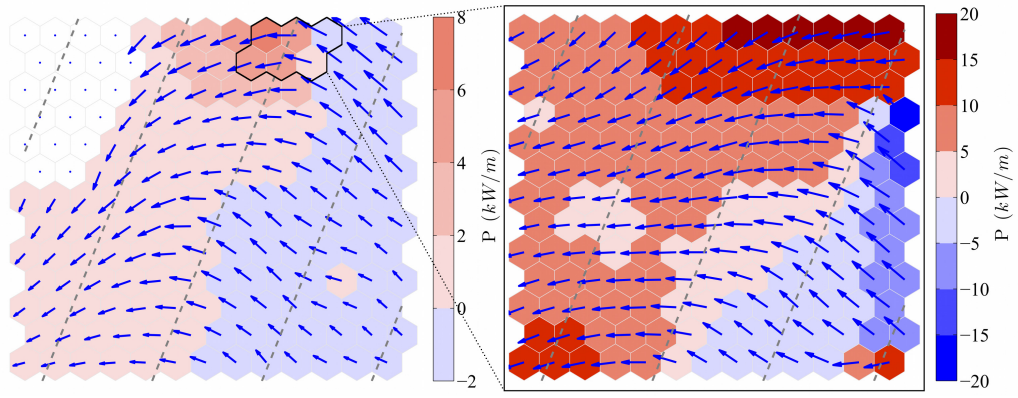

Figure 14. Application of TSOM: assessment of the longshore flux of wave energy P in shallow-water, after single-step SOM (left panel) and resulting from the TSOM extreme wave climate (right panel and cells within black solid line in the left panel). Mean wave directions at *Acqua Alta* tower (blue arrows) indicate contributions of different meteorological conditions: positive mainly due to Bora ($180 \leq \theta_m \leq 270°$N), negative to Sirocco ($270 < \theta_m \leq 360°$N). Land wind events (white cells) have been excluded, and the direction of the shoreline ($270°$N) is shown as gray dashed lines.