Ocean Science
Discussions

Open Access

Self-Organizing Maps approaches to analyze extremes of multivariate wave climate

F. Barbariol[1], F. M. Falcieri[1], C. Scotton[2], A. Benetazzo[1], S. Carniel[1], and M. Sclavo[1]

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

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

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

OSD

12, 1971–2006, 2015

SOM approaches for extremes of wave climate

F. Barbariol et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

In this paper the Self-Organizing Map (SOM) technique to assess the multivariate sea wave climate at a site is analyzed and discussed with the aim of a more complete representation which includes the most severe sea states that otherwise would be missed by the standard SOM. Indeed, it is commonly recognized, and herein confirmed, that SOM is a good regressor of a sample where the density of events is high (e.g. for low/moderate and frequent sea states), while SOM fails where the density is low (e.g. for severe and rare 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 the standard SOM have been tested in order to widen the range of applicability to extreme events. The first strategy contemplates a pre-processing of the input dataset with the Maximum Dissimilarity Algorithm; the second and the third strategies focus on the post-processing of SOM outputs, resulting in a two-steps SOM, where the first step is the standard SOM applied to the original dataset, and the second step is an additional SOM on the events exceeding a threshold (either taking all the events over the threshold or only the peaks of storms). Results suggest that post-processing strategies are more effective than the pre-processing one in representing the extreme wave climate, both in the time series and probability density spaces. In addition, a complete graphical representation of the outcomes of two-steps SOM as double-sided maps is proposed.

1 Introduction

The assessment of wave conditions at sea is fruitful for many research fields in marine and atmospheric sciences and for the human activities in the marine environment. To this end, in the last decades the observational network, that mostly relies on buoys, satellites and other probes from fixed platforms, has been integrated with numerical

models outputs, which allow to compute 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
5 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 hardly obtained. For individual waves, for instance, the bivariate joint pdf for wave height and period was derived by Longuet-Higgins (1983)
10 and the bivariate joint pdf for wave height and direction was obtained by Isobe (1988). A trivariate joint pdf for 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 for the significant wave height and the average zero-crossing wave period was derived by Ochi (1978),
15 and Mathisen and Bitner-Gregersen (1990).

Recently, the Self-Organizing Maps (SOM) technique has been successfully applied to represent the multivariate wave climate around the Iberian peninsula (Camus et al., 2011a, b) and the South America continent (Reguero et al., 2013). SOM (Kohonen, 2001) is an unsupervised neural network technique that classifies multivariate input
20 data and projects them onto a uni- or bi-dimensional output space, called a map. The SOM technique was originally developed in the 1980s, and over three decades has been largely applied in various fields, including oceanography (Falcieri et al., 2013; Solidoro et al., 2007). Typical applications of SOM are vector quantization, regression and clustering. SOM gained credit among other techniques with same applications due
25 to its visualization properties that allow to get multidimensional information from a 2-D 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. Hence, the most frequent conditions force the map to place clusters

there, while the most rare (i.e. the extreme events) are often missed. Consequently, it is highly unlike to find on the map a cluster representing the extremes.

In the context of ocean waves, drawing upon the works of Camus et al. (2011a, b) and Reguero et al. (2013), SOM input is typically constituted by a set of wave parameters measured or simulated at a given location and evolving over the time t, e.g. the triplet 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: selection of typical deep-water sea states for propagation towards the coast to study the longshore currents regime and coastal erosion, individuation of typical sea states for wave energy resource assessment and wave farm optimization. In addition the experimental 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 fact we have asked: how can we employ SOM with its visualization properties to improve representation of the edges of a multivariate wave climate at a location, i.e. extremes? To answer this question we have followed three different strategies. Firstly, we have pre-preprocessed the input data using the Maximum Dissimilarity Algorithm (MDA) in order to reduce the redundancies of 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 outputs. In this context, the improved extremes representation has been searched by running SOM on a dataset of events exceeding

Title Page

Abstract   Introduction

Conclusions   References

Tables   Figures

|◄   ►|

◄   ►

Back   Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

SOM approaches for
extremes of wave
climate

F. Barbariol et al.

a prescribed threshold. Hence, we have applied a two-steps SOM approach, by firstly running a standard SOM to get a representation of the low/moderate (i.e. the most frequent) wave climate, and then we have run a second SOM on a reduced input sample. This new sample has been obtained in one case by taking the events that exceed a prescribed threshold (e.g. 97th percentile of H_s, calling this strategy TSOM) from first step SOM results, and in another case by taking the peak of the storms, individuated by means of a Peak-Over-Threshold analysis (calling this strategy POT-TSOM, Boccotti, 2000). To present results of two-steps SOMs, we propose a double-sided map, which shows on the left the map with the low/moderate sea states, while on the right the map with the most severe sea states (i.e. the extremes).

2   Data

The dataset employed for the SOM analysis consists of wave time series gathered at the *Acqua Alta* oceanographic tower. *Acqua Alta*, owned and operated by the "Italian National Research Council – Institute of Marine Sciences" (CNR-ISMAR), is located in the northern Adriatic Sea (Italy, northern Mediterranean Sea), 8 miles off the Venice coast at 17 m depth (Fig. 1). The tower 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). 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, that were measured by 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 50 503 sea states.

Basic statistics of the time series, given in 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 can reach severe levels ($\max(H_s) = 5.23$ m). Such severe events

are quite rare, as confirmed by the 99th 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 experimental pdf of θ_m (Fig. 2, bottom panel), where the most frequent directions of propagation are indeed in the range $180 < \theta_m < 360°$ 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 attained from the bivariate diagram ($H_s - \theta_m$) representing the experimental joint pdf of H_s and θ_m (Fig. 2, upper panel): 247.5, 270 and 315° N. Mild sea states and calms ($H_s < 1.5\langle H_s \rangle$, Boccotti, 2000) are the most frequent conditions at *Acqua Alta*, with 80 % of occurrence during the 30 years. 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 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 results that "Bora" storms have an occurrence of 12 %, while "Sirocco" storms of 8 %.

The bivariate wave diagram (H_s–T_m) illustrates the experimental joint pdf of H_s and T_m (Fig. 3), where it emerges that the most common sea states are represented by $[H_s, T_m] = [0.2$ m, 3.6 s]. This diagram also displays that the $[H_s, T_m]$ pairs associated to extreme sea states are [4.8 m, 7.4 s], [5.0 m, 7.2 s] and [5.0 m, 5.8 s]. The experimental

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

marginal pdfs of H_s and T_m are also shown in Fig. 3, confirming that the most probable H_s at *Acqua Alta* is 0.2 m, and the most probable T_m is 3.6 s.

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 the 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 a 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 set as sheet, cylindrical or toroidal. The local structure of the lattice is defined by the shape of the elements, called units, that can be 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 the learning functions and parameters is crucial to obtain stable and reliable maps. In SOM, the classification of input data is performed by means of competitive-cooperative learning: 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, the chosen metric is the Euclidean distance, and 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 inhibi-

tion 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 as done by the original incremental SOM algorithm, where input data are presented one at a time to the map, or batchwise as done by the more recent batch algorithm, where the whole input data set is presented 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 2-D sequence of vectors obtained from the eigenvalues and eigenvectors of the covariance matrix of the data. In both the 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, the degree of optimality of the resulting map has to be assessed in other ways, e.g. by means of specific error metrics. The most common ones used for SOM 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 do not have adjacent first (the BMU) and second units in the map and is a measure of the topological preservation of the map. In this context, in order to achieve reliable and stable maps, it is crucial to follow practical advices and to keep error metrics under control.

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 the 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, BMUs (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, SOM have produced outputs as bi-dimensional squared lattices, 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 13×13 (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. The process has been performed once, without considering a rough phase followed by fine-tuning. Indeed, this is the condition that have produced the lowest mean quantization and topographic errors, hence the most reliable maps.

Input data have been normalized in oder that the minimum and maximum distance between two realizations of a variable are 0 and 1, respectively. To this end, according

SOM approaches for extremes of wave climate

F. Barbariol et al.

to Camus et al. (2011a), the following normalizations have been used:

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

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:

$$d_{ij} = \left\{ \left(H_i - \overline{H}_j\right)^2 + \left(T_i - \overline{T}_j\right)^2 + \left(\min\left[|\theta_i - \overline{\theta}_j|, \quad 2 - |\theta_i - \overline{\theta}_j|\right]\right)^2 \right\}^{0.5} \qquad (2)$$

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

4 Results

In this Section, results of the standard SOM approach and of the different strategies proposed to improve extremes representation are presented. The performances of standard SOM, MDA-SOM and TSOM are assessed by comparing the wave parameters time series and their experimental marginal pdfs to the time series reconstructed from the results of the different strategies and relative pdfs, respectively. POT-TSOM is treated separately because a direct comparison with the other strategies using the described methods is not possible.

4.1 Standard SOM

Standard SOM has been applied using the set-up illustrated in Sect. 3.2. The SOM output map in Fig. 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, θ_m] triplet. Hence, one can have an immediate sight on the wave climate features and on the experimental joint pdf thanks to the powerful visual properties of SOM's output.

5   Gradual and continuous changing in wave parameters over the BMUs 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 resembles the information that one could have gather from the bivariate (H_s–T_m) and (H_s–θ_m) diagrams, i.e. [0.2 m, 0.36 s, 315° N], though
10   these two diagrams are not formally related one to the other. The most of the BMUs show wave propagation directions that point towards the western quadrants, as displayed in the joint and marginal pdfs of θ_m. The BMUs denoting sea states forced by land winds (pointing toward east) are clustered in the top-left corner of the map and have very low frequencies of occurrence (individual and cumulated). The frequency of
15   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 Sect. 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.

20   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 3.0 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 Sect. 2 have shown that at *Acqua Alta* H_s can exceed 5.0 m. Therefore sea states with $H_s > 2.75$ m are represented by BMUs with lower H_s. This is clear in Fig. 5, where a se-
25   quence 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 representative BMU. In Fig. 5 sea states with $H_s > 2.75$ m are represented by BMU #118 (the one with the highest H_s), 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, standard 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 Fig. 6 displaying the marginal experimental pdfs of H_s, T_m and θ_m after SOM, compared to the original pdfs. Generally, marginal experimental 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 to 0.17–2.75 m, the range of T_m from 0.5–10.1 to 2.4–7.4 s, and the range of θ_m from 0–360 to 41–323° N. The maximum H_s value given by SOM (2.75 m) is pretty close to the 99th percentile value (2.68 m), pointing out that SOM provides a good representation of the wave climate up to the 99th 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 standard 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 99th 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.

SOM approaches for extremes of wave climate

F. Barbariol et al.

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 50 503 sea states (100 %) to 5050 (10 %), with step 10 %. The absolute errors on $\langle H_s \rangle$ and on the 99th 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 99th 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 $\langle H_s \rangle$, 40 % error on 99th percentile H_s). The corresponding MDA-SOM output map displayed in Fig. 7 is topologically equivalent to that produced by SOM (Fig. 4), except for minor differences on the location of some BMUs. However, the most frequent BMU has $[H_s, T_m, \theta_m] = [0.28\,\text{m}, 2.8\,\text{s}, 328^\circ\,\text{N}]$, which is more distant to what have emerged from diagrams of Sect. 2, with respect to standard SOM, especially for T_m. Also, the BMU with highest H_s has the triplet equal to $[2.8\,\text{m}, 6.0\,\text{s}, 275^\circ\,\text{N}]$, 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 Fig. 8, for the sequence of events already shown, the comparison has been extended to the results of 60 % MDA-SOM (smaller error on 99th percentile H_s, see Table 7) 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 $(2.93\,\text{m}, 5.8\,\text{s}, 258^\circ\,\text{N})$. 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 experimental pdfs after MDA-SOM are compared in Fig. 9 to the pdfs of the original dataset. The distributions are in good agreement and the representation is more complete with respect to standard SOM, especially concerning H_s. Neverthe-less, 10 % MDA-SOM distribution for H_s exhibits a larger departure from the original distribution at 1.7 m with respect to standard 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 observed at *Acqua Alta*.

4.3 Two-steps SOM (TSOM)

A two-step SOM (TSOM) has been then applied to provide a more complete descrip-tion 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^* was fixed, and the BMUs having $H_s > H_s^*$ have been considered to constitute a new input dataset that is composed of the sea states represented by the BMUs exceeding the threshold. Hence, a second SOM has been run on the new dataset (second step). The SOM set-up is the same as in the first step. At the end, we have obtained a two-sided map that represents the wave climate (Fig. 10, for instance): the first map on the left side describes the climate below H_s^*, the second map on the right side focuses on the climate over H_s^*. Three thresholds were tested that correspond to the 95th, 97th and 99th percentile of H_s: 1.80, 2.12 and 2.68 m, respectively. In the following, we have focused on the results with 97th per-centile 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 (97th percentile). The first map, on the left, is the map already shown in Fig. 4, representing the wave climate at *Acqua Alta*. On that map, the BMUs with $H_s > 2.12$ m have been encompassed by a black line.

Without such BMUs, the map on the left represents the low/moderate wave climate, 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 BMUs, 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 tight directional sector (50°) making hard to discriminate between "Bora" and "Sirocco" events. A more detailed representation of such extreme events is provided by the second map in Fig. 10, on the right, where extreme "Bora" and "Sirocco" events are more widely described by BMUs. Indeed, a sort of diagonal (from the top-right corner to the bottom-left corner of the map) divides the BMUs. "Bora" events are clustered on the left of this diagonal (top-left part of the map), while "Sirocco" events on the right of that (bottom-left part of the map). On the diagonal, BMUs 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 are $1.94 < H_s < 4.26$ m, $4.4 < T_m < 8.3$ s and $224 < \theta_m < 316°$ N.

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 Fig. 11. There, for the sequence of events presented in the previous Sections, the reconstructed TSOM time-series is compared to the original one. Also results with 95th and 99th percentile TSOM are plotted, and it clearly appears that the differences among the three tests (i.e. TSOM with H_s threshold on 95th, 97th and 99th percentiles) are very small, in particular for what concerns θ_m. Nevertheless, 95th percentile TSOM yield to a smaller estimate of the highest H_s peak with respect to the others, and 99th percentile TSOM deviates more than the others from the original T_m.

Such differences are found also in the marginal experimental pdfs of the wave parameters, shown in Fig. 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, 99th 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 97th percentile and the tightest with 99th percentile TSOM. Instead, $p(\theta_\mathrm{m})$ 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 Comparison among strategies

A summary of the performances of the different SOM strategies we tested until here is given in Table 3. There, standard SOM, MDA-SOM with 80 % reduction and TSOM with H_s threshold at 97th percentile are compared in their capabilities of representing the wave climate at *Acqua Alta* by means of the BMUs. 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 one, as shown by the high r_av (over 0.98) and r_SD (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). Though, the highest ratios and correlation coefficients, and lowest RMSE pertain to TSOM. Similar conclusions can be drawn for the pdfs, which are reproduced with very high CC (over 0.95) and $\mathrm{RMSE}_\mathrm{pdf}$ (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.

5   Peak-Over-Threshold Two-steps SOM (POT-TSOM)

As an additional strategy to provide a more complete representation of the wave climate through SOM, we tested a different approach. It is based on a two-step SOM applied initially to the whole dataset, to classify the low/moderate wave climate, and then applied to 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 h over a given threshold H_s^* that corresponds 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 storm constitutes a new dataset that has been analyzed by means of SOM. At the end, we have obtained a double-side map that represent both the low/moderate (on the left) and the "stormy" (on the right) wave climate. POT-TSOM is not directly comparable to the other strategies since the dataset used for the second step of the two-steps SOM is composed by the storms peaks only, thus the global dataset represented by SOM is not continuous.

POT-TSOM output map is shown in Fig. 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 99th 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 and standard SOM.

SOM approaches for extremes of wave climate

F. Barbariol et al.

6 Conclusions

In this paper, we have tested different strategies aimed at providing a complete description of the 1979–2008 trivariate wave climate at *Acqua Alta* tower using SOM. Indeed, we have verified that besides a satisfactory description of the multivariate low/moderate wave climate (in agreement with usual uni- and bivariate diagrams), standard SOM approach misses the most severe sea states, which are hidden in BMUs with H_s even considerably smaller than the extreme ones. For instance, standard SOM classified as [2.75 m, 5.9 s, 270° N] most of the sea states with $H_s > 2.75$. Hence, the most interesting part of the wave climate was condensed within a few BMUs which hardly allow to discriminate between "Bora" and "Sirocco" events. To increase the weight of the most severe and rare events, 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 standard SOM results (i.e. TSOM). Results presented showed that the post-processing technique is more effective than the pre-processing one. Indeed, a wider range of the wave parameters within BMUs was obtained, in particular for H_s. This allowed a representation of the sea states with TSOM more accurate and complete with respect to the one furnished by MDA-SOM, both for time series and marginal experimental 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 properties. In this context, we proposed a double-sided map representation for two-steps SOM strategies, which summarizes on the left the low/moderate wave climate and on the right the most severe one. This novel representation was employed also to show the results of the POT-TSOM, which classified the storms peaks at *Acqua Alta* tower.

References

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, doi:10.5194/nhess-12-2065-2012, 2012. 1976

Boccotti, P.: Wave Mechanics for Ocean Engineering, vol. 64, Elsevier Science, Amsterdam, the Netherlands, 2000. 1975, 1976, 1987

Camus, P., Cofiño, A. S., Mendez, F. J., and Medina, R.: Multivariate wave climate using self-organizing maps, J. Atmos. Ocean. Tech., 28, 1554–1568, doi:10.1175/JTECH-D-11-00027.1, 2011a. 1973, 1974, 1977, 1980, 1982, 1983

Camus, P., Mendez, F. J., Medina, R., and Cofiño, A. S.: Analysis of clustering and selection algorithms for the study of multivariate wave climate, Coast. Eng., 58, 453–462, doi:10.1016/j.coastaleng.2011.02.003, 2011b. 1973, 1974

Cavaleri, L.: The oceanographic tower Acqua Alta – activity and prediction of sea states at Venice, Coast. Eng., 39, 29–70, doi:10.1016/S0378-3839(99)00053-8, 2000. 1975

Falcieri, F. M., Benetazzo, A., Sclavo, M., Russo, A., and Carniel, S.: Po River plume pattern variability investigated from model data, Cont. Shelf. Res., 87, 84–95, doi:10.1016/j.csr.2013.11.001, 2013. 1973

Goring, D. G. and Nikora, V. I.: Despiking acoustic Doppler velocimeter data, J. Hydraul. Eng.-ASCE, 128, 117–126, doi:10.1061/(ASCE)0733-9429(2002)128:1(117), 2002. 1975

Isobe, M.: On joint distribution of wave heights and directions, in: Proceedings of the 21st International Conference on Coastal Engineering, 20–25 June 1988, Costa del Sol, Malaga, Spain, ASCE, New York, USA, 1, 524–538, 1988. 1973

Kohonen, T.: Self-Organizing Maps, vol. 30, doi:10.1007/978-3-642-56927-2, Springer-Verlag, Berlin-Heidelberg, Germany, 2001. 1973, 1977, 1978, 1979

Kohonen, T., Nieminen, I. T., and Timo, H.: On the quantization error in SOM vs. VQ: a critical and systematic study, in: Advances in Self-Organizing Maps, Springer, Berlin, Germany, 133–144, 2009. 1978, 1979

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: Proceedings of the 24th International Conference on Coastal Engineering, 23–28 October 1994, Kobe, Japan, ASCE, New York, USA, 370–383, 1994. 1973

5   Longuet-Higgins, M. S.: On the joint distribution of wave periods and amplitudes in a random wave field, P. Roy. Soc. A-Math. Phy., 389, 241–258, 1983. 1973

Mathisen, J. and Bitner-Gregersen, E.: Joint distributions for significant wave height and wave zero-up-crossing period, Appl. Ocean Res., 12, 93–103, doi:10.1016/S0141-1187(05)80033-1, 1990. 1973

10   Ochi, M. K.: On long-term statistics for ocean and coastal waves, in: Proceedings of the 16th International Conference on Coastal Engineering, 27 August–3 September 1978, Hamburg, Germany, ASCE, New York, USA, 1, 59–75, 1978. 1973

Reguero, B. G., Méndez, F. J., and Losada, I. J.: Variability of multivariate wave climate in Latin America and the Caribbean, Global Planet. Change, 100, 70–84, doi:10.1016/j.gloplacha.2012.09.005, 2013. 1973, 1974

15   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, J. Marine Syst., 53, 217–233, 2005. 1976

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, J. Geophys. Res., 112, 1–13, doi:10.1029/2006JC003553, 2007. 1973

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

25   techrep.pdf (last access: 16 July 2015), 2000. 1979

Yin, H.: The Self-Organizing Maps: background, theories, extensions and applications, in: Computational Intelligence: A Compendium, Springer, Berlin, 715–760, 2008. 1978

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Statistics of the 1979–2008 wave parameters dataset at *Acqua Alta* tower ($\langle - \rangle$: mean, SD: standard deviation, min: minimum, perc: percentile, max: maximum).

	$\langle - \rangle$	SD	min	50th perc	95th perc	97th perc	99th 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 99th 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
99th percentile	9	8	4	5	3	3	5	5	18	27

Table 3. Performance summary of different SOM approaches, through the comparisons of re-constructed to original time series, and resulting to original pdfs. r_{av}: time series ratio of averages, r_{SD}: time-series ratio of 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_{SD}	CC	RMSE (m)	range (m)	CC_{pdf}	$RMSE_{pdf}$
Standard 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 (97th perc)	0.99	0.95	0.96	0.16	0.17–4.26	1.00	0.04
T_m	r_{av}	r_{SD}	CC	RMSE (s)	range (s)	CC_{pdf}	$RMSE_{pdf}$
Standard 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 (97th perc)	1.00	0.90	0.95	0.32	2.4–8.3	0.99	0.02
θ_m	r_{av}	r_{SD}	CC	RMSE (°N)	range (°N)	CC_{pdf}	$RMSE_{pdf}$
Standard 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 (97th perc)	1.00	0.92	0.95	23	41–323	0.97	0.00

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
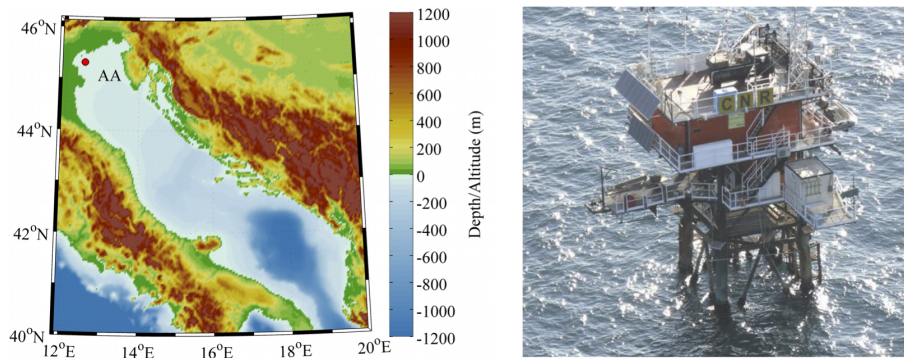
Figure 1. *Acqua Alta* (AA) oceanographic tower location in the northern Adriatic Sea, Italy.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper
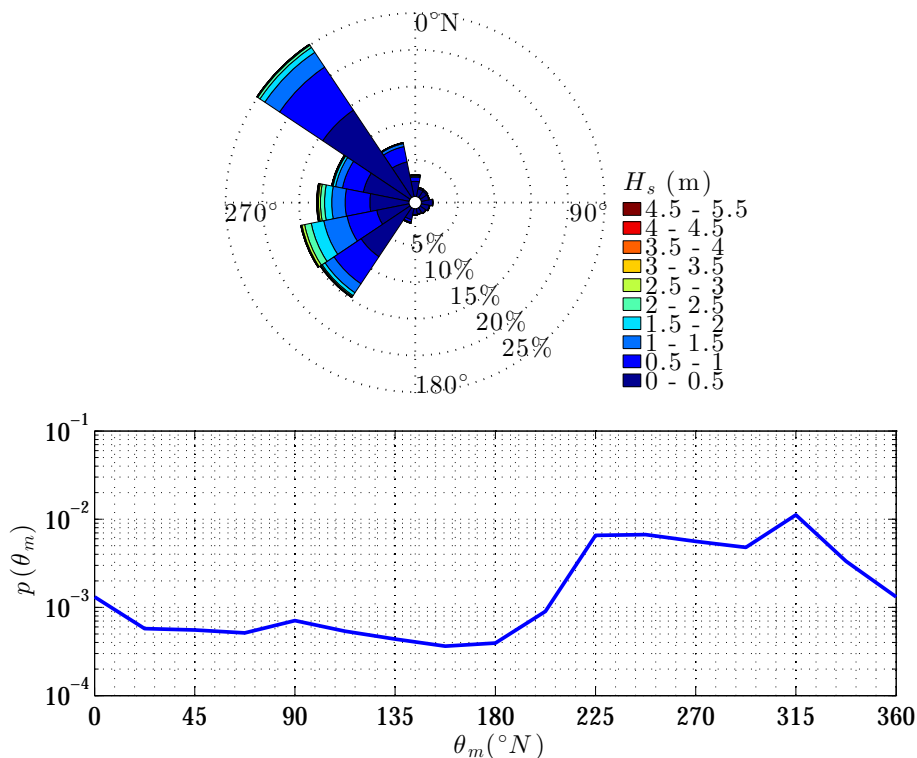
Figure 2. Bivariate wave climate at *Acqua Alta* tower representing the joint (H_s–θ_m) experimental pdf with radially distributed frequencies (top panel); resolutions are $\Delta H_s = 0.5$ m and $\Delta \theta_m = 22.5°$. Experimental marginal θ_m pdf at *Acqua Alta* tower (bottom panel).
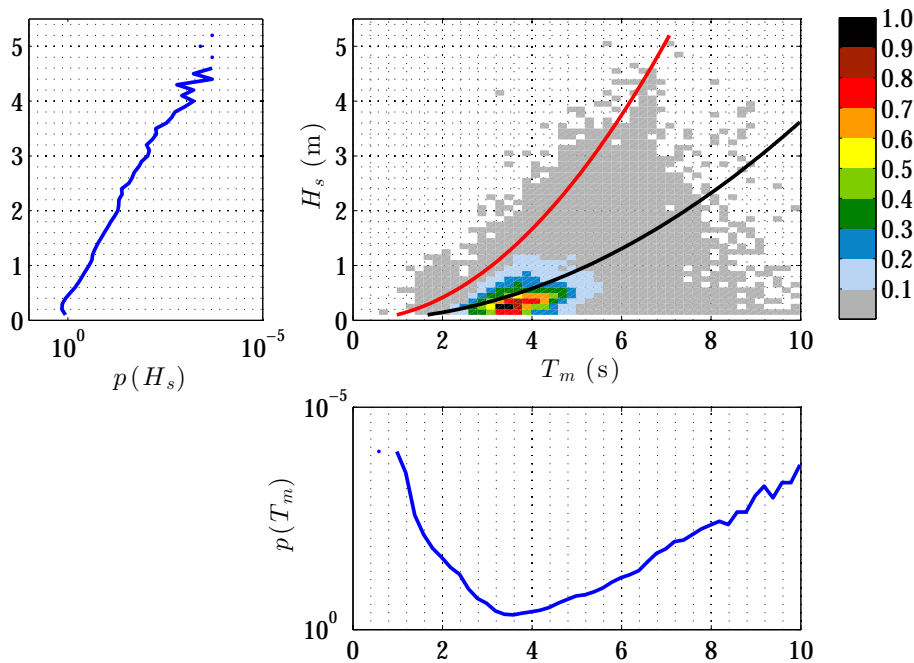
Figure 3. Bivariate wave climate at *Acqua Alta* tower representing the joint (H_s–T_m) experimental pdf (top-right panel); resolutions are $\Delta H_s = 0.1$ m and $\Delta T_m = 0.2$ s. Experimental marginal H_s and T_m pdfs at *Acqua Alta* tower (top-left and bottom-right panels, respectively). Solid lines in the top-right panel denote wave steepness $2\pi H_s/g/T_m^2$ equal to 2 % (average observed, black) and 7 % (breaking limit, red), *g* being gravitational acceleration.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

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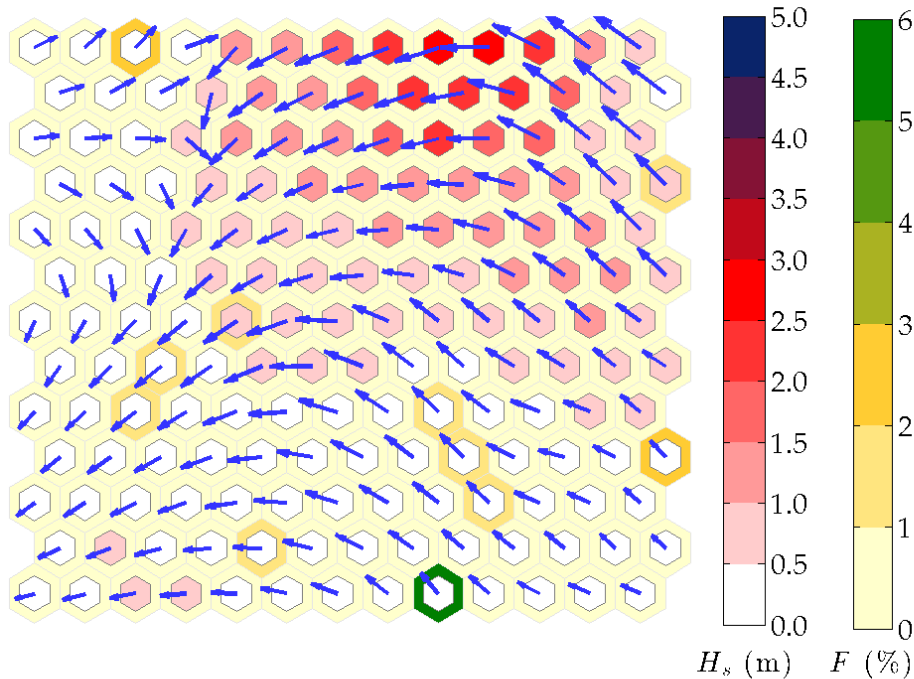
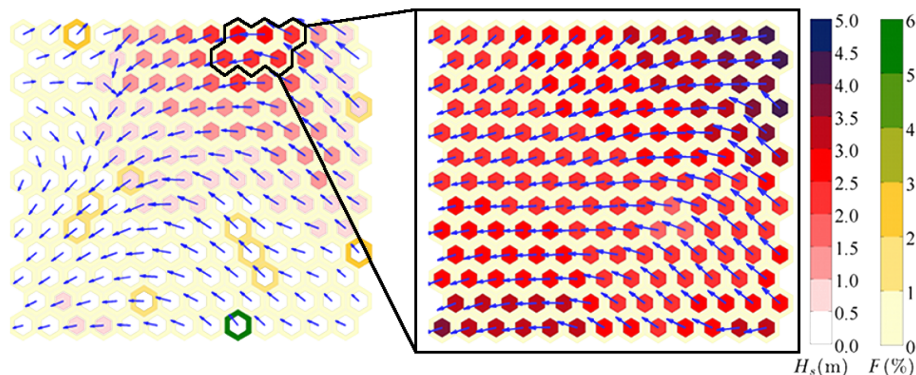**Figure 4.** Strandard SOM output map. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Mean quantization error: 0.06; topographic error: 22 %.

SOM approaches for extremes of wave climate

F. Barbariol et al.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
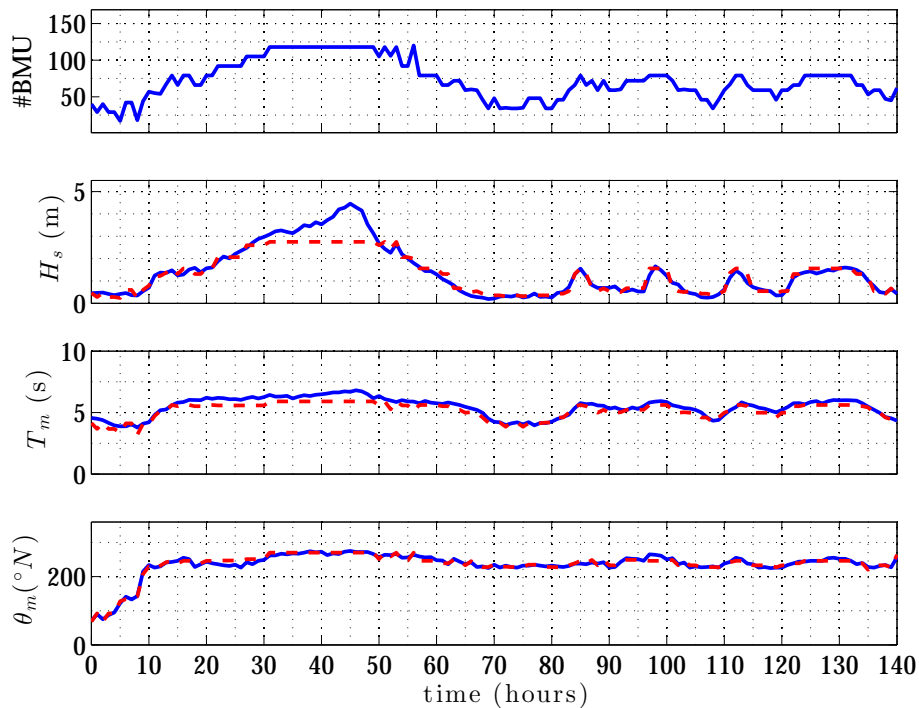
Figure 5. Strandard SOM: BMU (top panel) and comparison between original (blue solid lines) and reconstructed (red dashed lines) time series of H_s (central-top panel), T_m (central-bottom panel) and θ_m (bottom panel).

Full Screen / Esc
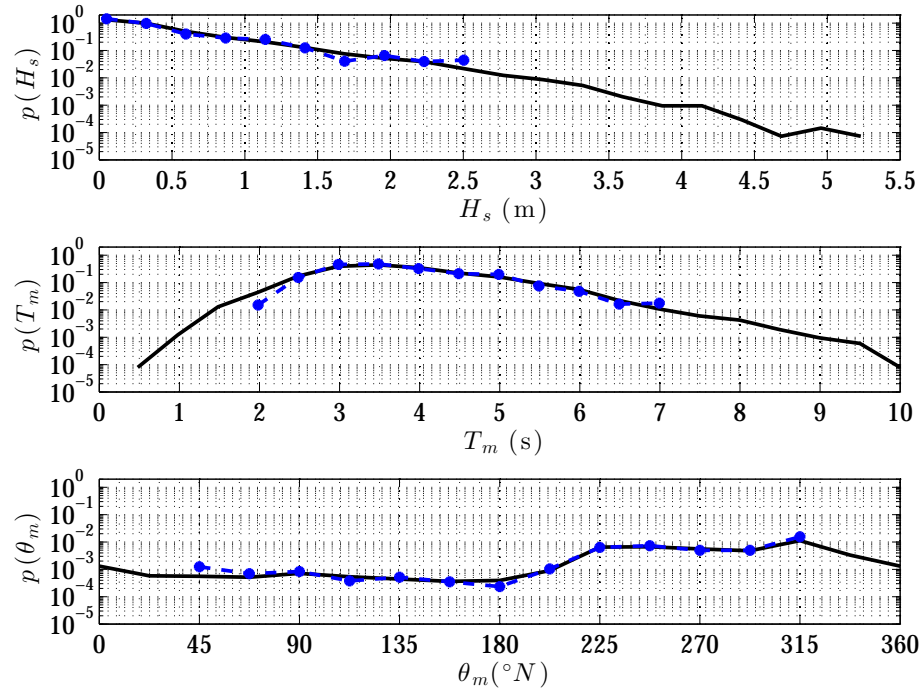
Figure 6. Strandard SOM: comparison of original (black solid line) and resulting (blue dashed-circles) pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel).
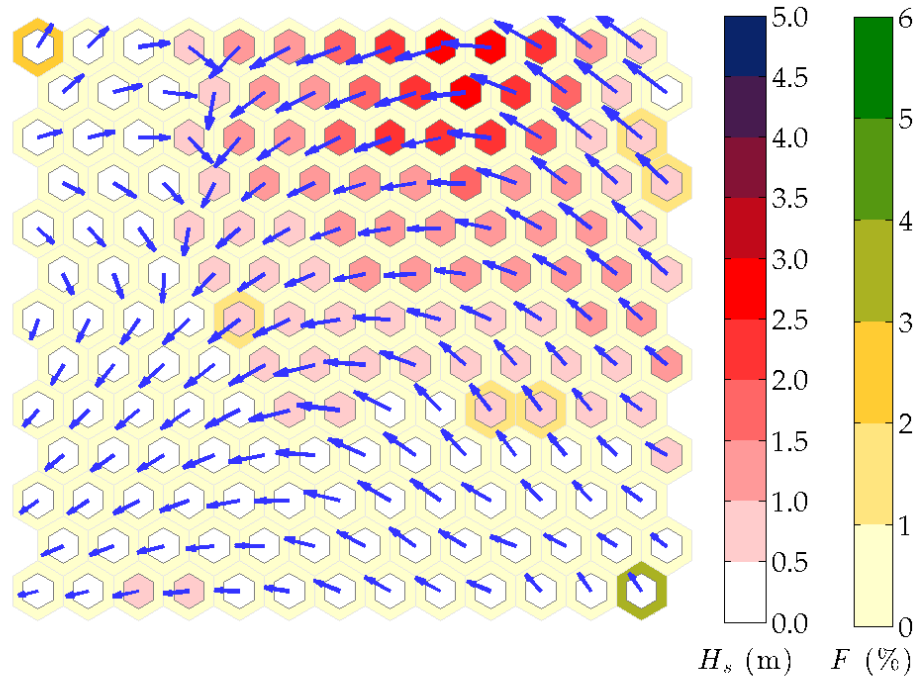
Figure 7. MDA-SOM output map, 80 % reduction of the original dataset. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Mean quantization error: 0.06; topographic error: 15 %.

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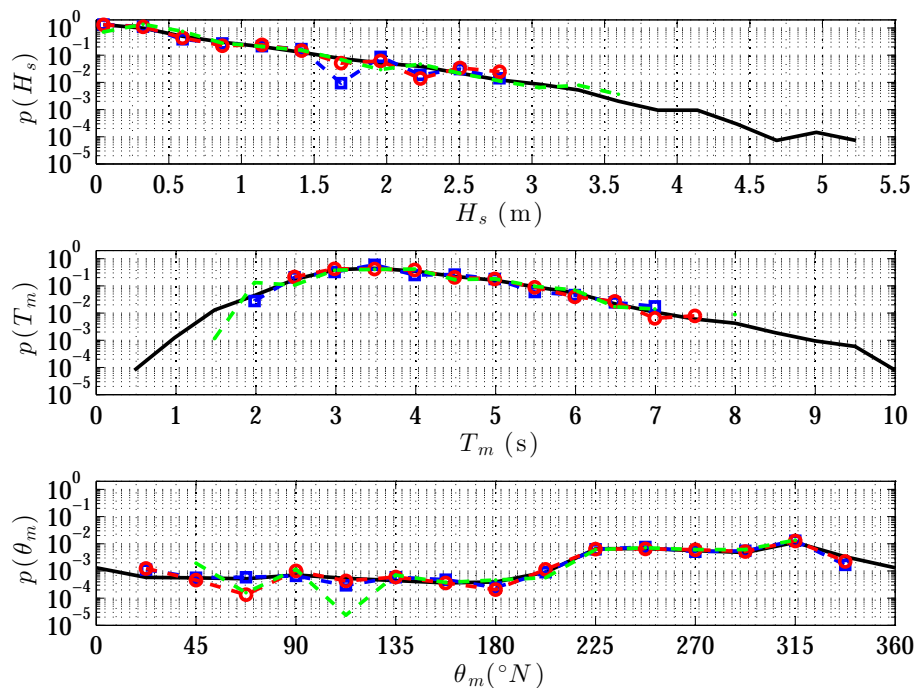
Figure 9. MDA-SOM: comparison between original (black solid lines) and resulting pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel): 80 % (blue dashed-squares line), 60 % (red dashed-circles line) and 10 % (green dashed-dots line).

Figure 10. TSOM output map with threshold $H_s^* = 2.12\,\text{m}$ (97th percentile of H_s). H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Wave climate below the threshold (left panel) and over the threshold (right panel, BMUs within black solid line in the left panel). For the right panel map, mean quantization error: 0.04; topographic error: 6%.
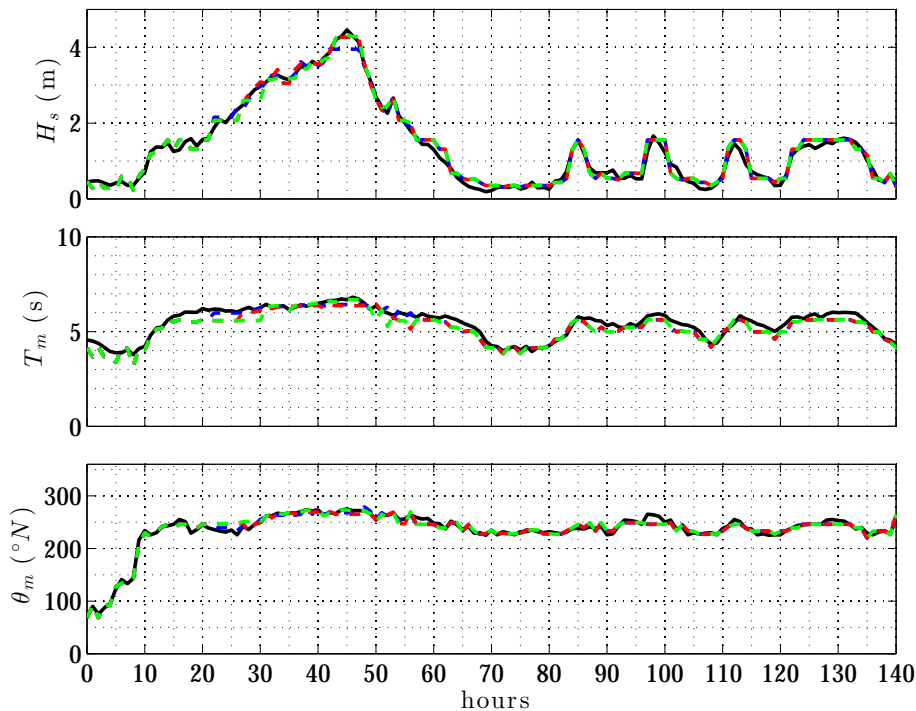
Figure 11. TSOM: comparison between original (black solid lines) and reconstructed time series of H_s (top panel), T_m (central panel) and θ_m (bottom panel): 95th percentile (blue dashed line), 97th percentile (red dashed line) and 99th percentile (green dashed line) H_s thresholds.
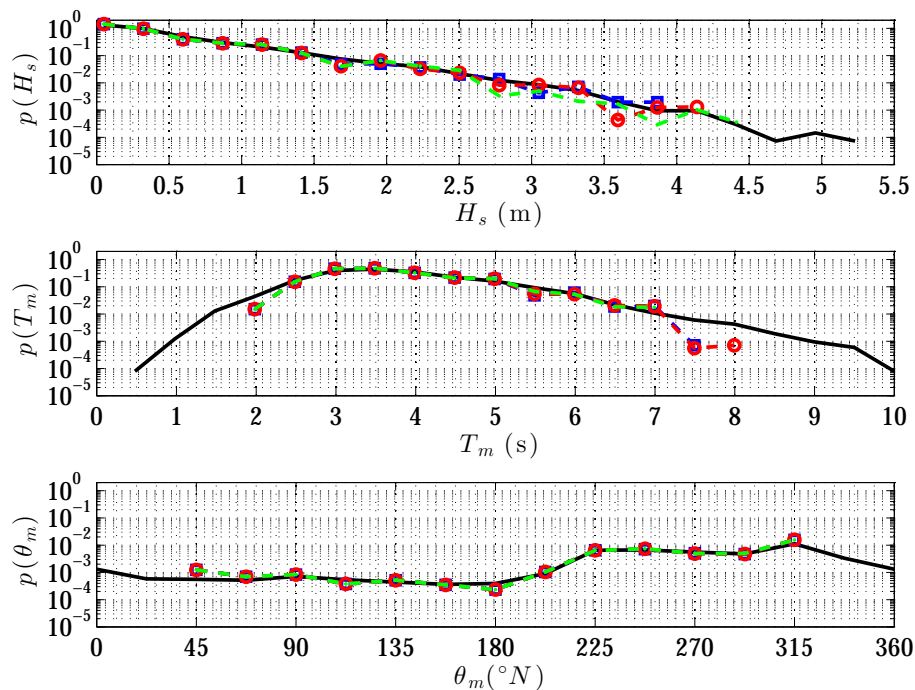
Figure 12. TSOM: comparison of original (black solid line) and resulting pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel): 95th percentile (blue dashed-squares line), 97th percentile (red dashed-circles line) and 99th percentile (green-dots dashed line) H_s thresholds.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

OSD

12, 1971–2006, 2015

SOM approaches for extremes of wave climate

F. Barbariol et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

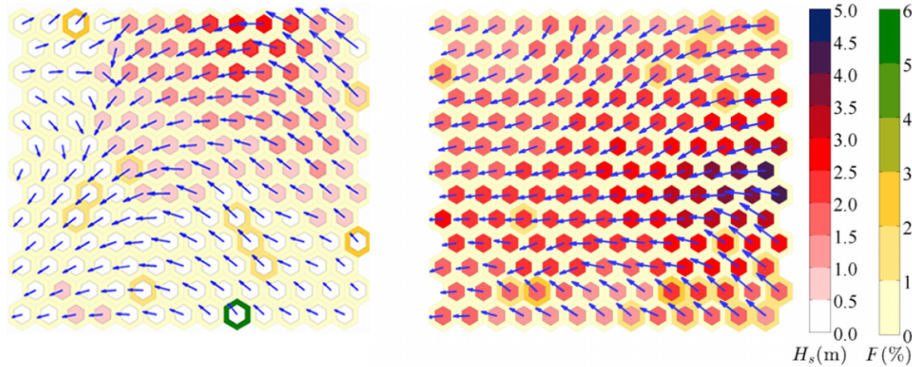Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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